NEXT GENERATION INTERNET

Grant Agreement No.: 957216
Call: H2020-ICT-2018-2020

Topic: ICT-56-2020
Type of action: RIA

inGenious

D4.1 Multi-Technologies Network for IoT

Revision: v1.0

Work package	WP 4
Task	Task T4.1, T4.2 and T4.3
Due date	30/04/2021
Submission date	30/04/2021
Deliverable lead	TUD
Version	1.0
Editors	Roberto Bomfin (TUD), Javier Renart (UPV)
Authors	Roberto Bomfin (TUD), Ivo Bizon (TUD), Ahmad Nimr (TUD), Christos Politis (SES), Juan Jose Garrido Serrato (SES), Miguel Cantero (5CMM), Manuel Fuentes (5CMM), David Martín-Sacristán (5CMM), Gino Ciccone (TEI), Giuseppina Carpentieri (TEI), Anton Luca Robustelli (TEI), Jose Costa-Requena (CMC), José Luis Cárcel (FV), Pietro Piscione (NXW), Erin E. Seder (NXW), Giacomo Bernini (NXW), Eddy Higgins (iDR), Jahveen Davis (iDR), Shane Bunyan (iDR), Javier Renart (UPV), Nuria Molner (UPV)
Reviewers	Mika Skarp (CMC), Waqar Anwar (TUD), Joe Cahill (iDR), Carlos Alcaide Pastrana (TID), Gino Ciccone (TEI), Carsten Weinhold (BI), Nuria Molner (UPV)

Abstract	This document describes the approach of iNGENIOUS to develop a heterogeneous connectivity architecture by aggregating different existing and forthcoming IoT technologies. In short, this deliverable targets at explaining the state of the current technologies applied to internet-of-things (IoT) networks and the planned innovations of iNGENIOUS. In particular, this deliverable covers three aspects of the network, namely, i) radio access network, ii) core network, and iii) slicing and orchestration.
Keywords	Radio Access Network, Core Network, Network Slice Orchestration, 5G, Internet-of-Things

Document Revision History

Version	Date	Description of change	List of contributor(s)
V1.0	30/04/2021	EC Version	Roberto Bomfin (TUD), Javier Renart (UPV)

Disclaimer

Copyright notice

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:	R	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to iNGENIOUS project and Commission Services	

** R: Document, report (excluding the periodic and final reports)*

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

Executive Summary

This document describes how iNGENIOUS plans to develop a heterogeneous connectivity solution that aggregates different existing and forthcoming IoT technologies. In particular, this deliverable has the objective of describing the state of the art (SoA) technologies applied to internet-of-things (IoT) networks and the planned innovations of iNGENIOUS in this regard, which are also mapped to the use cases of iNGENIOUS. The document covers three aspects of the network, namely, radio access network (RAN), core network and network slicing, which are briefly summarized in the following

- Regarding the RAN, several radio access technologies (RATs) are considered to support a variety of IoT devices with diverse data usage, coverage and power consumption. This leads to a number of radio access network (RAN) technologies such as NB-IoT, 5G, LoRaWAN, Sigfox, flexible RAN and satellite communications.
- In the Network Core side, additional functions for time sensitive networking (TSN) are considered to integrate TSN fixed devices as part of the next generation of 5G-IoT networks. Moreover, this document provides detail on the required interfaces between 5G Core (5GC) and software defined network (SDN) components to deliver the required adaptive networking mechanisms of IoT applications.
- Lastly, network slicing solutions for such a complex network with high number of devices is covered. For instance, a network slicing extension of current Management and Orchestrations (MANO) architectures are considered, where workflows for slicing and integration with 5GC and next generation (NG)-IoT are described as well as machine learning slice optimization strategies.

Co-funded by the Horizon 2020
Framework Programme of the European Union

Table of Contents

Co-funded by the Horizon 2020 Framework Programme of the European Union

List of figures

Co-funded by the Horizon 2020 Framework Programme of the European Union

Co-funded by the Horizon 2020 Framework Programme of the European Union

List of tables

Co-funded by the Horizon 2020
Framework Programme of the European Union

Abbreviations

AD	Analog-to-Digital
AF	Application Function
AI	Artificial Intelligence
AMF	Access Mobility Function
AGV	Automated Guided Vehicle
BBU	Baseband unity
CDMA	Code-Division Multiple Access
COTS	Commercial off-the-shelf
CNFDs	Containerized Network Function Descriptor
CP	Cyclic prefix
C-RAN	Cloud RAN
CRC	Cyclic Redundancy Check
CSMA	Carrier Sense Multiple Access
CU	Centralized Unity
DA	Digital-to Analog
DFT	Discrete Fourier Transform
DL	Downlink
DN	Data Network
DU	Distributed Unit
DVL	Data Virtualization Layer
eCPRI	Common Public Radio Interface
Edge-NFVI	Edge-Network Functions Virtualization Infrastructure
EIRP	Effective Isotropic Radiation Power
ENI ISG	Experiential Networked Intelligence Industry Specification Group
EM	Eletromagnetic
EMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
FDMA	Frequency-Division Multiple Access
GEO	Geostationary

Co-funded by the Horizon 2020 Framework Programme of the European Union

GFDM	Generalized Frequency Division Multiplexing
gNB	gNodeB
GPU	Graphics Processing Unit
HTS	High Throughput Satellite
IGW	Intelligent Gateway
IoT	Internet of Things
IP	Internet Protocol
IQ	In-phase and Quadrature
LoRa	Long Range
MA	Multiple Access
MANO	Management and Orchestration
MBO	Mobile Backhaul Orchestrator
MCS	Modulation and Coding Scheme
MEC	Multi-access Edge Computing
MEO	Medium Earth Orbit
MIMO	Multiple-Input Multiple-output
MIOT	Massive Internet of Things
ML	Machine Learning
MMTC	Massive Machine Type Communication
MNO	Mobile Network Operator
NEF	Network Exposure Function
NFV	Network Functions Virtualization
NFVO	Network Functions Virtualization Orchestrator
NG-IoT	Next-Generation IoT
NG-RAN	Next Generation Radio Access Network
NMS	Network Management System
NR	New Radio
NRF	Network Repository Function
NSDs	Network Service Descriptors
NSIs	Network Slice Instances
NSMF	Network Slice Management Function

Co-funded by the Horizon 2020 Framework Programme of the European Union

NST	Network Slice Template
NSSF	Network Slice Selection Function
NWDAF	NetWork Data Analytics Function
OFDM	Orthogonal Frequency Division Multiplexing
PDCP	Packet Data Convergence Protocol
PC	Personal Computer
PCF	Policy Control Function
PDU	Protocol Data Unit
PLCP	Physical Layer Convergence Protocol
PLMN	Public Land Mobile Network
PNFDs	Physical Network Function Descriptors
QoS	Quality-of-service
RAN	Radio Access Network
RIC	RAN Intelligent Controller
RT	Real-time
RU	Radio Unity
SA	Stand-alone
SatRAN	Satellite Radio Access Network
SBA	Service Based Architecture
SDMA	Space-Division Multiple Access
SDN	Software Defined Network
SDX	Software Defined Everything
S-GW	Serving Getaway
SLA	Service Level Agreement
SMF	Session Management Function
SMO	Service Management and Orchestration
SNO	Satellite Network Operator
SoA	State of the art
SVM	Support Vector Machine
TDMA	Time-Division Multiple Access
UC	Use Case

Co-funded by the Horizon 2020
Framework Programme of the European Union

UDM	Unified Data Management
UE	User Equipment
UL	Uplink
UPF	User Plane Function
URLLC	Ultra Reliable Low Latency Communications
VIM	Virtualized Infrastructure Manager
VM	Virtual Machine
VNFDs	Virtual Network Function Descriptors
V-RAN	Virtual RAN
VSB	Vertical Service Blueprint
VSD	Vertical Service Descriptor
VSI	Vertical Service Instance
VSMF	Vertical Service Management Function
WP	Work Package
5GC	5G Core
5G-NR	5G New Radio

Co-funded by the Horizon 2020 Framework Programme of the European Union

1 Introduction

1.1 Objective of this Deliverable

This deliverable has the objective of describing the state of the art (SoA) of current technologies applied to internet-of-things (IoT) networks and the planned innovations of iNGENIOUS within work package (WP)4. In iNGENIOUS, different network requirements are covered in different Use Cases (UCs). Therefore, several radio access technologies (RATs) are explored in this document such that these requirements can be fulfilled. In this way, a Smart Multi-Technology IoT radio access network (RAN) is achieved which supports many kinds of IoT devices with a variety of data usage, coverage and power consumption. Furthermore, the Core Network is responsible for providing end-to-end data transmission by bridging different RAN to applications. Basically, the Core Network has the control over the RAN and devices by managing the traffic among the different nodes in the network. As the number of user equipment (UEs) start to increase, the resource management and allocation in the network becomes more critical. To this end, the Next-Generation IoT (NG-IoT) network slice orchestration supported by an artificial intelligence (AI) and machine learning (ML) platforms play a decisive role, performing semi-automated decisions on resource allocation and management. In summary, this document described the planned work of iNGENIOUS related to three aspects of the network, i) radio access networks (RAN), ii) Core Network, and iii) network slice and orchestration. These components are part of the iNGENIOUS network architecture depicted in Figure 1.

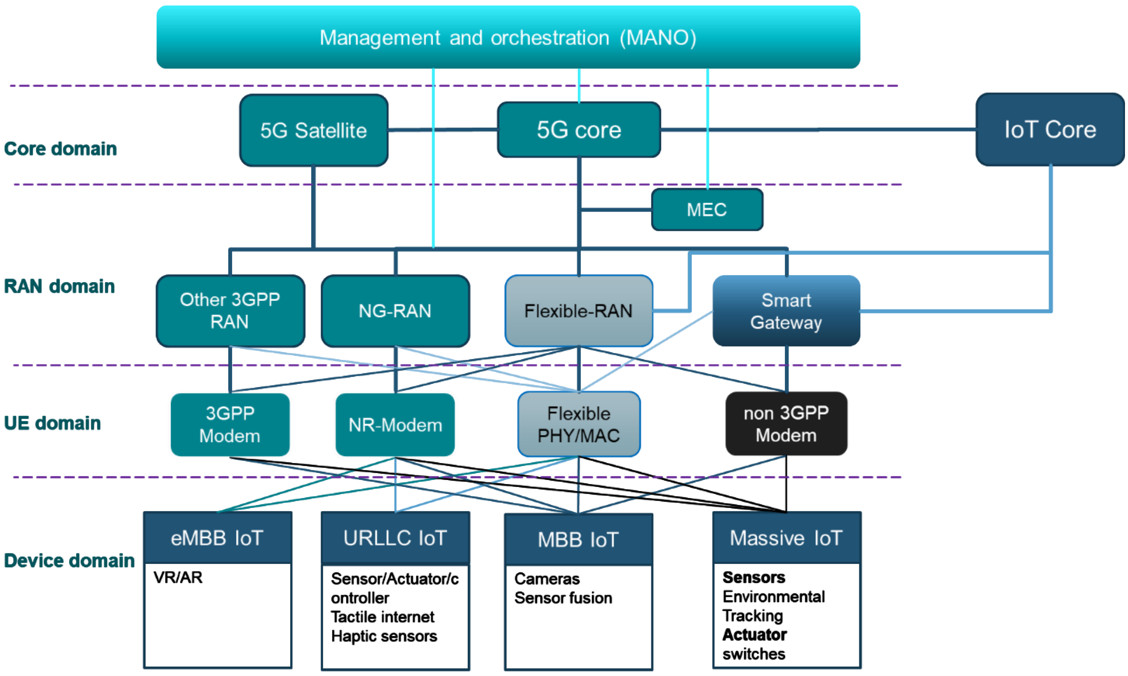


Figure 1: iNGENIOUS network architecture

Lastly, this document also has the objective of describing how the work of WP4 in iNGENIOUS will contribute to the development of use cases (UCs).

Co-funded by the Horizon 2020 Framework Programme of the European Union

1.2 Role of WP4 in iNGENIOUS

WP4 aims at developing a heterogeneous connectivity architecture in order to aggregate different existing and forthcoming IoT technologies, such that the requirements of various traffic classes involved in the use cases are assessed. Figure 2 depicts the UC scenarios of iNGENIOUS with various traffic classes, which are described in detail in Deliverable D2.1 [1].

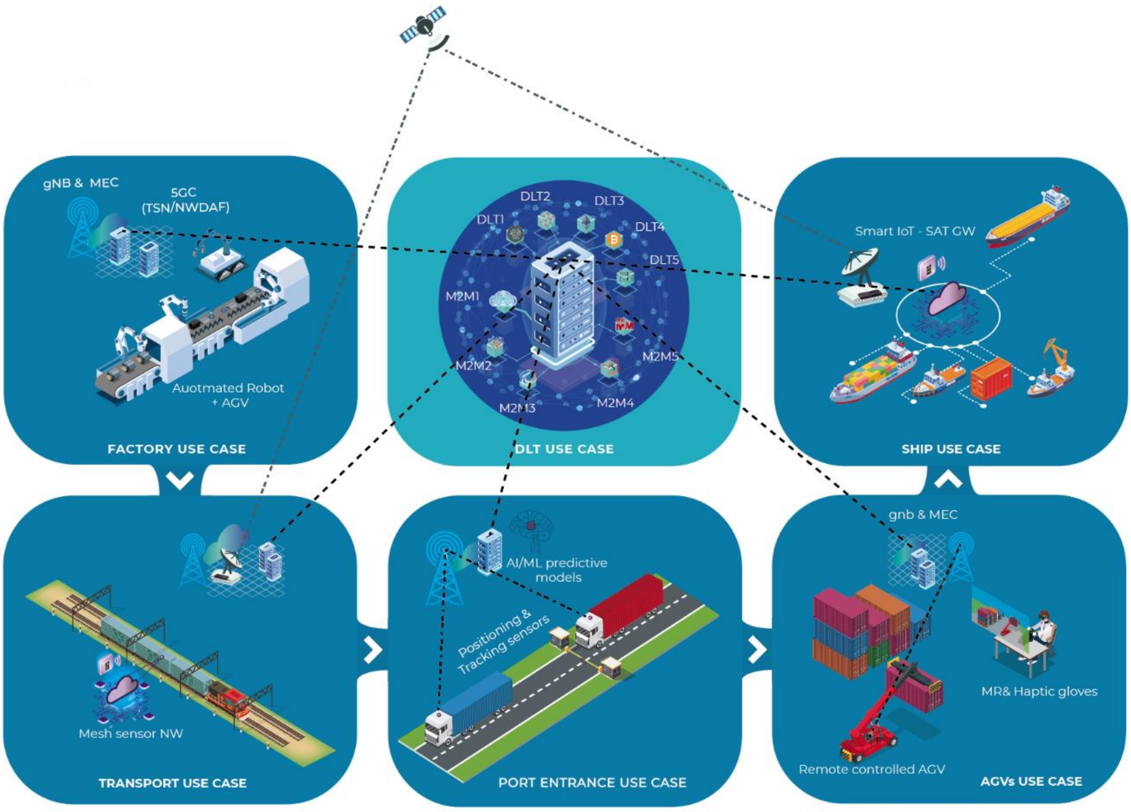


Figure 2: iNGENIOUS next generation supply chain use cases

In order to construct such a heterogeneous network, iNGENIOUS approaches this challenge by integrating a variety of radio access technologies (RATs) and communication protocols in a single and unified network architecture. In this manner, the network resources can be cooperatively shared among the different RATs. Naturally, the concept of such network requires a new level of flexibility, reconfigurability, resilience and resource awareness. These requirements are tackled by software defined network (SDN) techniques to enable smart and semiautonomous network slice orchestration.

Additionally, a set of objectives has been defined in WP4, which are:

- to design and develop a **flexible RAN architecture** to support the forthcoming IoT network requirements.
- to design **new functional blocks** to support smart and semiautonomous decisions based on ML algorithms.

Co-funded by the Horizon 2020
Framework Programme of the European Union

- to design and develop **extensions to current slicing and orchestration frameworks** on IoT centric scenarios.
- to develop **new protocols and architectural blocks** to improve the support of IoT UC over shared infrastructures.

As it has been described in the last section and Is seen In Figure 1, iNGENIOUS approaches the IoT networking by developing innovations for the RAN, Core Network, and network slice and orchestration. In the follows the approach of iNGENIOUS with respect to these components are briefly commented.

Regarding the RAN, this document provides details on how iNGENIOUS plans to improve the network support of diverse radio technologies such as NB-IoT/LTE-M, 5G, LoRaWAN, SigFox, flexible RAN, etc. In addition, a special attention is given to satellite networks in order to enable connectivity in remote areas. Lastly, Itis shown how machine learning (ML) can be applied for network optimization.

The aim of iNGENIOUS with respect to the core network is to provide network functions that deliver a dynamic IoT platform. As such, it is considered existing modules of the 5GC control plane which follow a service based architecture (SBA). Then, additional functions for time sensitive networking (TSN) are considered to integrate TSN fixed devices as part of the next generation of 5G-IoT networks. Moreover, this document covers the required interfaces between 5G Core (5GC) and SDN components to deliver the required adaptive networking mechanisms of IoT applications.

The work of iNGENIOUS related to network slicing will extend current Management and Orchestrations (MANO) architectures to support a new set of requirements defined in the project. As such, the workflows for slicing and integration with 5GC and NG-IoT is described as well as ML slice optimization strategies considered in this project.

This deliverable provides a state of the art on all the specific technologies adopted by iNGENIOUS, but most importantly, the innovations the project expects to introduce in these technologies to fulfil the use cases requirements. WP4 will work on implementing and improving three main aspects which affect network performance and capabilities: radio access network, core network and network slice orchestration.

1.3 Role WP4 in the Use Cases

The goal of this section is to describe how WP4 is related the the UCs depicted in Figure 1 and defined in the deliverable 2.1 (D2.1) of iNGENIOUS [1]. The six UCs combined have the object of demonstrating how the technical innovations of iNGENIOUS will enable digitalization and monitoring of the supply chain ecosystem as a whole, enhancing factory operations, transportation and maritime port operation. In the following, the six UCs and their relations to WP4 are briefly described. A more detailed description of the UCs is found in D2.1 [1].

1.3.1 Automated robots with heterogeneous networks - Factory UC

Co-funded by the Horizon 2020 Framework Programme of the European Union

This UC considers automated robots in smart factories that make use of the tactile internet. In this scenario, the sensors and machinery synchronously work with latencies of few milliseconds, where the interoperability with wired TSN environments are also explored. This UC will enable an automated robot control in industrial environments based on the design of a smart distributed application that will leverage different types of sensors, actuators and parallel control loops for connecting machines and humans.

The contributions of WP4 to this UC include the 5G IoT layer, that has a special focus on software defined communications in PHY and MAC, NFV and time synchronization. Additionally, the WP4 will consider MANO and Slicing Service for the automated deployment of slices for covering TSN requirements. The particular innovations are:

- Flexible PHY/MAC (Section 2)
- AI/ML for RAN (Section 2)
- Network Slice Manager (Section 3)
- 5GLAN AF (Section 3)
- TSN AF (Section 3)
- SDN MBO (Section 3)
- End-to-end network slice provisioning (Section 4)
- Orchestration and slicing of Flexible RAN and 5G RAN (Section 4)
- Orchestration and data collection from 5G Core (Section 4)
- Orchestration of applications and NFs at edge/MEC (Section 4)
- AI-assisted (network-data based) slice optimization (Section 4)

The SoA technologies related to WP4 used in this UC are given below:

- 5G NR-RAN (Section 2)
- 5G UE connectivity for automated guided vehicle (AGV) (Section 2)

1.3.2 Transportation platforms health monitoring - Transport UC

This UC has the objective of showing that asset health tracking can lead to low operational costs and high asset availability, which uses new data-based services provided by low-power edge distributed network and intelligent sensor modules installed in the transportation platforms. Specifically, this UC will enable the monitoring of health for transportation platforms with a particular focus on railway transportation. The monitoring of health conditions will reside in the design of neuromorphic sensors able to gather and process data on the edge of the network while bringing low cost and power and high life expectancy. For enabling the exchange of data between smart edge sensors and platforms, near continuous connectivity to the edge will be enabled by exploiting terrestrial and non-terrestrial networks.

Co-funded by the Horizon 2020 Framework Programme of the European Union

The technical development of WP4 that will contribute to this UC include the Smart IoT GW, that has the following innovation:

- Network Slice Manager (Section 3)
- End-to-end network slice provisioning (Section 4)
- Orchestration of satellite backhaul (Section 4)

The SoA technologies related to WP4 used in this UC are given below.

- LoRa (Section 2)
- Satellite Backhaul (Section 2)
- Orchestration of applications and NFs at edge/MEC (Section 4)

1.3.3 Situational understanding and predictive models in smart logistics scenarios - Port Entrance UC

This UC targets the development of AI/ML-based predictive models to estimate and optimise truck turnaround times for optimising the access and reduce the wait for vehicles at the port accesses, leading to corresponding savings on direct costs for carriers. This will enhance the situational understanding of events in maritime ports and terminals by means of collecting and aggregating data processing. By analysing the available data sources, the use case will optimize and predict processes for reducing the time that trucks spend inside the port and terminal facilities, i.e., truck turnaround times (TTT).

The contribution of WP4 in this UC consists of setting the connectivity of IoT tracking devices. The particular innovations are:

- Network Slice Manager (Section 3)
- 5GLAN AF (Section 3)
- SDN MBO (Section 3)
- End-to-end network slice provisioning (Section 4)
- AI-assisted (IoT/DVL-data based) slice optimization (Section 4)

1.3.4 Improved driver's safety with mixed reality and haptic solutions - AGVs UC

This UC aims at improving the safety conditions of workers in maritime ports and terminals, by keeping the employees away from hazardous environments such as fuel port terminals using innovative wireless networks. The idea is that the workers operate AGVs remotely using 5G wireless communications, MR and haptic solutions, which will be integrated in an immersive remote indoor cockpit. 5G will ensure high throughput and low latency connectivity with the cockpit, while MR and haptic solutions will provide an immersive experience to the worker. Additionally, the project will also explore the deployment of a novel TSN application function that maps 5G network slicing with fixed TSN Central Network Controller (CNC) polices.

Co-funded by the Horizon 2020 Framework Programme of the European Union

The technical innovation of W4 in this UC consists of:

- mmWave deployment (Section 2)
- End-to-end network slice provisioning (Section 4)

Additionally, WP4 will contribute to this UC with SoA 5G infrastructure through the deployment of a 5G-based private network for enabling AGV remote connectivity. In particular, the SoA technologies employed in this UC are:

- 5G modem for the connection of the AGVs to the 5G network in the sub-6GHz band (Section 2)
- Network Slice Manager (Section 3)
- Orchestration and slicing of 5G RAN (Section 4)
- Orchestration of applications and NFs at edge/MEC (Section 4)

1.3.5 Inter-modal asset tracking via IoT and satellite - Ship UC

The goal of this UC is to provide end-to-end (E2E) intermodal asset tracking with IoT and satellite connectivity for enabling enhanced real-time monitoring of shipping containers when they are transported in both terrestrial and maritime segments. In order to achieve this objective, this UC will explore communications via satellite backhaul and IoT terrestrial infrastructure, enabling real-time monitoring of cargo parameters when containers are sailing on the sea and when they approach coast. To enable the ubiquitous coverage, sensors and IoT tracking devices will be installed on the shipping containers transported by ships and trucks on both segments, whose connectivity will be provided by the Smart IoT GW.

The contribution of WP4 to this UC consists of providing the Smart IoT GW for enabling satellite and IoT connectivity. In particular, the technical innovations involved in this UC are:

- The Smart IoT Gateway (Section 2)
- 5GLAN AF (Section 3)
- SDN MBO (Section 3)
- End-to-end network slice provisioning (Section 4)
- Orchestration of satellite backhaul (Section 4)

The SoA technologies related to WP4 employed in this UC are:

- LoRa (Section 2)
- Satellite Backhaul (Section 2)
- Network Slice Manager (Section 3)

Co-funded by the Horizon 2020 Framework Programme of the European Union

1.3.6 Supply chain ecosystem integration - DLT UC

In this UC, the project aims at overcoming the absence of a virtual interoperability between existing M2M and DLT solutions through the development of interoperable IoT and DLT layers. The idea is to provide two different interoperable layers in order to abstract the complexity of the underlying M2M platforms and DLT solutions, guaranteeing at the same time data privacy and security by means of encoding and anonymization techniques. These layers will be capable of securely and semantically exchange the information flows between the different actors that can take part along the supply chain ecosystem. This could be used in different segments of the supply chain such as manufacturing, transportation or logistics.

Although most of the technical contribution in this UC will come from WP5 as stated in D2.1, WP4 will contribute with:

- Orchestration and data collection from 5G Core (Section 4)
- AI-assisted (IoT/DVL-data based) slice optimization (Section 4)
- AI-assisted (network-data based) slice optimization (Section 4)

1.4 Structure of the Document

This deliverable is organized in three main sections, which are followed by a conclusion in the end of the document. The three main sections are described below.

- Section 2, Multi-Technology IoT Radio Access Networks: this section describes the Radio Access Technologies that will be used in the iNGENIOUS project, which will be different according to each Use Case requirements.
- Section 3, Smart IoT Core Network: this section describes the functionality provided by the mobile packet core to deliver reliable connectivity for IoT devices to deliver data to either local edge computing, cloud services or fixed local area networks.
- Section 4, NG-IoT Network Slice Orchestration: this section describes the key innovation for what concerning the NG-IoT Network Slice Orchestration, the building block composing the orchestrator and the main interaction with the iNGENIOUS components realised in the supply chain and industrial IoT scenarios.

Co-funded by the Horizon 2020
Framework Programme of the European Union

2 Smart Multi-Technology IoT Radio Access Networks

The wide variety of use cases in INGENIOUS have different network requirements. The Factory Use Case needs high data rates and low coverage, while the Ship Use Case requires low data rates and coverage in remote areas where cellular connectivity is not available. To accomplish all the KPIs, which are fully described In D2.1, we must explore different Radio Access Technologies (RATs) with the goal of implementing the most relevant to us, and innovating in the technologies which are not mature enough to fulfil iNGENIOUS objectives. By following this approach, we'll achieve a Smart Multi-Technology IoT RAN, which will support all kind of IoT devices with divergent requirements in terms of data usage, coverage and power consumption.

2.1 Access Technologies

One technology is not cable of serving all the projected applications and volumes for IoT. WiFi and BTLE are widely adopted standards and serve the applications related to communicating personal devices quite well. Cellular technology is a great fit for applications that need high data throughput and have a power source. LPWAN offers multi-year battery lifetime and is designed for sensors and applications that need to send small amounts of data over long distances a few times per hour from varying environments.

2.1.1 3GPP RAN

Radio Access Networks (RAN) provide radio access and assist to coordinate network resources across wireless devices [1]. User equipment (UE) connects to the cellular network via the RAN. The RAN architecture has evolved with each new generation of mobile networks up to what today is known as 5G or NG-RAN. Some of the main novelties that were introduced over the years were the use of multiple antennas (MIMO), ultra-high bandwidth, extremely low latencies, carrier aggregation or edge computing, among others.

Traditional 3GPP radio networks consist of three major components [3]:

- The baseband, responsible of a series of computing and signal processing functions to enable wireless communications.
- The radio, which translates the digital generated signal into analogue RF signals, modulated and transmitted in a particular frequency band.
- The antenna, which radiates electrical RF signals into the air.

One of the key elements of the NG RAN architecture is the support of centralised processing in Cloud-RAN, with protocol functionality split of NR base station, i.e., gNB, in central units (CUs) and distributed units (DUs). Such novel architecture enhancements provide a significant opportunity to design an innovative RAN architecture for delivering content in 5G [4]. Figure 3 shows the different modules in the RAN architecture for 4G LTE and 5G NR.

Co-funded by the Horizon 2020 Framework Programme of the European Union

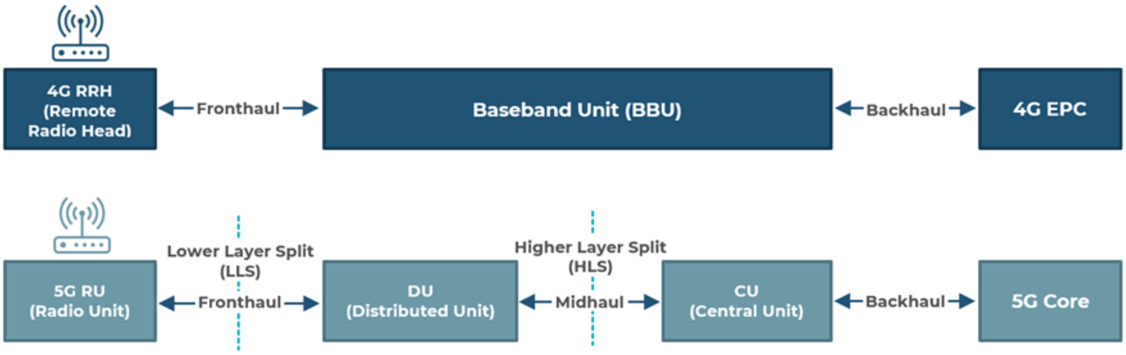


Figure 3: 4G vs 5G RAN architecture

In the figure, this new DU/CU separation can be clearly observed. In 4G, all layers are managed by the BBU (Baseband Unit). This unit handles the functions of the PDCP, RLC, MAC and PHY, while the RF functions are done in the RRH (Remote Radio Head). In 5G, the functions carried out by the RAN are divided into CU, DU and RU (Radio Unit). The Fronthaul is between the RU and DU, while the Midhaul is between the DU and CU and the Backhaul between the CU and 5G Core.

The gNB functions are split into CU and DU, where CU covers higher layer protocol functions of Service Data Adaptation Protocol (SDAP) and Packet Data Convergence Protocol (PDCP), and DU covers lower layer protocol functions such as RLC, MAC and PHY (high-level functions). At the same time, the RU not only performs the RF functions but also the low-level part of the PHY, see Figure 4.

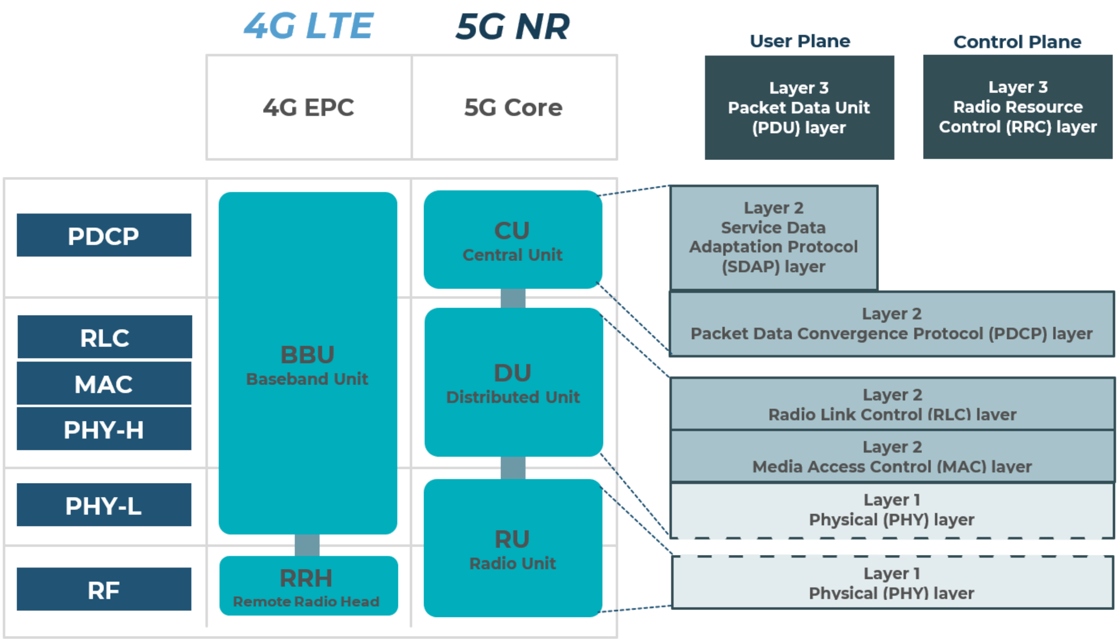


Figure 4: RAN architecture and layers

The fact that the protocol layers are divided into different units provides flexibility to the NR architecture. This consequently creates scenarios where parts of the RAN and 5GC are geographically co-located, and parts of the RAN are placed close to the edge.

Co-funded by the Horizon 2020
Framework Programme of the European Union

In the following pages, we provide a comprehensive description of both 4G and 5G RAN architectures and their interfaces.

2.1.1.1 4G LTE RAN

The 4G LTE RAN, also known as E-UTRAN (Evolved UTRAN), is formed by a group of eNBs interconnected with each other by means of the X2 interface [7][8]. As briefly introduced before, an eNB consist of two parts, i.e., BBU and RRH. The BBU manages almost all RAN functions, while the RF layer is handled by the RRH. The BBU consist of digital modules that process all the signals transmitted and received by the air interface and acts as an interface to the core through the Backhaul. The RRH is typically mounted to the radio tower and directly connected with the antennas and is responsible for modulation and demodulation all signal transmitted or received on the air interface.

eNBs send in turn information to the Evolved Packet Core (EPC), more concretely control data to the Mobility Management Entity (MME) and user data to the Serving Gateway (S-GW), through the S1 interface. This control and user plane separation (CUPS) is key in LTE. X2 and S1 Interfaces are divided into UP and CP, each one dedicated to the user and control aspects, respectively. This layout of the different architectural elements and their connections is illustrated in Figure 5.

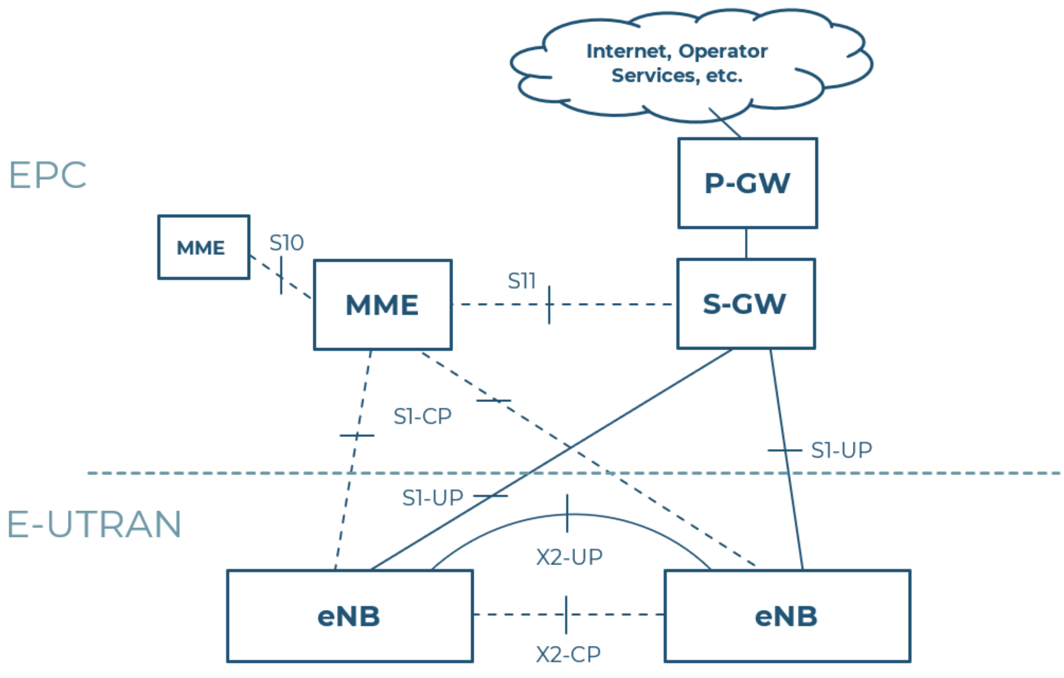


Figure 5: 4G RAN architecture

User plane protocols include PDCP, RLC, MAC and PHY, while the control plane part implements NAS, RRC, PDCP, RLC, MAC, PHY [7]. Figure 6 shows the protocol stacks for both cases.

Co-funded by the Horizon 2020 Framework Programme of the European Union

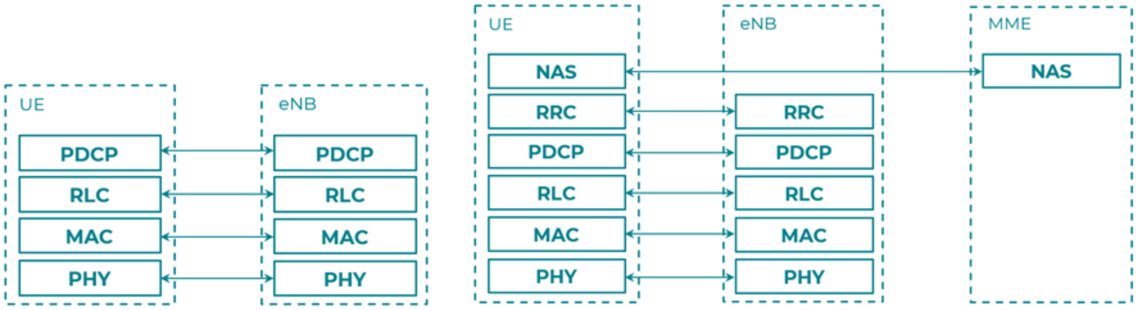


Figure 6: User plane (left) and control plane (right) protocol stacks in 4G LTE RAN [7]

2.1.1.2 5G NR-RAN

In 5G NR, the NG-RAN is defined as the system that interconnects the 5GC and the UEs. The 3GPP defines different deployment configurations, dividing them into two groups: SA (Stand-Alone) and NSA (Non-Stand-Alone). SA deployments use one cellular technology while NSA consist in a combination of technologies in the same network. This enables the transition of the standards. Here 5G SA will be considered combining 5G RAN and 5G Core.

The 5G RAN is formed by a group of gNBs, which communicate to each other through the Xn interface. Each gNB communicates in turn to the 5GC, more specifically to the AMF for control data and to the UPF for user data, by using the NG-C and NG-U interfaces, respectively. The NG-RAN architecture is illustrated in Figure 7.

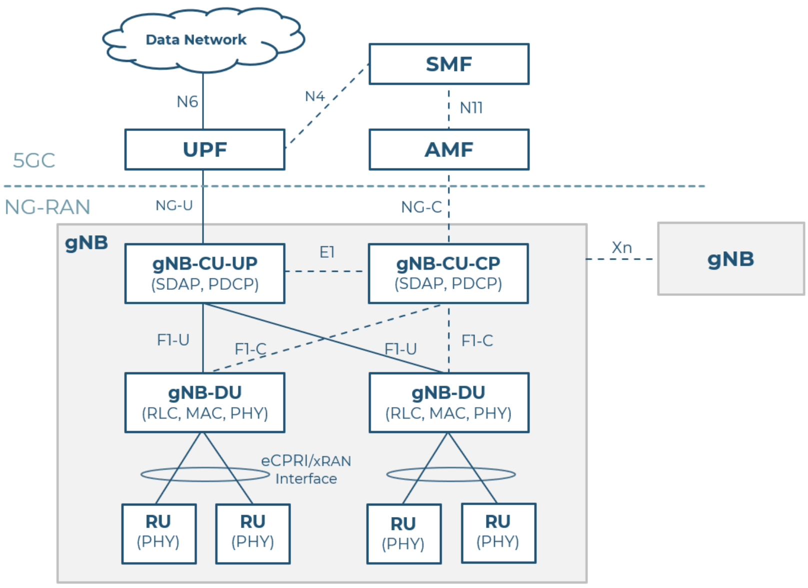


Figure 7: 5G RAN architecture

One of the main novelties in 5G is that each gNB is split into two parts, i.e., the gNB-CU (gNB-Central Unit) and the gNB-DU (gNB-Distributed Unit), connected to each other by the F1 interface. Note that a gNB only contains one CU but supports the use of multiple DUs. Through the eCPRI/xRAN interface, the DU is connected to one or several RUs, which manage the lower PHY layer and the RF functions. The RUs are directly connected to the antennas.

Co-funded by the Horizon 2020
Framework Programme of the European Union

A gNB may divide the CU into CU-CP and CU-UP for control plane and user plane usage, respectively. Each DU is connected to the CU-CP via the F1-C interface, and to the CU-UP via the F1-U interface. Note that there is just one CU-CP that controls the control plane, but there could be multiple CU-UPs for user plane utilisation.

The protocol stack in 5G includes SDAP, PDCP, RLC, MAC and PHY layers for user plane, while NAS, RRC, PDCP, RLC, MAC and PHY are implemented for the control plane. The protocol stacks are shown in Figure 8.

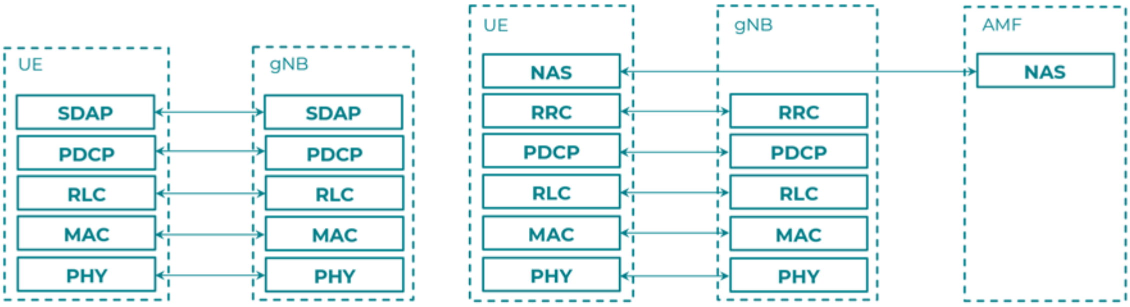


Figure 8: User plane (left) and control plane (right) protocol stack in 5G NG-RAN [5]

As it can be observed, the protocol stack in 5G is similar to 4G except for the addition of the Service Data Adaptation Protocol (SDAP), which is exclusive to 5G. Its main service and functions include: (i) mapping between a QoS flow and a data radio bearer, and (ii) marking QoS flow ID (QFI) in both DL and UL packets.

- **NG Interface**: NG is defined as the interface between the RAN and 5GC. It is divided into NG-U and NG-C, specific for the communication with UPF and AMF, respectively. A detailed description of NG-U and NG-C can be found in TS 38.410 [8]. The protocol stacks for the NG-U and NG-C interfaces are shown in Figure 9.

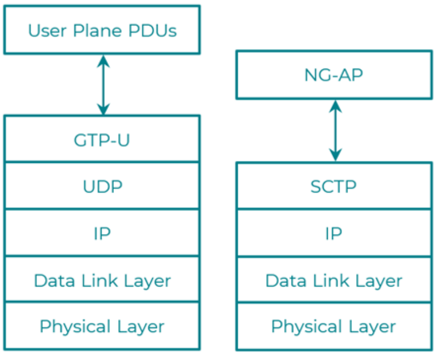


Figure 9: NG-U (left) and NG-C (right) protocol stacks in 5G NG-RAN [4]

Within the NG-U, the transport network layer is built on top of the IP layer, while GTP-U is used on top of UDP/IP to carry the user plane PDUs between the NG-RAN node and the UPF. NG-U provides non-guaranteed delivery of user plane PDUs between the NG-RAN node and the UPF.

The NG-C, on the other hand, builds the transport network layer on top of IP, but in order to ensure reliable transport of signalling messages,

Co-funded by the Horizon 2020 Framework Programme of the European Union

Stream Control Transmission Protocol (SCTP) is added on top of IP. The SCTP layer provides guaranteed delivery of application layer messages. The application layer signalling protocol is referred as NGAP (NG Application Protocol). At transport level, an IP layer point-to-point transmission is used to deliver the signalling PDUs.

- **Xn interface**: the Xn is defined as the interface between two gNBs. Following the same control and user plane separation, it is divided into Xn-C and Xn-U, respectively. Further details of Xn-U and Xn-C can be found in TS 38.420 [9]. The protocol stacks for these interfaces are shown in Figure 10.

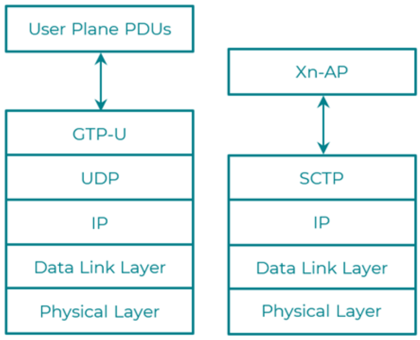


Figure 10: Xn-U (left) and Xn-C (right) Protocol Stack [4]

As done in the NG interface, in the Xn-U interface the transport network layer is built on IP transport and GTP-U is used on top of UDP/IP to carry the user plane PDUs. Xn-U provides non-guaranteed delivery of user plane PDUs and supports data forwarding as well as flow control.

In the Xn-C interface, on the other hand, the transport network layer is built on SCTP on top of IP. The application layer signalling protocol is called XnAP (Xn Application Protocol). The Xn-C interface supports the Xn interface management; UE mobility management, including context transfer and RAN paging; and dual connectivity.

2.1.2 Non-3GPP RAN

2.1.2.1 LoRa

Low Power Wide Area (LPWAN) solutions are able to offer multi-year battery lifetime for sensors and applications that need to send small amounts of data over long distances a few times per hour from varying environments. Typically, LPWAN solutions have to address critical aspects like network architecture, battery lifetime, robustness to interferences, network capacity (maximum number of nodes in a network), network security, variety of applications served, etc.

Within existing LPWAN solutions, LoRa (Long Range) is a physical layer solution created to enable long range communications while keeping low power consumption on devices. The main innovation of LoRa physical layer is the use of a chirp spread spectrum modulation (SS Chirp), which maintains the same low power characteristics as FSK modulation but significantly increases the communication range. Thanks to this feature, LoRa systems are

Co-funded by the Horizon 2020
Framework Programme of the European Union

able to cover entire cities or hundreds of square kilometres by deploying a single gateway or base station.

LoRA physical layer is complemented by LoRaWAN, the communication protocol and system architecture for the network. LoRaWAN aspects have the most influence in determining the battery lifetime of a node, the network capacity, the quality of service, the security, and the variety of applications served by the network.

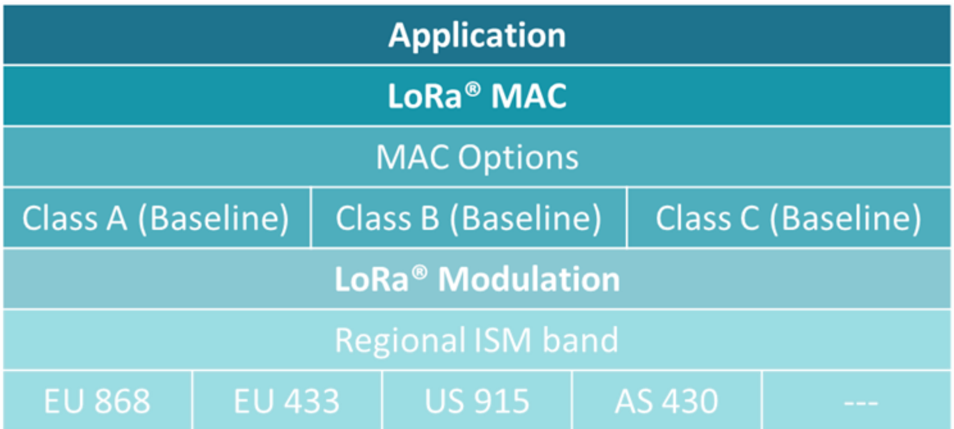


Figure 11: LoRaWAN System Architecture

Regarding network architecture, LoRa aims at exploiting long range star architectures since it helps to preserve battery life of devices when long-range connectivity is required. Long range star typologies are seen as an alternative to mesh network architectures, since the latter add complexity, reduce the network capacity and decrease the battery lifetime of devices as nodes have to receive and forward information for increasing the range.

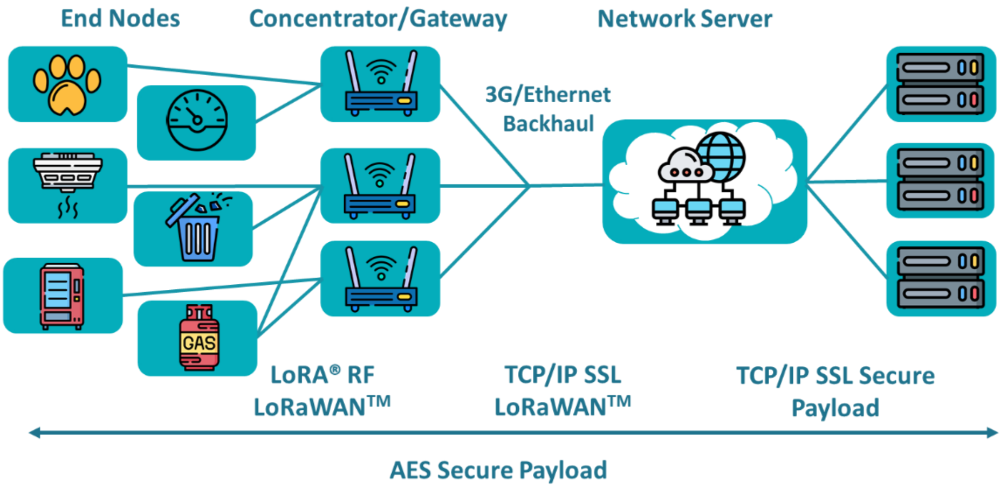


Figure 12: LoRaWAN Network Architecture

LoRaWAN end-nodes are designed to serve different applications and meet different battery life and latency requirements. As a consequence, three different devices classes are considered (see Figure 13):

- **Class A or Bi-directional end-devices** allow for bi-directional communications whereby each end-device's uplink transmission is

Co-funded by the Horizon 2020 Framework Programme of the European Union

followed by two short downlink receive windows. Class A is proposed as the best power consumption solution for applications that only require downlink communication from the server shortly after the end-device has sent an uplink transmission.

- **Class B or Bi-directional end-devices with scheduled receive slots** open extra receive windows at scheduled times. In order for the end-device to open its receive window at the scheduled time, it receives a time-synchronized beacon from the gateway. This allows the server to know when the end-device is listening.
- **Class C or Bi-directional end-devices with maximal receive slots** have almost continuously open receive windows, only closed when transmitting.

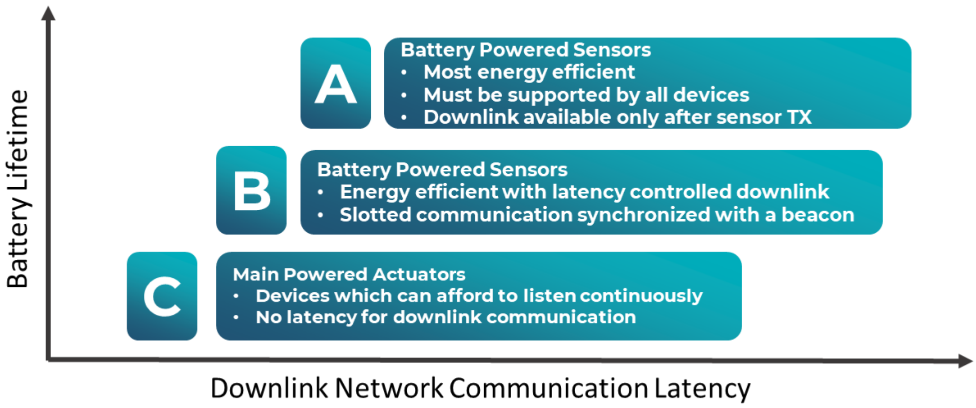


Figure 13: LoRaWAN End-Device Classes

Within the proposed star architecture, LoRaWAN network end-devices are not associated with a specific gateway. As a consequence, the data transmitted by a node is typically received by multiple gateways. Thanks to the proposed solution, if a node is moving there is no handover needed from gateway to gateway. At the same time, LoRaWAN nodes operate under the Aloha protocol, meaning that all nodes communicate the data asynchronously when the data is ready to be sent. This aspect contributes to reduce battery consumption associated with synchronization aspects.

In order to make a long range star network viable, gateways must have a very high capacity or capability to receive messages from a very high volume of nodes. LoRaWAN gateways exploit adaptive data rates and multichannel multi-modem transceivers to enable the reception of simultaneous messages on multiple channels. These features enable LoRaWAN networks to offer very high capacities while keeping network scalability through the addition of multiple gateways.

After receiving data from multiple nodes, gateways forward the received packet to the cloud-based network server via backhaul technologies like cellular, Ethernet, satellite, or Wi-Fi. The network server is the entity of the network in charge of managing and filtering redundant packets, performing security checks, schedule acknowledgments through the optimal gateway, and perform adaptive data rate, etc. Finally, network servers send the data to application servers by exploiting TCP/IP secure protocols.

Co-funded by the Horizon 2020 Framework Programme of the European Union

Regarding security, LoRaWAN utilizes two different layers: one for the network and one for the application. The network security ensures authenticity of the node in the network while the application layer of security ensures the network operator does not have access to the end user's application data. AES encryption is used with the key exchange utilizing an IEEE EUI64 identifier.

Finally, in terms of spectrum usage and regulatory requirements, LoRaWAN specification varies slightly from region to region as shown in Figure 14.

	Europe	North America	China	Korea	Japan	India
Frequency Band	867-869 MHz	902-928 MHz	470 – 510 MHz	920 – 925 MHz	920 – 925 MHz	865 – 867 MHz
Channels	10	64 + 8 + 8	In definition by Technical Committee	In definition by Technical Committee	In definition by Technical Committee	In definition by Technical Committee
Channel BW Up	125/250 kHz	125/500 kHz				
Channel BW Dn	125 kHz	500 kHz				
TX Power Up	+14 dBm	+20 dBm typ (+30 dBm allowed)				
TX Power Dn	+14 dBm	+27 dBm				
SF Up	7-12	7-10				
Data rate	250bps – 50 kbps	980 bps – 21.9 kbps				
Link Budget Up	155 dB	154 dB				
Link Budget Dn	155 dB	157 dB				

Figure 14: LoRaWAN Spectrum Allocation

In particular, in Europe, LoRaWAN defines ten channels, eight of which are multi data rate from 250bps to 5.5 kbps, a single high data rate LoRa channel at 11kbps, and a single FSK channel at 50kbps. The maximum output power allowed by ETSI in Europe is +14dBM, with the exception of the G3 band which allows +27dBm.

2.1.2.2 Sigfox

Sigfox is an ultra-narrowband and low power IoT technology used for transmitting non-periodic small data volumes (12 bytes) from sensors and devices. Sigfox exploits Binary Phase-Shift Keying (BPSK) and Gaussian Frequency Shift Keying (GFSK) modulation in tiny slices of spectrum within 868 MHz frequency band in Europe and 902 MHz in US.

The Sigfox communication stack is shown in Figure 15.

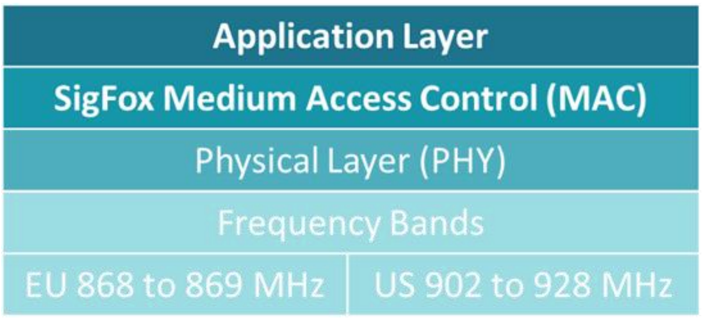


Figure 15: Sigfox Protocol Stack [10]

Co-funded by the Horizon 2020 Framework Programme of the European Union

Sigfox radio frequency layer includes communication channels. The PHY and MAC layers include the modulation scheme, the medium access control, as well as the error detection and the channel access. Finally, the application layer is defined according to user's requirements and specifications. Sigfox transmission mechanism uses redundancy so the data is transmitted for 3 times on 3 different channels (to ensure frequency diversity) at different time intervals of 400 ms (to ensure time diversity). This redundancy makes the communication more robust to interferences. Initially, Sigfox configuration was unidirectional where only the node could send data to the Gateway (uplink). The communication from the Gateway to the node (downlink) for ensuring bi-directional communications was added later [11].

Sigfox nodes can send a number of 140 messages per day with a maximum payload of 12 bytes and can receive from the application via the Gateway 4 messages per day with a maximum payload of 8 bytes. The Sigfox protocol aims at providing large autonomy through an extremely low energy consumption based in 'wake-up' strategy for sensors, allowing to achieve years of battery life. Additionally, devices do not require configuration and connection request or signalling procedures. Thanks to its low cost and ease of configuration, Sigfox is sometimes used as secondary solution to any other type of network, e.g.: Wi-Fi, Bluetooth, cellular, etc.

Sigfox relies on a dedicated radio-based network where devices and sensors exchange lightweight messages (12 bytes, excluding payload headers) with multiple Sigfox base stations.

Figure 16: Sigfox Network Architecture

In particular, the life cycle of transmissions over Sigfox architecture is the following:

1. A device wakes up and emits a message using its radio antenna.
2. Multiple Sigfox base stations in the area receive the message and send the message to the Sigfox Cloud. Base stations are composed of: local Sigfox antennas, a low-noise amplifier to amplify the signal and filter the noise, and an access point to understand Sigfox messages and send them to the cloud.
3. The Sigfox Cloud sends the message to a customer's backend platform.

All base stations are part of the Sigfox public network, which aims at providing global connectivity for this solution in exchange for a subscription fee (operator-based role).

2.1.2.3 Eltres

Co-funded by the Horizon 2020 Framework Programme of the European Union

ELTRES is Sony's proprietary solution to LPWA technologies. Its name comes from "The three" in Spanish, as it has 3 main characteristics, namely, i) log distance, ii) high-speed mobile performance, and iii) low power consumption.

ELTRES can cover long distances of over 100km, reducing the cost of deploying multiple base stations. It can also achieve good performance even when the transmitter is moving at speeds higher than 100km/h, which makes it especially useful for devices in cars or drones. Furthermore, its low power consumption makes it possible to use coin batteries as its power supply, maintaining a very small blueprint.

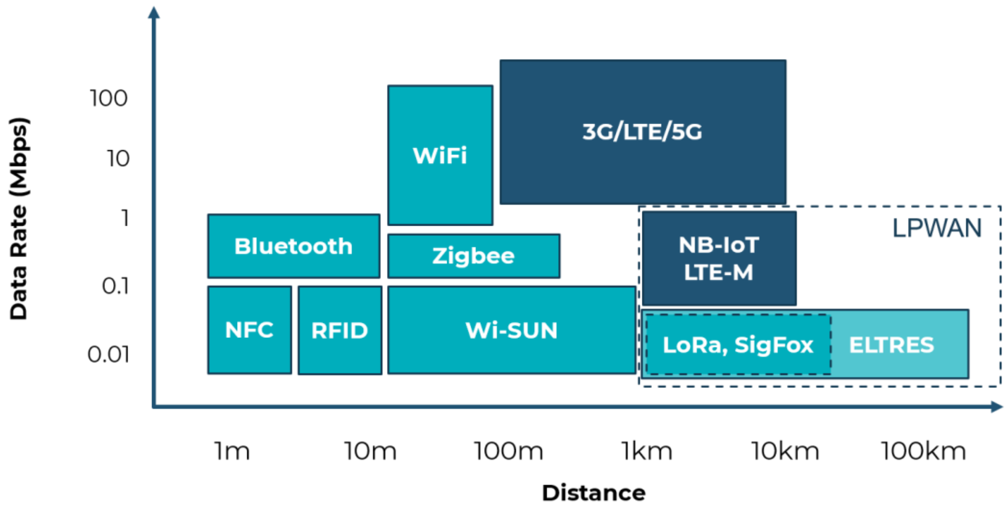


Figure 17: Wireless technologies comparison.

This technology aims at a similar market that Lora and Sigfox, but it brings some additional improvements that will make it more appropriate for certain use cases. It uses the open standard ETSI TS 103 357 and has built in GNSS both to track the IoT devices and to synchronise with the Base Stations thanks to the Time Of Day (TOD).

ELTRES is designed to work in scenarios with high Interference, where a large number of devices are connected via ELTRES or other LPWA technologies. It Is able to achieve a very low error rate of transmission, thanks to packet redundancy and maximum ratio combining, achieving up to 6dB signal gain.

2.1.3 Flexible RAN

RAN industry is slowly changing from vendor specific software/hardware to open Radio Access Networks which offer interoperable networks and will lower the barrier of entry for new vendors. Before reaching this point, it has gone through several iterations, like C-RAN and V-RAN.

2.1.3.1 RAN Virtualisation and centralisation

Cloud RAN (C-RAN) refers to a cloud-based evolution of the traditional RAN. C-RAN moves the Base Band Units (BBU) from the base station to the cloud, allowing for multiple remote radio heads (RRU) to be connected to a single BBU. C-RAN adds a high speed fronthaul (usually fibre optics) to interconnect these two entities.

Co-funded by the Horizon 2020
Framework Programme of the European Union

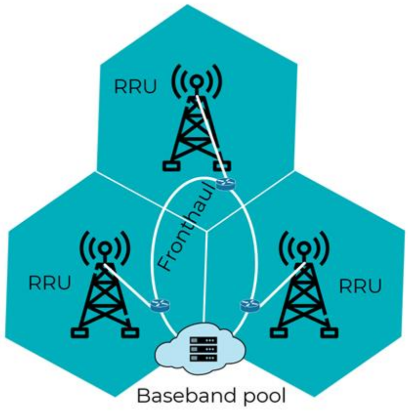


Figure 18: C-RAN concept

The next iteration of C-RAN appears by exploiting Software Defined Everything (SDx) and virtualisation which is called Virtual RAN (VRAN). Virtualisation facilitates the creation of logically isolated instances over abstracted physical hardware, which can be shared in a flexible, dynamic and efficient manner. This ability to share resource allows for flexible control, efficient resource usage and low cost.

There are two main virtualisation techniques that can be used in V-RAN: Hypervisor-based and container-based virtualisation. Hypervisor-based virtualisation i.e Virtual Machine (VM) runs a complete guest operating system which leads to a computationally resource intensive operation. On the other hand, the container-based virtualisation (usually Docker and Kubernetes) fosters the execution of specific software applications in isolated system environments called containers. This type of virtualisation is much lighter and efficient as it does not require a full deployment of an OS, as it runs on top of the host's OS kernel, independently of the operating system used in the container.

Currently, V-RAN is evolving towards the concept of Open RAN. This new concept looks to standardise the hardware and offer open-source software, which can be used on commercial, off-the-shelf hardware (COTS).

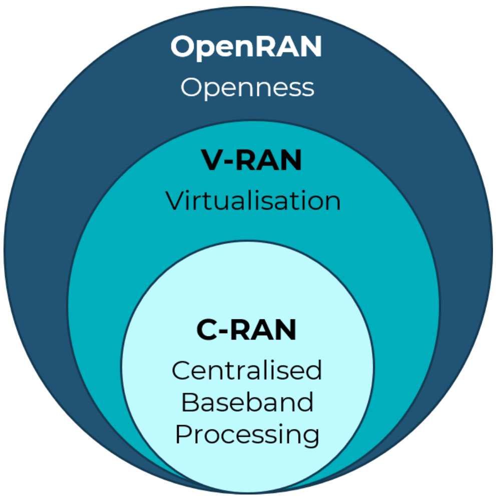


Figure 19: C-RAN evolution towards OpenRAN

The main goal of Open RAN is to have open interfaces, which allow equipment in every part of the network to interoperate between them, independently of which vendor creates it. There have been great standardisation efforts lately in this area, which have all conveyed to form the O-RAN alliance.

2.1.3.2 O-RAN

The O-RAN alliance is formed by more than 20 operators and 200 contributors from all over the world, which seek to build an open RAN for 5G networks. O-RAN aims to drive the mobile industry towards an ecosystem of innovative, interoperable, multi-vendor and autonomous RAN, with the benefit of reduced cost, improved performance and greater agility.

The success of O-RAN alliance in providing the evolution of the radio access networks, have been demonstrated recently. Four of the European largest operators signed a Memorandum of Understanding to express their commitment to rolling out Open RAN technology in the European markets. Orange for example, committed to only buy Open RAN technology from 2025. The O-RAN alliance has standardised all its architecture, including the different entities and open interfaces which communicate between them.

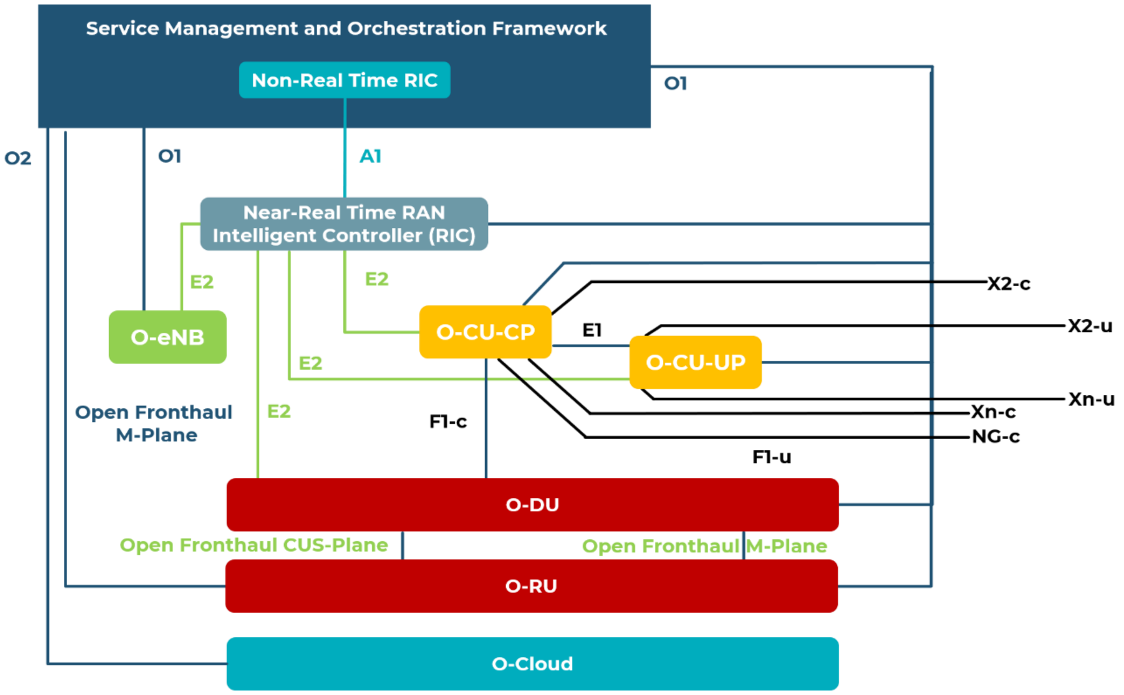


Figure 20: O-RAN overall logical architecture

The O-RAN architecture consists of 5 main entities, which are the Service Management and Orchestration Framework (SMO), the Near-Real Time RAN Intelligent Controller (RIC), the Centralised unit (O-CU), the Distributed unit (O-DU) and the Radio Unit (O-DU).

- **SMO:** Provides management services like core management, E2E slice management, orchestration, o-cloud management etc. It also hosts the Non-RT RIC.

Co-funded by the Horizon 2020 Framework Programme of the European Union

- **The Non-RT RIC:** Supports intelligent RAN optimization by providing policy-based guidance, ML model management and enrichment information to the near-RT RIC.
- **Near-RT RIC:** Enables near real time control and optimization of the E2 nodes functions and resources thanks to data collection and actions over the E2 interface.
- **O-CU:** Performs the layer 3 functions like RRC, PDCP, SDAP and it is divided in two logical nodes, one with the control plane and another with the user plane.
- **O-DU:** Connects multiple Radio Units to a Centralised unit and it performs different functions of high PHY layer, MAC and RLC, synchronization, Ethernet etc. (Downlink and uplink baseband processing, synchronization, signal processing)
- **O-RU:** Converts radio signals sent to and from the antenna to a digital signal that can be transmitted over the open fronthaul to the O-DU.

The main difference of O-RAN's architecture with the 3GPP architecture, is the inclusion of intelligent controllers, which seek to improve the RAN's autonomy by introducing machine learning algorithms that will allow the network to take smarter decisions. Furthermore, the O-DU has been divided in two different entities, the O-DU and the O-RU. This allows to reduce the hardware at the base stations, as the O-DU is brought to the edge cloud. Furthermore, several O-RU will be able to connect to a single O-DU over the open fronthaul, reducing the resources needed to manage and process the data coming from the O-RUs.

INGENIOUS' RAN architecture will consist of an O-RAN implementation using COTS hardware. This will allow to implement Machine Learning models to the decisions via Near-RT RIC applications, called xApps, which will be explained in further detail in section 2.4.

2.1.3.3 Flexible PHY/MAC

The BS functions can be split into processing and RF modules, as illustrated in Figure 21. The processing functions include logical protocols, MAC procedures for controlling the access to the wireless channel, resource allocation, and scheduling. Moreover, the PHY function performs the baseband processing and exchanges the IQ samples with the RF frontends, which include the DA/AD converters, analogue transceivers, and antennas. The flexible RAN is achieved by means of flexible RF frontends and a flexible PHY/MAC design. The frontend flexibility is supported in 5G and other standards by providing multiple frequency bands, where an RF frontend can be tuned to a carrier frequency and operates with different bandwidth. Thus, we focus on flexible PHY/MAC designs, which allow the reconfiguration of signals and multiuser multiplexing.

Co-funded by the Horizon 2020 Framework Programme of the European Union

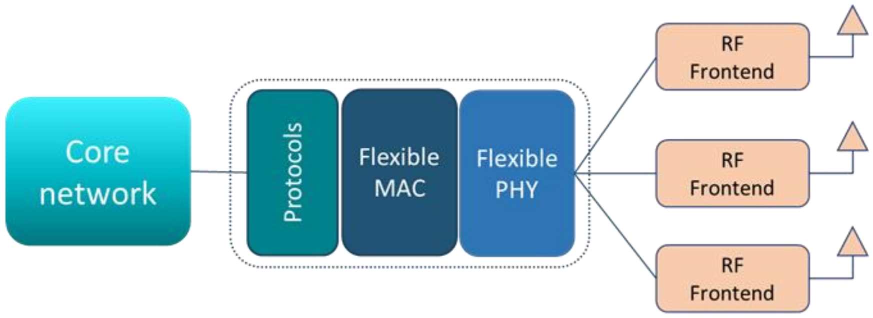


Figure 21: Flexible RAN architecture

Multiple access schemes: The MA schemes are part of the MAC sublayer and can be classified in two groups, packet-based, where one user gains the access to the channel for transmitting a packet of data within a short time interval such as in IEEE 802.11 standards. To coordinate the access, random access protocols are employed, e.g. ALOHA, CSMA and its variants, and token ring. The other category is channel-based MA, where the signals from different users are multiplexed in time, frequency and space. This scheme is used in 4G and 5G cellular networks. First, random access techniques are used to initiate the access to the network, and after that resource allocation and scheduling are employed for sharing the resources. The MAC also provides mechanisms to exchange control information about the data frame, which is influenced by the MA scheme.

In generic DL multiplexing, as illustrated in Figure 22, the data $\{\boldsymbol{b}_u\}$ from different U users are multiplexed in a scheduling slot, where the transmitted signal $x[n]$ is received through the corresponding u-th user channel h_u as $y_u[n]$. Based on prior information, the user decodes its targeted data. In the UL, each user encodes the data in a signal $x_u[n]$, and at the BS, the signals superimpose in a received signal $y[n]$, from which the BS decodes and demultiplex the individual user data. common multiplexing schemes are TDMA, CDMA, FDAM, and SDMA.

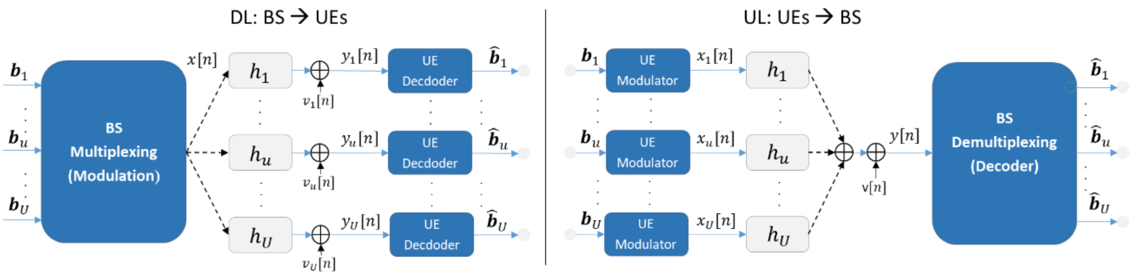


Figure 22: Generic multiple access

Flexible TDMA-based PHY/MAC: In TDMA schemes, a single transmitter, either the AP or the UE, occupies the whole spectrum for the transmission of a frame. The physical frame consists of preamble for synchronization and channel estimation, PLCP, and PUD signals. The PLCP, which convey information about the PHY parameters is encoded with predefined structure, whereas the PDU signal is flexibly generated by changing the PHY parameters, such as the MCS, and waveform parameters. The PDU is the logical MAC frame which contains the MAC header and payload from the upper layers.

Co-funded by the Horizon 2020 Framework Programme of the European Union

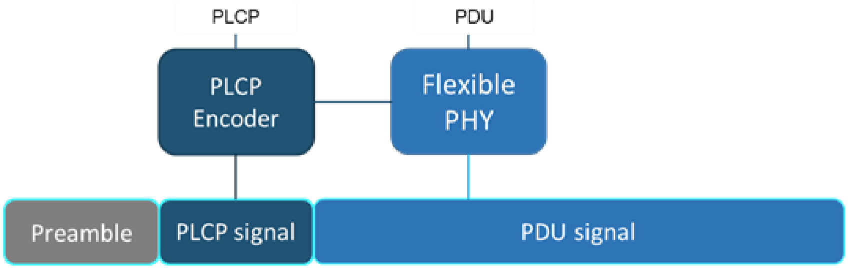


Figure 23: TDMA MAC

Extended Flexible PHY/MAC for 5G-NR: 5G-NR provides PHY flexibility by employing different numerologies for subcarrier spacing (OFDM symbol length), CP, different MCSs with predefined channel coding and mapping, as well as DFT precoding in the UL. The 5G MAC is based on combination of FDMA and TDMATDMA with flexible resource allocation. Previous studies show the potential of other waveforms in decreasing the OOB emissions and thus, reducing the required guard subcarriers [13], decreasing the PAPR [14], and achieving higher reliability by exploiting time and frequency diversity [15]. In fact, the linear waveforms can be seen as preceded-OFDM [16], and accordingly, the integration with 5G-NR requires elaborating flexible precoding. Employing this concept, the extended GFDM framework [17], which was developed in ORCA project [18], provides flexible precoding schemes inherited from the GFDM modem flexibility [19]. As shown in Figure 24, U users are multiplexed in a scheduling slot. The u-th user is assigned a set of time-frequency resources, $\mathcal{N}_u^{(d)}, \mathcal{N}_u^{(p)}$ for data and pilots, respectively. The bit stream from the upper layers is mapped to complex symbols vector $\boldsymbol{d}_u$using the bit-to-symbols block. This block summarizes the operations of channel coding, interleaving, and mapping. The MCS is selected from a predefined set depending on the channel status. Prior to the mapping to the resource grid, precoding is applied.

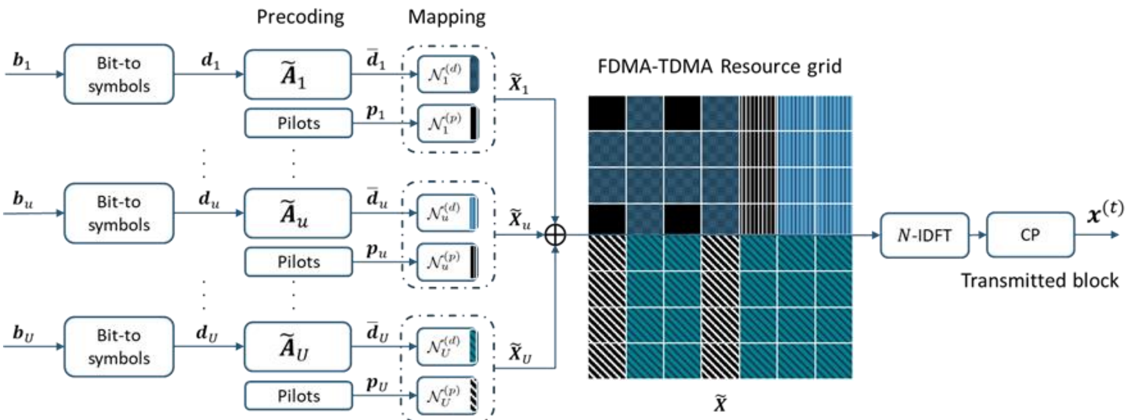


Figure 24: Flexible precoding in 5G-NR

The precoding matrix $\tilde{\boldsymbol{A}}_u$ follows the GFDM architecture shown in Figure 25 which is defined by the parameters M_u and K_u for the transform, in addition to the widowing matrices $\boldsymbol{W}_{u,tx}$ and $\boldsymbol{W}_{u,rx}$ for the precoding and postvoiding. Moreover, one or more of the DFT/IDFT transforms can be bypassed and, therefore, additional precoding schemes are directly supported by the GFDM extended precoder, such as OTFS [20]. This model extends the flexibility of 5G-NR, with minimum changes to the MAC procedures to exchange the

Co-funded by the Horizon 2020 Framework Programme of the European Union

precoding information in the control channel. The concept can be extended to MIMO spatial precoding.

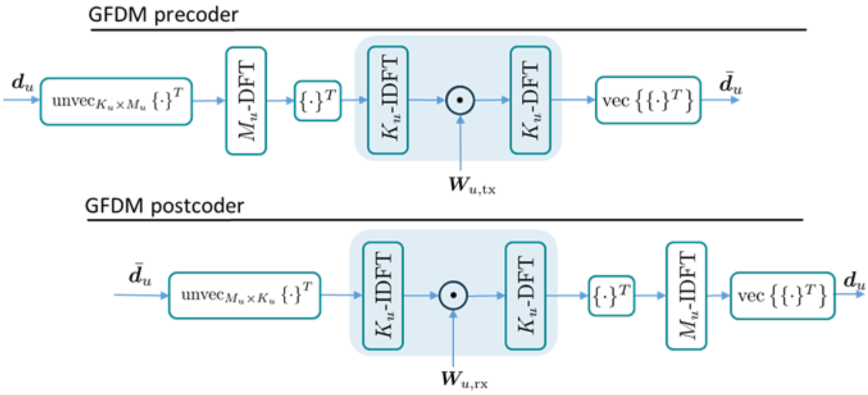


Figure 25: GFDM precoding

Flexible RAN architecture: In iNGENIOUS, we consider shared processing resources and multiple RF frontends, as shown in Figure 26. The baseband PHY/MAC functions are realized by means of software and hardware codesign, where heavy processing is implemented by hardware, whereas software approaches are used for lighter processing.

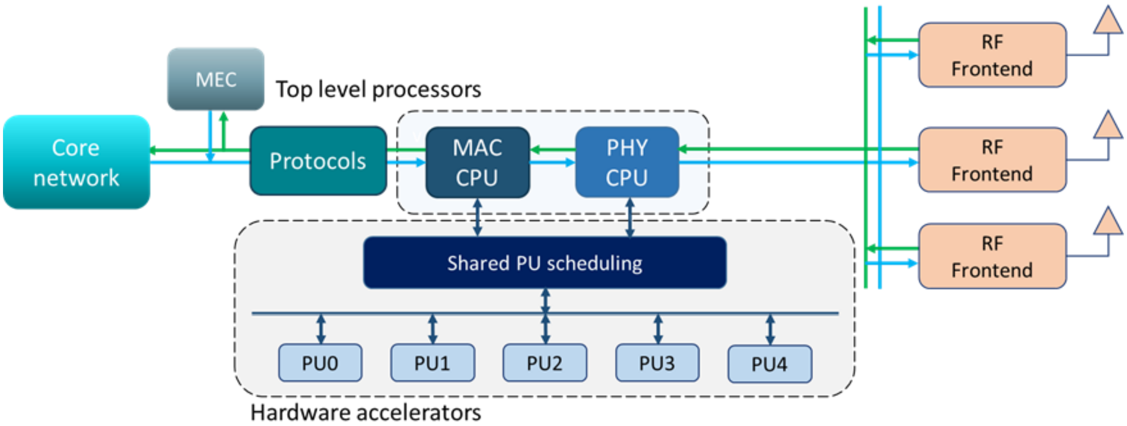


Figure 26: Flexible RAN architecture

This architecture can be used to realize different RAN configurations:

- **Multiple RATs**: to connect heterogeneous IoT devices of standard 3GPP and non-3GPP. In this case, the frontends are tuned to certain frequency and bandwidth, and corresponding PHY/MAC and other layer 2 protocols are realized on the shared processing.
- **Different MIMO configurations**: for spatial multiplexing, diversity, and mu-MIMO. The frontends are configured similarly while the corresponding PHY/MAC implementation should be adjusted.
- **Carrier aggregation**: to exploit multiband availability, each frontend is tuned to the corresponding band, and the PHY/MAC implementation needs to be modified accordingly.
- **Customized RAN**: besides the implementation of standard specifications, customized PHY/MAC design allows private covert

Co-funded by the Horizon 2020 Framework Programme of the European Union

communications, where only devices with the same customized PHY/MAC are able to access to the network.

All these configurations can be exploited to support different application scenarios, especially, in private industrial network. Considering the Factory UC presented in D2.1, the network consists of varieties of IoT devices. The applications require different number of devices with different link requirements. The RAN can be flexibly reconfigured for optimized exploitation of the radio and computation resources to ensure efficient and reliable execution.

2.2 Network Coverage and Deployment Options

2.2.1 URLLC, eMBB and mMTC

The 5G standard for broadband cellular networks aims to succeed the 4G network which provides connectivity to most current cell phones [21]. Like its predecessors 2G-to-4G, the service area of a 5G network is divided into cells. 5G enabled wireless devices to communicate wirelessly to cells that are supported by 5G New Radio (NR) gNodeB (gNB) base stations with huge bandwidth (BW) capabilities; this allows connectivity to the Internet and the telephone network making new internet of things (IoT) applications feasible. Radio waves carry the data of the communications. Due to the virtual future of live broadcasting [22] of immersive experiences with volumetric video (an innovative technology similar to holographic) live streaming with simplified processing and low latency time is desired for the three-dimensional (3D) creation; this would allow an experience of the event, happening right in front of your eyes, in real-time. 5G will enable this broadcasting, overcome some of the limitations of 4G, and guarantee available bandwidth for the provision. It additionally relieves the weight of forthcoming wearable devices as the processing can be moved to remote computers in the 5G network (a cloud).

The central focus of 5G's development has been the integration of different mobile applications under a common framework. With this goal, the standardization community has identified three generic technologies that shall cover all requirements introduced by the diversity envisioned in the future deployment scenarios. These generic technologies are named enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC). The recommendation ITU-R M.2083 0 provides the minimum performance requirements for eMBB, URLLC and mMTC. Briefly, eMBB covers the classical broadband cellular communications where mobile users communicate large data payloads, but with increased data rates and spectral efficiency when compared to previous generations of cellular communications. Under the URLLC umbrella, applications that require high reliability and low latency are to be covered. This high reliability translates into a high success probability when transmitting a certain payload with a predefined time duration. The ITU recommendation states that 99.999% of the time a 32 bytes protocol data unit (PDU) has to be received correctly with maximum latency of 1 ms. Lastly, most IoT applications will be supported via mMTC, which shall support a connection density of 1 million devices per km^2

Co-funded by the Horizon 2020 Framework Programme of the European Union

while keeping an acceptable QoS indicator. Narrow band IoT (NB-IoT) is the current scheme within 3GPP's standards that will support mMTC applications. NB-IoT operates within the former 2G bands, and therefore requires a spectrum license for operation.

Moreover, critical mMTC has been identified as an emerging service category that will handle applications that share requirements between mMTC and URLLC. Exemplary applications include: complete automation of remote area industries such as mining and oil extraction, and in hazardous industries, where human activities are necessary, a massive network of monitoring devices can provide alarms about the quality of air, predict the presence of toxic or explosive gases, and seismic activities. While RAN slicing may enable coexistence of mMTC and URLLC, simultaneous guarantees for both are not achievable under the current 5G standard [23]. As it has been identified in in the work [23] related to enabling critical mMTC, latency around 10 to 30 ms remain the main bottle neck in performance, since the required values are 4 ms for mMTC and eMBB and 1 millisecond for URLLC. Moreover, flavours can be distinguished from the combination of mMTC and URLLC requirements. Recent efforts have been made to extend the capabilities of the 5G framework for supporting these combinations. Although the focus of such efforts has been the physical and link layers, the network and upper layers remain much the same of what has been proposed for eMBB scenarios [23].

In order to address the demands of diverse applications, radio access network (RAN) slicing is presented as the solution to aggregate the aforementioned heterogeneous services under the common 5G framework [24]. In this case, isolated (orthogonal) resources are assigned to different applications based on the corresponding requirements. However, theoretical analysis on non-orthogonal RAN has shown promising -trade-offs on spectral efficiency and keeping the minimum performance requirements for each type of communication service [25].

It is also important to realize that the currently available mMTC enablers, i.e., NB-IoT and LoRa, present limitations that hinder the realization of critical mMTC applications. With respect to the limitations of NB-IoT, experimental investigations show that the long battery life time requirement may be achieved only in specific situations. The study presented in [26] points out that the 10 years battery life requirement is only achieved if a single transmission occurs within 24 hours, which hinders its employment for latency sensitive applications. On the other hand, as NB-IoT is a 3GPP technology, the already available infrastructure in urban and suburban areas is an appealing feature of NB-IoT, since the deployment of LoRa gateways is far less developed around the world.

Achieving the necessary performance to attend URLLC demands still remains an active topic of PHY-MAC design. For addressing the latency requirements, short frames with small code block length have been proposed. However, very high reliability and small code words are opposing design criteria. Therefore, analysing the performance of different channel coding schemes with short code words is relevant for URLLC applications. Furthermore, scheduling schemes that have latency as their main optimization design criteria are still under development. The 5G standard reduces the PHY latency by increasing subcarrier spacing, and hence decreasing the transmission time interval,

Co-funded by the Horizon 2020 Framework Programme of the European Union

while the reliability can be improved via a Modulation Coding Scheme (MCS) table specifically tailored to a target block error rate of 10^{-6}. Moreover, on the MAC level, granting free access for uplink transmissions and latency optimized scheduling policies are solutions to enable URLLC applications [27].

The current 5G NR is able to address the PHY latency, but bottlenecks in uppers layers, such as queuing, processing and access delays, still remain. The classical approach in communications systems design is to optimize the functionality of each layer almost independently of its relations on the following or previous layers. Cross-layer design presents potential to address the bottlenecks [28]. This approach has been employed based on models that capture the interactions among different layers. However, the existing solutions are usually overcomplicated, and do not have feasible real-time implementations. Consequently, the recent advances on machine learning techniques and its integration with model-based knowledge represents an attractive solution to enable the aforementioned services [29]. For instance, the complex relations associated with cross-layer design can be embedded into a neural network (NN) via the off-line training process, which then allow a real-time solution. Nevertheless, NNs still suffer from the long training procedure, and the need to access a large data set for yielding well performing solutions. This issue can be addressed by a two-step strategy. First, a NN model is trained with synthetic data from mathematical models, and secondly, a fine-tuning training step is implemented within real world systems and data. As a result, the time and data required for training based solely on real world data are largely reduced [30].

2.2.2 mmWave Deployment

Currently, the cell coverage is accomplished using the 5G NR allocated spectrum as determined by the 3GPP working group (WG), whose frequency bands are separated into two main frequency ranges, bands from 410MHz to 7125MHz, and those from 24.25GHz to 52.6GHz. Typically, higher frequency radio waves have a shorter coverage (require smaller cells) but provide higher available bandwidth. Sub-6 GHz bands are traditionally used by previous standards like the 4G long term evolution (LTE) and can be conveniently reconfigured for a 5G-like tech approach using a Non-Stand Alone (NSA) configuration supported by a Dynamic Spectrum Sharing (DSS) mechanism. On the other hand, millimetre waves (mmWs) is a new technology enabled by 5G which makes use of unprecedented large bandwith and data rate.

Thus, unlike sub-6 GHz bands, mmWs can provide greater bandwidths and so higher download speeds that could reach up to 10Gbit/s [32] at incredibly low latency time levels. Although bandwidth, speeds and latency are promising at mmWs, we have limitations due to the equipment available nowadays. Therefore, to avoid these equipment limitations, it is desirable to use the lowest mmW 25-39 GHz band frequency Instead of using 410 MHz to 24.25GHz. However, 5G networks operate in up to three frequency bands, low, medium and high. This allows us to provide services with a trade-off between speed versus distance for 5G devices seeking the highest speed gNB within range [33]; this is achieved through beamforming.

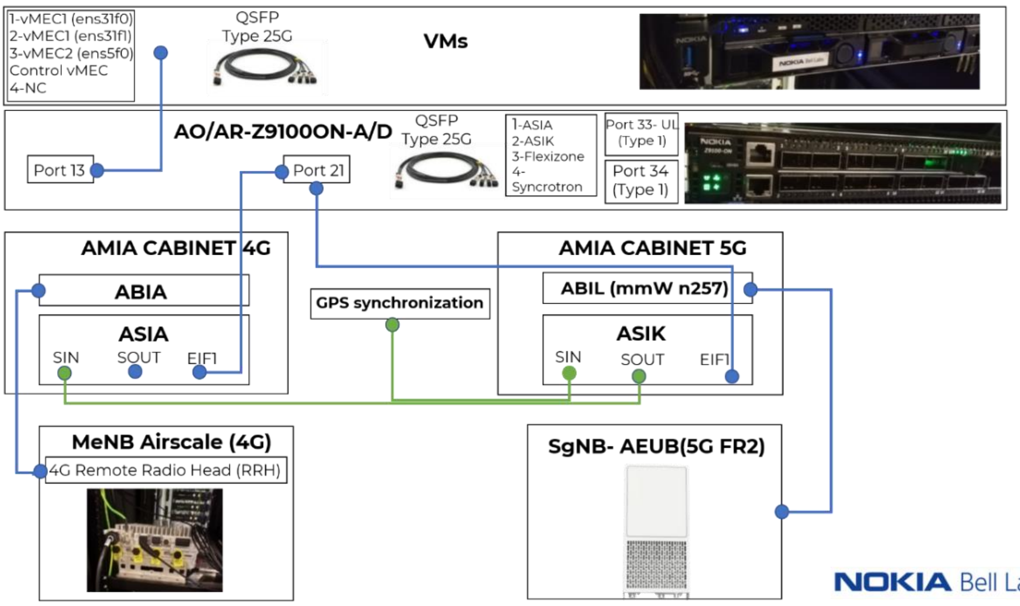
Low-band 5G uses the 600-850MHz Ultra High Frequency (UHF), with expected download speeds of 30-250Mbit/s (slightly higher than that of 4G

Co-funded by the Horizon 2020
Framework Programme of the European Union

and similar macrocell sizes of several kilometres wide). Mid-band 5G uses the 2.5-3.7 GHz microwave (μW) at 100-900Mbit/s (microcell sizes of ≤2km wide) and the high-band 5G uses the 25-39GHz mmW at Gbit/s (picocell sizes of ≤200m wide) [33]. Using high frequencies of the mmW spectrum (20 Gbit/s, as specified in the ITU document IMT-2020) [34], higher download speeds will be possible. But due to the nature of electromagnetism at higher frequencies, increased propagation losses with higher levels of diffraction are expected. Since waves at higher frequencies have difficulties in passing through obstacles, it will be possible to achieve higher download speeds. A trade-off that needs to be accounted for the project goals.

In the AGVs UC, iNGENIOUS plans to employ a private 5G mmW radio coverage in order to enable high data rates. The private mmW gNB will be integrated to the far-edge MEC (cloud-based digital platform) up to either a Standalone / Non-standalone (SA/NSA) 5G Core network for data processing. For cost reasons, the employment of mmW antennas with low cost Is considered. Therefore, experiments will be carried out to measure the coverage the base transceiver station and adapt for the mmW solution. If a single antenna propagation field does not illuminate the desired area in the UC, the use of multiple antennas cannot be seen as an alternative since this solution is very costly and requires heavy engineering works for the wiring with associated additional costs. For this reason, reflector panels are proposed to redirect EM waves towards the desired area, such that coverage can be improved at moderate costs.

The configuration used will be NSA 3X, that is based oh the New Radio - Dual Connectivity (EN-DC), a technology that enables introduction of 5G services and data rates in a 4G network. In addition to the 5G antenna, the structure of radio 5G is composed of ABIL and ASIK modules. ABIL is the base band module and ASIK module is the controller. The frequency for the Radio Remote Head (RRH) 4G antenna will still be defined. The structure of radio 4G is also composed of ABIA and ASIA, which are the base band module and controller, respectively. And we use the Switch A0/AR Z9100ON-A/D for connecting the different radio modules. An schematic diagram of this setup is shown In Figure 27.



Co-funded by the Horizon 2020
Framework Programme of the European Union

Figure 27: Schematic diagram of connections between radio modules and antennas

Initial tests that we have carried out until now registered 2.5 Gbps per modem in the downlink, and 160 Mbps per modem in the uplink. Unit equipment which support dual connectivity can connect simultaneously to LTE Master Node and 5G Secondary Node gNB. The current plan is to employ the RTL6305 Askey device which support the mmWave and Sub 6GHz for 5G NR NSA and SA core network. This device supports several bands, but the band we are interested in is the same as the band of our mmW antennas, i.e., n257.

2.2.3 Wide Area and Private Networks

Among the most common types of area networks are personal area networks (PANs) and local area networks (LANs). LANs are usually privately owned networks. Metropolitan area networks (MANs), which are usually privately or publicly owned, or a combination of both. And Wide Area Networks (WANs), which are typically a combination of private and public ownership, covering geographic areas ranging from 10 m to 1.5 km indoors and 5 to 60 km outdoors.

Unlike LANs, the networking elements that serve the driver's safety use case setting are expected to be more than 1km apart, therefore laid within the MAN type, using a combination of optical fibre and wireless communications, but not necessarily within a WAN because they are not that far apart. Because the gadgets of this UC connect, using a combination of wired and wireless means, to a Personal Computer (PC) located inside the cockpit that would be treated as a private PAN. Outdoors automated guided vehicles (AGVs) communicating to the indoor cockpit (supported by a private network gateway) and to the backhaul form a private LAN prior extending to the embracing MAN.

Easy installation/maintenance is expected for the PAN configuration with moderate difficulty for the MAN due to the location of its networking elements. The driver's safety use case aims to have low latency low propagation delays at high speeds, which is typically addressable in PANs but more moderate for the MAN. The use of a private Passive Optical Local Area Network (POLAN) allows a single optical fibre to connect the cockpit to the backhaul and to serve multiple devices for the hosting of Multi-access Edge Computing (MEC) applications. Unlike PANs, congestions are more susceptible in MANs and more in public than private networks. To achieve leading Key Performance Indicators (KPIs) of low latency at high speeds in this driver's safety use case it is preferable using a private network and its respective Standalone (SA) 5G core network. For the adoption, a permission must be granted by the current publicly owned network provider.

Since the driver's safety use case relies on a combined network ownership and uses User Datagram Protocol (UDP) over Internet Protocol (IP), several IPs will need to be reserved for the data transfer over private hosts and the sent packets not routable on the public internet and ignored by all public routers. For these private hosts to communicate via the public network, a network address translation at a routing gateway must be accomplished. This is

Co-funded by the Horizon 2020 Framework Programme of the European Union

performed using a bridge, across the internet, between the networks via a Virtual Private Network (VPN) or IP Tunnel that encapsulates and encrypts the transmitting packets for a secure delivery; this enhances the driver's safety use case network security generally.

2.3 Satellite Networks

In the iNGENIOUS project, satellite backhaul will be used to enable real time/periodic monitoring of predetermined parameters (temperature, humidity, accelerometer, etc.) of shipping containers when they are sailing on the sea, as the satellite networks provide the only means of ubiquitous connectivity. Cargo vessels with international routes far beyond the coverage of terrestrial cellular networks, can use satellite networks for connectivity allowing shipment information to be ubiquitously available across all connected platforms and interested parties in real-time. Data analytics on this rich and timely data would further allow supply chain players to achieve operational excellence, major reductions in operational uncertainties, and significantly increased revenues.

However, the installation of a satellite terminal on a ship is a very complicated process and needs the authorization of the owner of the ship and the captain. The physical installation may not take a lot of time, but a huge amount of considerations (safety regulations, site survey to decide where to put the satellite terminal, opening holes and passing cables in the ship, etc.) should be taken into account in order to get the approval. For this reason, it is considered as alternative that the satellite terminal will be installed on the port of Valencia.

In this case, the satellite architecture to be used in the iNGENIOUS project is shown in Figure 28.

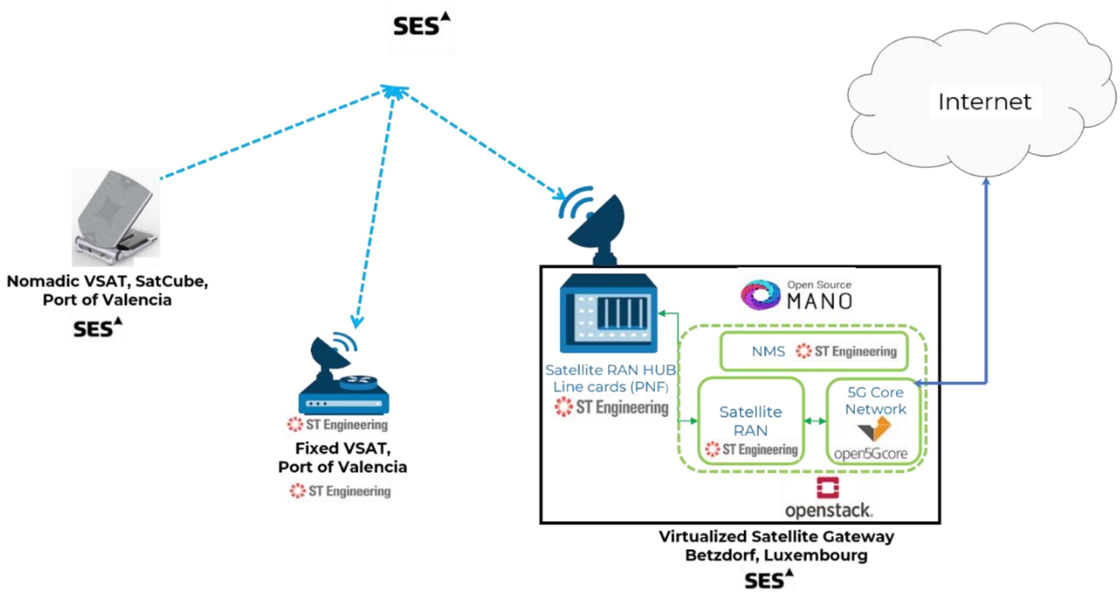


Figure 28: Satellite architecture in the INGENIOUS project

Co-funded by the Horizon 2020
Framework Programme of the European Union

2.3.1 Space Segment

In particular, the space segment builds upon the existing SES's global multi-orbit (geostationary (GEO)/medium earth orbit (MEO)) and multi-band transparent (bent-pipe) satellite fleet with over 70 commercial telecom in-orbit satellites covering 99% of the globe and world population.

2.3.2 Satellite Hub

The baseline satellite ground segment corresponds to the ST Engineering iDirect's VelocityTM Intelligent Gateway (IGW) system residing at the SES teleport site in Betzdorf, Luxembourg introducing a standard 3GPP Core Network (SatCore) to the satellite hub platform. Functions of the existing satellite network are offloaded, allowing SatCore to operate and manage network like a standard terrestrial 3GPP network. The existing satellite network is modified to comply standard radio access network (RAN), referred to as Satellite RAN (SatRAN), node to the SatCore. In addition, the remote satellite terminal is modified to present itself as a standard user equipment (UE) to the network, which allows it to connect to a SatCore in order to access network services.

In the specific implementation illustrated in Figure 28, the 5G Core (5GC) solution Open5GCore [35] of Fraunhofer FOKUS has been integrated within the SatCore and it is used for the management of the Satellite Network. Thus, the existing satellite network has been modified to comply with the standard 3GPP Release 15 compliant 5GC Network architecture. That is, SatRAN has been modified and Ni interfaces to SatCore have been aligned to the 5GC. As such, SatRAN presents itself as a standard 5G RAN whereas the remote satellite terminal presents itself as a standard 5G UE to the network. This enables support for 3GPP services, such as authentication, billing, policy and charging control. This also enables the management of the satellite network slice by the Mobile Network Operator (MNO) in a seamless way, as if the MNO manages a standard 3GPP mobile network. As such, roaming between satellite and terrestrial mobile networks is enabled by such architecture.

In fact, the SatCore may be operated by the MNO or the Satellite Network Operator (SNO), depending on the business model, and there may even be separate 3GPP core networks for the MNO and SNO networks. Initially, the existing non-3GPP based physical layer (e.g., DVB-S2x/RCS2 based) is used over the satellite, which allows faster satellite integration into 5G. However, work is currently ongoing in 3GPP to investigate support for 5G New Radio (NR) air-interface over satellite. This would allow the SatRAN function to be fully 3GPP compliant.

The satellite network functions are virtualised by transferring their execution environment from a dedicated server to a Virtual Machine (VM) using the OpenStack Pike VIM [36]. Satellite VNFs include the SatRAN software element, the satellite 3GPP Core Network function (SatCore), the satellite Network Management System (NMS) as well as additional auxiliary VNFs deployed on the same system using the OpenStack VIM.

2.3.3 Satellite Terminal

Co-funded by the Horizon 2020 Framework Programme of the European Union

Furthermore, various satellite terminals can be used:

- **Fixed VSAT:** the remote very small aperture terminal (VSAT) terminal is fixed/stationary, located at the port of Valencia which corresponds to a 1.2-m Ku-band VSAT terminal.
- **Nomadic VSAT:** the remote VSAT terminal corresponds to the SatCube Ku-band small-factor transportable terminal, which will be also deployed in the port of Valencia, but It can be easily transferred at any place.

The SatCube Ku-band small-factor transportable terminal [37] corresponds to a light weight, compact, portable satellite terminal that enables broadband connectivity almost anywhere on earth. It is a compact, user-friendly device that delivers quick connectivity empowering people at work to communicate and deliver critical services. The highly intuitive user interface makes its operation simple. SatCube's technology is specifically developed to tap High Throughput Satellites (HTS) to fulfil high capacity communication needs, targeting industry verticals and workplaces requiring broadband connectivity where limited or no cellular connectivity is available. The SatCube Ku terminal includes a specially designed battery pack providing 3-hr transmit time; additional batteries are available and are hot-swappable, allowing continuous operation. The SatCube Ku terminal includes an embedded iDirect iQ200 modem, Wi-Fi LAN, flat antenna, amplifier (high efficiency GaN SSPA), positioning system, upconverter, downconverter, heat pipe cooling system, and all necessary user interfaces such as graphical interfaces; all-in-one compact terminal, weighing just 8 kg and the size of a large laptop.

Figure 29: SatCube Ku-band small-factor lightweight transportable terminal with embedded ST Engineering iDirect's iQ200 modem

2.3.4 Satellite Backhaul

In terms of satellite communication, iNGENIOUS will focus primarily on the satellite backhaul solutions. A satellite backhaul connectivity deployment includes an edge node and a central node connected using a satellite backhaul link. The UEs (IoT devices) connect to the edge node which connects to the central node through a GEO backhaul link. The backhaul is seen as a transport layer for the messages between the edge and the central node. Because of this, the backhaul should be as transparent as possible, while at the same time being able to assure a guaranteed communication quality.

Co-funded by the Horizon 2020 Framework Programme of the European Union

With other words, the heterogeneous IoT devices send regular status update and a Smart IoT GW gathers and processes the data and the connectivity with the IoT cloud/Data centre is obtained through satellite backhaul as shown in Figure 30.

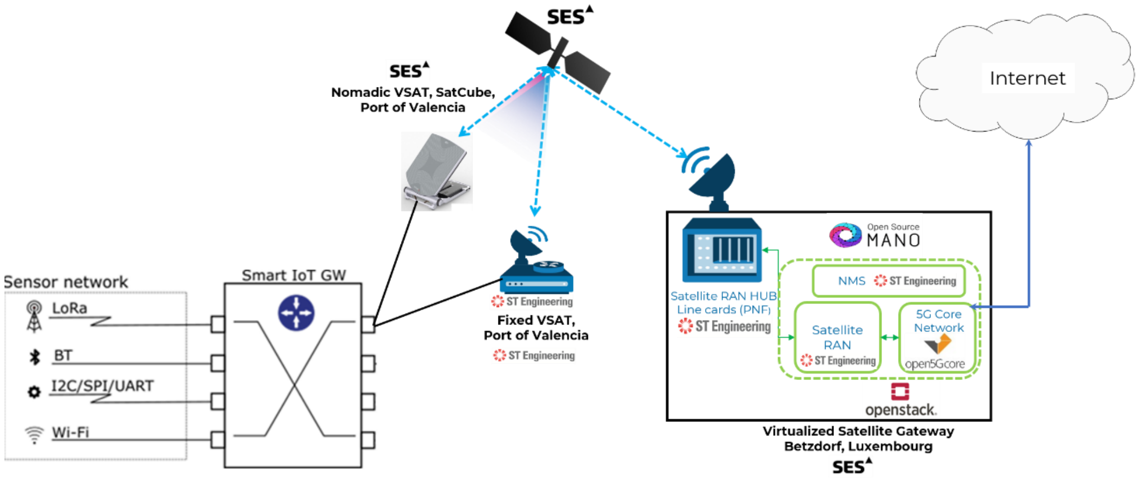


Figure 30: Satellite backhaul connectivity architecture

2.3.5 Smart IoT Gateway

The Smart IoT Gateway (GW) is the system element responsible for the appropriate routing and sorting of sensor data, coming from one or more sensor networks to higher layer data consolidation services and machine-to-machine (M2M) platforms as presented in Figure 31. For performing these operations, the Smart IoT GW is able to interconnect multiple physical interfaces, extracting and transforming messages as data traverses from one side to the other.

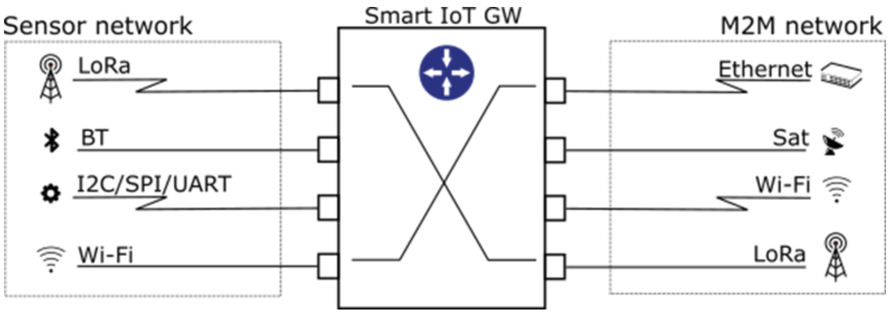


Figure 31: Smart IoT GW physical interfacing

Taking the Open Systems Interconnection (OSI) model as a reference, the Smart IoT GW will expose several physical and data-link interfaces to receive sensor data. Sensors can send messages to the Smart IoT GW either wirelessly (with technologies such as IEEE 802.11, LoRa or Sigfox), or directly connected to the device (via Ethernet, I2C or SPI). The Smart IoT GW will be smart enough to manage the routing and direct the received messages to the right output interface in the right timing. Several factors will be taken in consideration in this operation:

Co-funded by the Horizon 2020 Framework Programme of the European Union

- **Context**: such as the current geographical localization of the Smart IoT GW or its situation relative to potential recipients of messages;
- **Message prioritization**: due to urgent messages that need to be forwarded immediately over other messages that can be grouped together for channel usage optimization;
- **Channel availability**: like in cases where constrained communications impose a specific interface linked to a channel, such satellite link in situations the satellite networks provide the only means of ubiquitous connectivity.

Physical interfaces are added to the Smart IoT GW as plug-in modules, that allows to abstract most of the device functionality from the number and type of the interfaces installed in the Smart IoT GW.

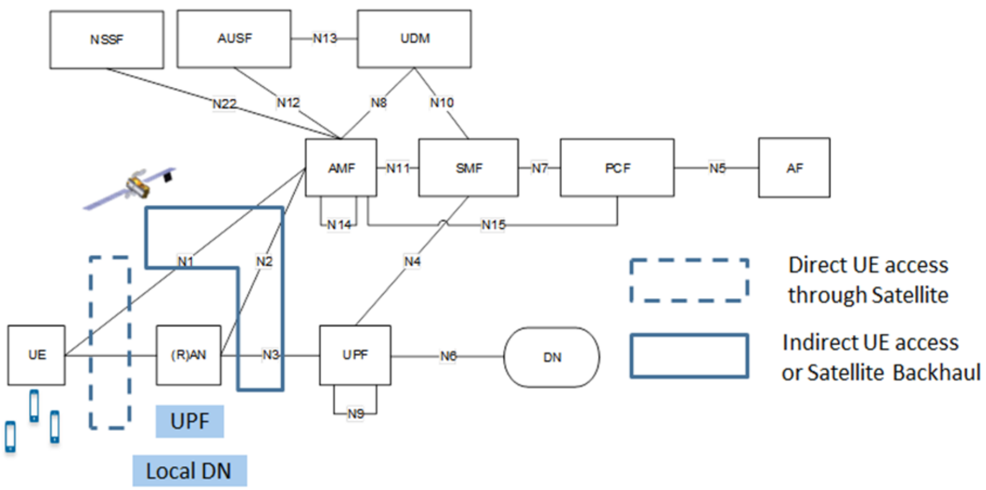
2.3.6 Satellite Integration into 5G

Finally, the iNGENIOUS project may evaluate the scenario where the IoT devices send messages wirelessly through Cellular IoT (5G) which reach to the 5G Core Network via satellite link.

In this context, the outputs from the EU H2020 5GPPP Phase 2 5G-VINNI [38], EU H2020 5GPPP Phase 2 project SaT5G [39] and the ESA ARTES project SATis5 [40], as well as the works in Timeola et al [41] and Corici et al [42], aim at fostering seamless integration of satellite usage within terrestrial 5G networks. With this way, they will enable high-value attractive solutions for Satellite communications and terrestrial actors.

The positioning of the satellite link in 5G system architecture, as defined in the 3GPP TS 23.501 [43], is depicted in Figure 32 including two main possibilities:

- **Direct access:** satellite-capable UE has a direct access to the 5G network through a satellite link;
- **Indirect access or backhaul:** UE connects to (R)AN via 3GPP or non-3GPP access technologies. (R)AN is connected to the 5G Core through a satellite link.



Co-funded by the Horizon 2020 Framework Programme of the European Union

Figure 32: Satellite positioning into 5G System Architecture (Source: SaT5G [39])

For each integration type, several "extensions" can be derived, depending on specific implementation options. Figure 33 provides a taxonomy of the various implementation options as well as support of some additional features (multi-access edge computing (MEC) and Multilink for instance) identified by the EU H2020 5GPPP Phase 2 project SaT5G [39], [41] project in terms of satellite integration into 5G.

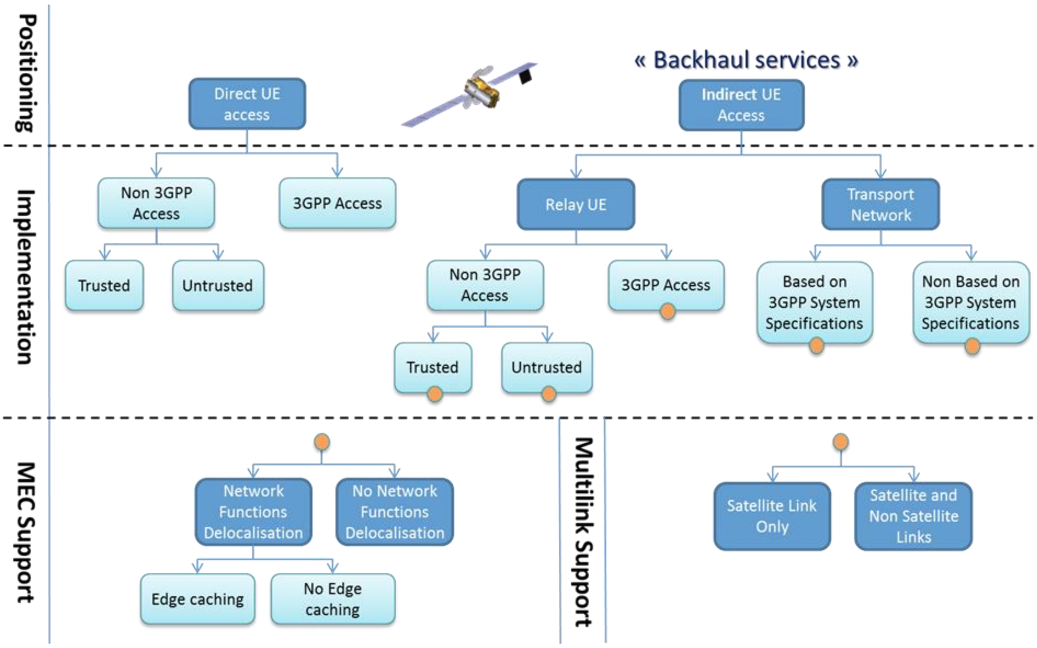


Figure 33: Taxonomy of Various Implementation Options for Satellite Integration into 5G System (Source: SaT5G [39])

Management and orchestration (MANO) is of paramount importance for achieving a seamless integration, and support of Network Slicing, is a "must have" feature introduced for 5G. Two main approaches can be envisaged in this domain – either a single orchestrator for both satellite access network and terrestrial network or, as a more likely option, specific orchestrators are developed for each segment, and appropriate interfaces need to be defined to enable overall orchestration. In the case of the single orchestrator, the Mobile Network Operator (MNO) most likely will consider the Satellite Network Operator (SNO) network as one of the possible transport networks that it may use and it will see the transport network operator (in this case the SNO) as a provider of connectivity. It is therefore most likely that the MNO management and orchestration network will consider the SNO management and orchestration network as hierarchically lower, i.e. the management of the MNO will give orders to the management of the SNO and expect to get responses (e.g. in the form of connectivity) and reporting (e.g. on the status of the connectivity) from the SNO. In the case of specific orchestrators for the MNO and SNO network, the MNO and SNO network will have to manage its own access and core network separately and it may need to cooperate in some cases.

Satellite communications integration into 5G in the iNGENIOUS project will focus on the indirect access or backhaul scheme described in ETSI TR 103 611 [44] and presented in Figure 34, whereby the UE accesses to RAN via 3GPP or

Co-funded by the Horizon 2020 Framework Programme of the European Union

non-3GPP access technologies, and the RAN is connected to the 5G Core through a satellite link.

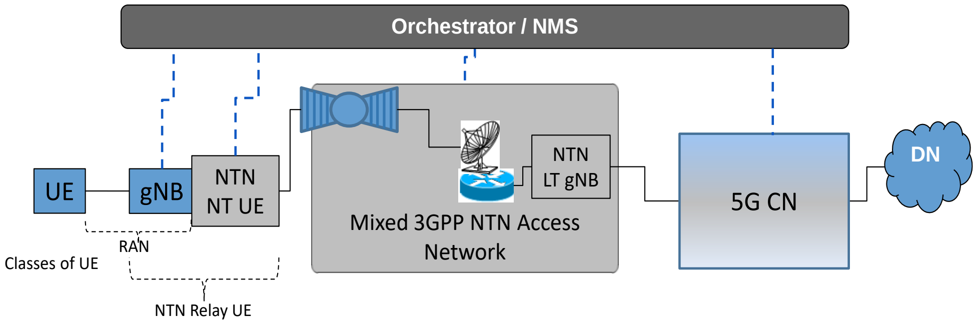


Figure 34: Indirect Mixed 3GPP Non-Terrestrial Network (NTN) Access with Bent-Pipe Payload [44]

2.3.7 Satellite Direct Access

For the iNGENIOUS project, satellite direct access will be investigated as a complementary method for IoT devices to connect to the IoT cloud/Data centre, where it may prove possible to deliver IoT content in a more efficient and cost-effective manner by utilising a Direct-to-satellite approach rather than through traditional backhauling means.

Currently, pre-existing narrow-band, and some broadband satellite services are being marketed as satellite direct access IoT services. Relatively new entrants to this market have identified the gaps in the Direct-To-Satellite IoT market (in terms of message size, service cost, etc.) and are providing alternative products that utilise existing and new satellite constellations. These alternative products use a variety of communication media between the IoT device and the satellite, including both a 3GPP and a non-3GPP based physical layer.

The investigation will include the following elements, and will conclude with a vision of how to implement improvements:

- A detailed analysis of the current state of the art of Direct-To-Satellite IoT.
- Considerations of the link budget requirements of a Direct-To-Satellite IoT solution.
- A description of the components that are required to make up the solution.
- Examination of radio technologies over satellite, including LoRa and CIoT (CIoT includes NB-IoT and LTE-M).
- Applicability of modifications to existing wavefoms, and introduction of new waveforms for IoT over satellite support.
- Grant-free access considerations for Direct-to-Satellite IoT.
- The effect of the different satellite constellation types on a Direct-To-Satellite IoT solution.

Co-funded by the Horizon 2020 Framework Programme of the European Union

2.4 AI/ML for RAN

Cellular networks are becoming more complex in each generation. With the increase in bandwidth of 5G, and the increasing number of highly demanding devices with various quality of service (QoS) requirements, network operators will have to manage their networks with a very high efficiency in order to reduce costs and maintain a high QoE for the users. Intelligence can be useful in different layers and entities of the network to take certain decisions. The O-RAN Alliance has introduced 5 different control loops where ML can be applied for network optimization, as can be seen in Figure 35.

Nº	Control and learning objective	Scale	Input data	Timescale	Architecture
1	Policies, models, slicing	> 1000 devices	Infrastructure-level KPIs	Non real-time > 1s	
2	User Session Management e.g., load balancing, handover	> 1000 devices	CU-level Kpis e.g., number of sessions, PDCP traffic	Non real-time 10-1000ms	
3	Medium Access Management	>100 devices	MAC-level KPIs e.g., PRB utilization, buffering	Near real-time 10-1000ms	
4	Radio Management e.g., resource scheduling, beamforming	~10 devices	MAC/PHY-level KPIs e.g., PRB utilization, channel estimation	Real-time <10ms	
5	Device DL/UL Management e.g., modulation, interference, blockage detection	1 device	I/Q samples	Real-time < 1 ms	

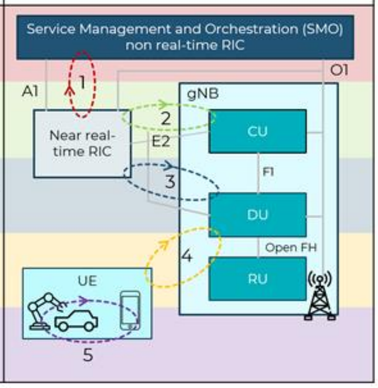


Figure 35: Closed control loops in O-RAN

Each closed control loop runs with different number of users, at different timescales and uses different sources for the input data. This allows to optimize all RAN operations and parameters with a different approach in each control loop. Currently, the O-RAN Alliance is working on 3 of the control loops, introducing xApps in the near RT-RIC which allow to implement ML algorithms which take decisions in a control loop.

- **Non-Real-Time Control Loop**: the non-RT Control Loop operates on a timescale of at least one second. This control loop involves the coordination between the Near RT RIC and the Non-RT Ric through the A1 interface. This control loop can manage the orchestration of resources at the infrastructure level, applying policies and making decisions that impact thousands of devices. An example of this control loop management would be orchestrating and instantiating network slices.

- **Near Real-Time Control Loops**: these control loops operate on a timescale between 10ms and 1s and run between the O-CU and O-DU and the Near-RT RIC. These control loops can make decisions which affect hundreds or thousands of UEs, as one near RT RIC can be associated to multiple gNBs. ML algorithms are implemented as external applications called xApps and are deployed on the near RT RIC to optimize the per-user QoE, controlling handover processes and load balancing, or the beamforming design and resource scheduling.

- **Real-Time Control Loops**: these control loops operate at a sub 10ms or even sub ms timescale and mainly concert interactions between

Co-funded by the Horizon 2020 Framework Programme of the European Union

elements in the O-DU. Control Loops at a similar timescale could also operate between the O-DU and the UE, or even at the UEs, but these cases are not currently covered by O-RAN.

2.5 Relation to UCs and Expected Innovation

This section refers to the relationship of the state-of-the-art Radio Access Network technologies and innovations mentioned above with each use case in iNGENIOUS D4.1. Table 1 below lists the main NG-IoT network slice orchestration functionalities introduced in the previous sections, and for each of them how they map with the various relevant iNGENIOUS UCs in terms of innovation maturity. For this mapping, three options are available: i) state-of-the-art (SoA), meaning that the related functionality is implemented with state-of-the-art technologies, ii) innovation concept, for those innovative functionalities relevant to the UCs but that will not be demonstrated, iii) innovation demonstrated, for those functionalities that will be implemented and showcased as part of the UC. Notice that the mapping provided in Table 1 can only be considered valid at the submission date of this document, and it may evolve in the future according to the evolution of UCs implementation and requirements in terms of network slice provisioning and orchestration. Further details of each use case description can be found in D2.1 [1].

- **5G NR-RAN**: System that interconnects the 5G core and the UEs. (Section 2.1.1)
- **Flexible PHY/MAC**: software defined radio which supports frontend flexibility by providing multiple frequency bands. (Section 2.1.3)
- **AI/ML for RAN**: O-RAN introduces intelligence in the 5G RAN by implementing ML models for decision making. (Section 2.4)
- **LoRa**: physical layer solution created to enable long range communications while keeping low power consumption on devices. (Section 2.1.2)
- **mmW deployment**: 5G NR can use high frequencies in the mmW bands to achieve better performance. (Section 2.2.2)
- **Smart IoT Gateway**: the system element responsible for the appropriate routing and sorting of sensor data, coming from one or more sensor networks to higher layer data consolidation services and machine-to-machine (M2M) platforms. (Section 2.3.5)

Table 1: NG-IoT RAN functionalities mapped to UCs

Functionality	Factory UC	Transport UC	AGVs UC	Ship UC
5G NR-RAN	SoA		SoA	
Flexible PHY/MAC	Innovation concept			
AI/ML for RAN	Innovation concept			

LoRa		SoA		SoA
mmW deployment			Innovation demo	
Satellite Backhaul		SoA		SoA
Smart IoT Gateway				Innovation demo

Co-funded by the Horizon 2020
Framework Programme of the European Union

3 Smart IoT Core Network

The RAN requires the Core Network to deliver device authentication and authorization as well as providing end to end connectivity. Thus, Core Network will interconnect the RAN with application and services consuming and processing the data collected by the end devices. The Core can connect multiple Radio Access Technologies (RAT) to the applications and services. Moreover, the Core Network applies a set of priorities and traffic management to different devices based on profile added by the network operator. Therefore, the Core Network has the ultimate control of the radio access and the devices as well as managing the traffic from end devices towards data networks such public internet, cloud or private local area network.

3.1 5GC with SBA Architecture

The 5GC follows several principles that are mainly targeted for reaching higher flexibility, supporting many different use cases. This includes the introduction of service-based principles, where network functions (NF) provide services to each other. A clean control plane/user plane split allows independent scaling of control plane and user plane functions and supports flexible deployments in terms of where the user plane can run.

The 5GC control plane is based on the SBA. In SBA, the NFs communicate with each other via a logical communication bus and NFs can provide services to each other. An NF instance is registered to a Network Repository Function (NRF). Using the NRF, a NF instance can discover other NF instances providing a certain service. The goal of such architecture is to get a higher flexibility in the overall system, and to make it easier to introduce new services. The SBA enables one core to connect to different RATs which enables easier management of the system and new business process.

In the 5G core, the AMF provides the interfaces towards the RAN, the Session Management Function (SMF) keeps track of the ongoing sessions for a user, and the Unified Data Management (UDM) keeps the subscriber profiles. The UPFs implement the user plane between the RAN and the Data Network (DN) (which can be the Internet, an operator services network, or a 3rd party services network). The Network Slice Selection Function (NSSF) is used to assist slice selection. The Network Exposure Function (NEF) is mainly responsible of exposing capabilities and events of internal NFs to external Application Functions. The Policy Control Function (PCF) governs the network behaviour via policy decisions. The AF (Application Function) provide a way for applications to interact with the 5GC.

Co-funded by the Horizon 2020
Framework Programme of the European Union

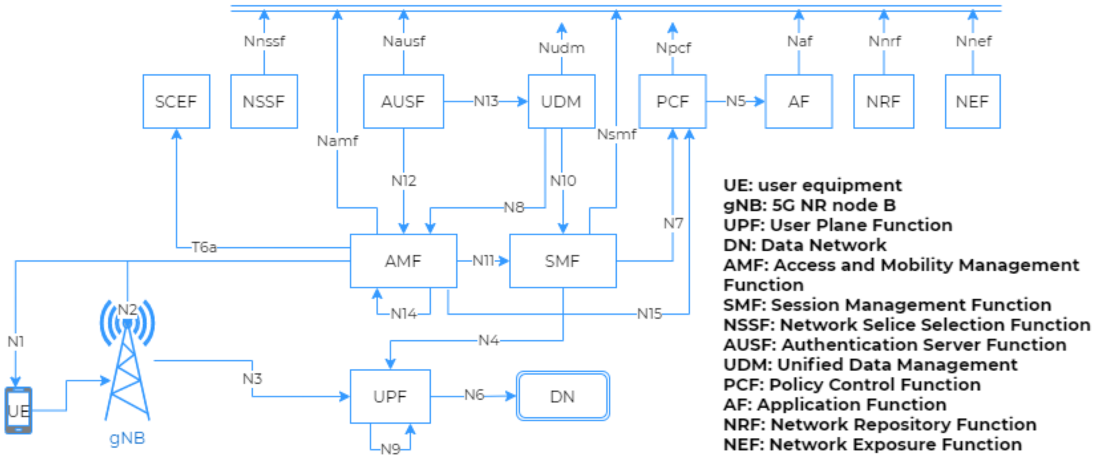


Figure 36: Diagram of Service Based Architecture Arquitecture (SBA)

The 5GC also includes the support for deploying Multi-Access Edge Computing (MEC). Specifically, in those cases where low latency for processing the data from end devices is required. MEC is an evolution of cloud computing that pushes applications from centralized data centres to the network edge near the end-users. MEC is indeed one of the key pillars for meeting the demanding Key Performance Indicators (KPIs) of 5G, especially as far as low latency and bandwidth efficiency are concerned. The 5G system provides a set of new functionalities that serve as enablers for edge computing. These enablers are essential for integrated MEC deployments in 5G networks. A brief explanation can be found in clause 5.13 of 3GPP Technical Specifications 23.501:

- The ability of an Application Function to influence UPF (re)selection and traffic routing via the Policy Control Function (PCF) (clause 5.6.7 of 23.501).
- The Session and Service Continuity (SSC) modes for different UE and application mobility scenarios (clause 5.6.9 of 23.501).

Cloud computing offers storage, computational, and networking facilities within a single or multiple virtualization platforms for enabling different services for mobile networks. Such infrastructure services can be offered by separate service providers. However, cloud computing has shortcomings with regards to emerging applications that require ultra-short latency. It is also limited with regard to computation-intensive applications running on power/CPU-constrained UE. These applications can partially run their computation in the cloud (i.e., other parts of the code are running on the user equipment) while ensuring response times in the range of milliseconds. These limitations are principally due to the centralized cloud computing architecture. The MEC represents a vital solution to these limitations. By pushing the computing resources to the edge servers that are near to users, it allows reducing the delay and enables applications requiring response time in the range of milliseconds. This section provides an overview on MEC integration with 5GC.

Co-funded by the Horizon 2020 Framework Programme of the European Union

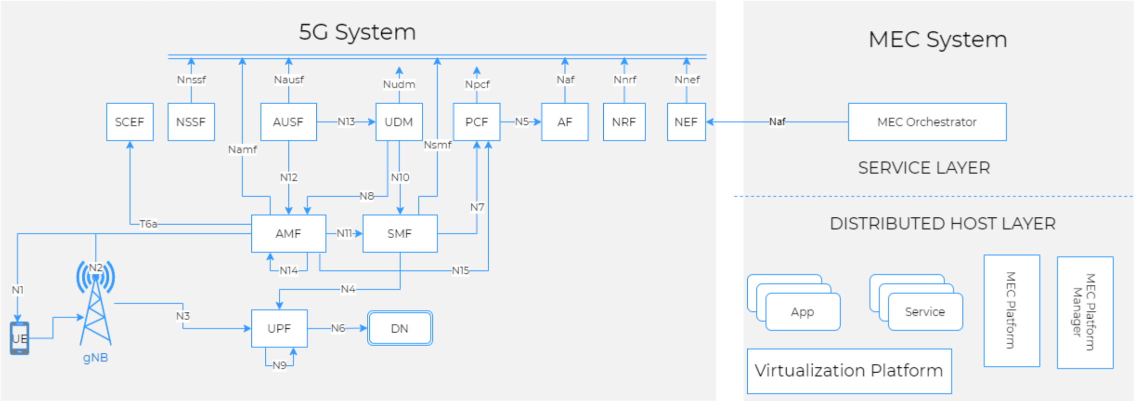


Figure 37: 5GC architecture for Multi-Access Edge Computing (MEC)

The MEC system reference architecture, defined in [45], consists of MEC hosts and functional entities required to run the MEC applications within an operator network. Two levels are distinguished: distributed host level and system level. The distributed host level includes MEC hosts and management entities. Each MEC host contains a virtualization infrastructure, which provides compute, storage and network resources, and a MEC platform. The latter ensures the required functionalities to run a MEC application on a particular virtualized infrastructure (this include discovering and advertising MEC services, instructing data plane according to traffic rules, configuring DNS proxy/server, etc.). The management of the host level consists of the MEC platform manager and the virtualization infrastructure manager. They are responsible of handling the functionalities of a particular MEC host and the application running on it. The system level includes the MEC orchestrator as a core functional entity. The MEC orchestrator has an overview of the complete MEC system. Its functionalities include triggering application instantiation and termination, triggering application relocation, etc.

The figure above shows the 3GPP 5G system architecture proposed for MEC deployment. The MEC system architecture is defined in [46] and [47] clause 5.13. The design approach of 5G architecture 5G system allows mapping MEC onto Application Functions (AF). At the MEC system level, the MEC orchestrator interacts with the NEF of the 5G system through the Nef interface (interface exhibited by NET to the AF). At the MEC host level, the MEC platform is the entity that interacts with 5G NFs and takes care of controlling the traffic steering to the MEC applications. The host level functional entities are deployed in a Data Network (DN) that could be external to the 5G system.

Another feature introduced as part of 5GC consists of network slicing. A network slice is a logical network that provides specific network capabilities and network characteristics that can be dynamically created. A given UE may give access to multiple slices over the same Access Network (e.g. over the same radio interface). Each slice may serve a particular service type with agreed upon Service-level Agreement (SLA). A Network Slice is defined within a Public Land Mobile Network (PLMN) and includes the Core Network Control Plane and User Plane NFs as well as the 5G Access Network (AN).

The 5GC can allocate network slice instance for different devices. A network slice instance consists of a set of NF instances and the required resources (e.g.

Co-funded by the Horizon 2020 Framework Programme of the European Union

compute, storage and networking resources) which form a deployed Network Slice.

The network slicing has been defined part of 5G architecture in 3GPP [48]. A network slice is considered a logical end-to-end network that can be allocated to several UEs and it can be dynamically created and modified. A Network Slice is defined within a PLMN and shall include the Core Network Control Plane and User Plane NFs and the NG RAN. A UE may access multiple slices that are linked to a PLMN where each UE is registered. The slices are associated to given Service-level Agreement (SLA) based on bit rate, latency and packet loss.

Each slice is identified by Single Network Slice Selection Assistance Information (S-NSSAI). 3GPP has defined eight (8) S-NSSAIs in the NSSAI which is the group of S-NSSAI that is sent between the UE and the network during the registration and signalling procedure. The UE provides the network the NSSAI which then must allocate the required resources at radio, network and mobile core NFs.

The S-NSSAI consists of following elements:

- **A Slice/Service type (SST)**, defined the expected requirements in terms of features and services associated to the network slice.
- **A Slice Differentiator (SD)**, is optional and provides additional information to differentiate each slice amongst multiple slices with the same SST to e.g. isolate traffic to different services into different slices.

In this first release the following basic slice IDs have been identified:

Table 2: Basic Slice IDs

Slice/Service type	SST value	Characteristics
eMBB (enhanced Mobile Broadband)	1	Slice suitable for the handling of 5G enhanced Mobile broadband, but not limited to the general consumer space mobile broadband applications including streaming of High Quality Video, Fast large file transfers etc.
URLLC (ultra- Reliable Low Latency Communications)	2	Slice suitable for the handling of ultra-reliable low latency communications.
MIoT (Massive IoT)	3	Slice suitable for the handling of massive IoT.

Co-funded by the Horizon 2020 Framework Programme of the European Union

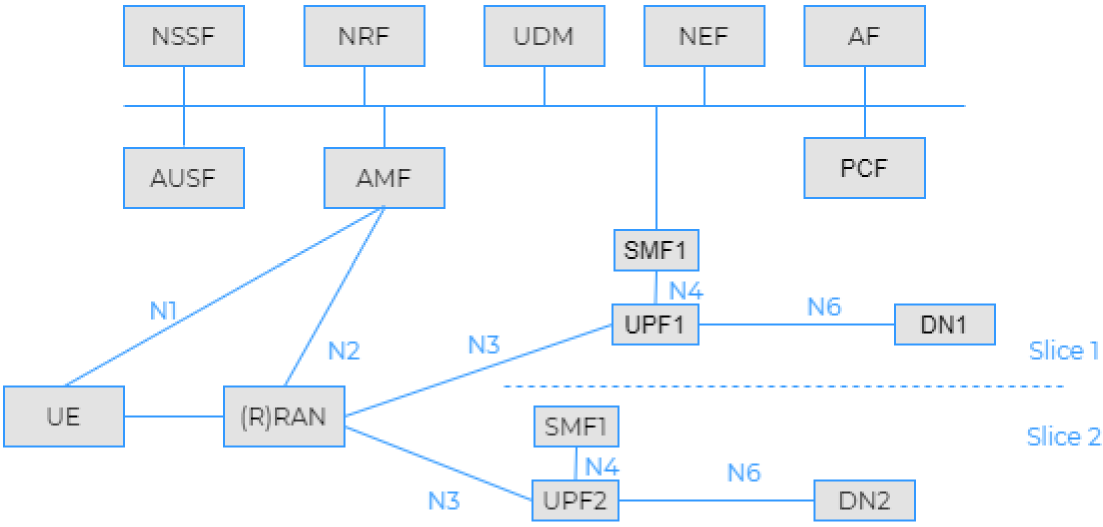


Figure 38: Diagram of Network Slices

3.2 Network Functions for TSN Integration

The TSN functionality included in the 5GC allows to connect mobile with fixed devices as they were physically connected through standard TSN bridges.

TSN is a LAN technology which extends Ethernet to make it more deterministic in order to support time sensitive applications. Its use is typically on the factory floor. It is a complex technology involving many IEEE/IEC standards (more than 10) for things like bounded latency, reliability, time sync and resource management. A TSN bridge separates the control and data plane. However, there are many interactions between the two in order to setup filtering, traffic shaping, scheduling, time sync etc. The control flow is from the bridges to a CNC/CUC. Figure 39 shows the centralized model of TSN management.

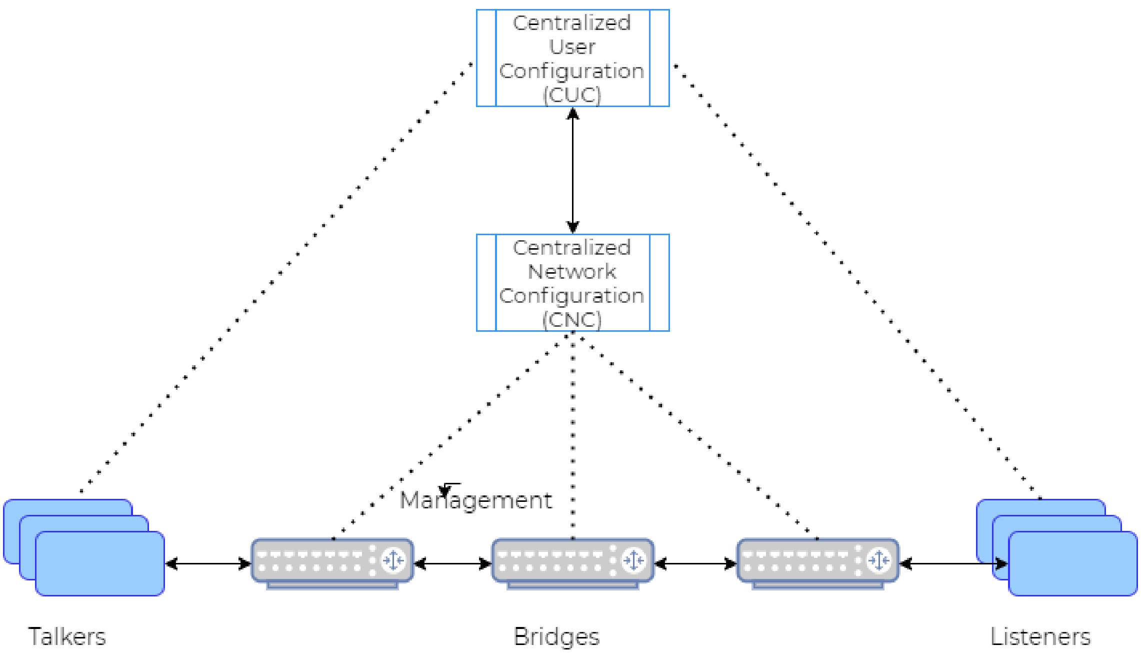


Figure 39: TSN architecture

Various low level (L2) traffic flows through the bridges to the CNC. These include LLDP and ARP for the support of discovery of devices and their capabilities. It also includes neighbour information. These low-level protocols reveal a lot of information on the underlying network which are not secured and hence are not meant to leave the local LAN.

The control room is where centralized control and management functions are located, like centralized PLCs, CNC and automation data collection.

5GS does not support all the 10+ standards of TSN but does supports the features mentioned in Table 3.

Table 3: Features supported by 5GS

Bridge capability	Reference
Bridge ID, Chassis ID, etc.	IEEE 802.1Q
Traffic forwarding information	IEEE 802.1Q clause 8.8.1
Bridge delay, propagation delay related information	IEEE 802.1Qcc clause 12.32.1, 12.32.2
5GS bridge Topology and Neighbor discovery	IEEE 802.1AB (**LLDP**)
Traffic class related information	IEEE 802.1Q clause 12.6.3 and clause 8.6.6.
Bridge enhancements for support of scheduled traffic	IEEE 802.1Q clause 8.6.8.4, 12.9, Annex Q.2 (Qbv) (optional)
Per-Stream Filtering and Policing (PSFP) information	IEEE 802.1Q clause 8.6.5.1 (Qci)
Time aware system	IEEE 802.1AS

The Integration of mobile devices that require TSN functionality with fixed devices require NFs that will adapt the requirements from TSN into the 5G networks.

With this NF, the 5G system is integrated with the external network as a TSN bridge. This NF provide a "logical" TSN bridge that includes TSN translator functionality for interoperation between TSN system and 5G system both for user plane and control plane. These NF that provides the TSN translator functionality consists of Device-side TSN translator (DS-TT) and Network-side TSN translator (NW-TT).

Co-funded by the Horizon 2020 Framework Programme of the European Union

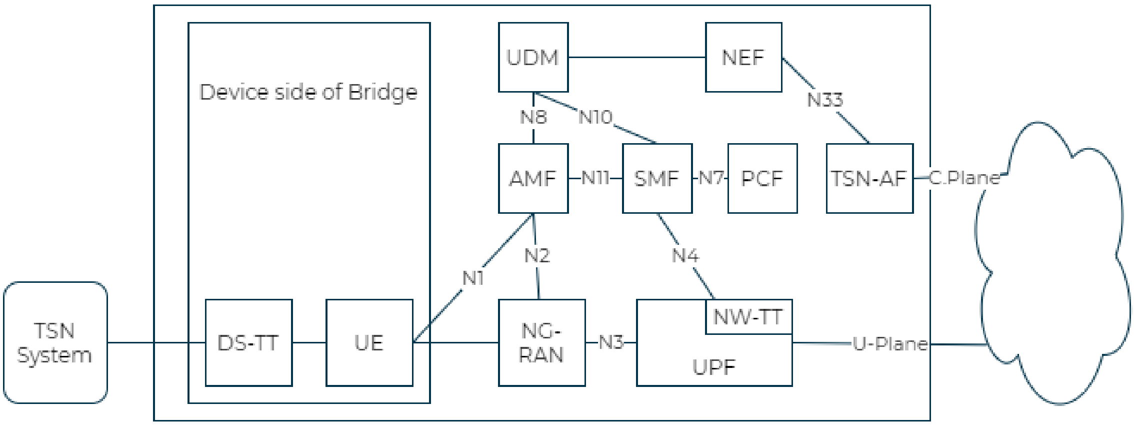


Figure 40: TSN network functions (Ref: TR 22.821 Figure 5.21-2)

The E2E 5G system can be considered as an IEEE 802.1AS "time-aware system". The TSN Translators (TTs) at the edges of the 5G system need to support the IEEE 802.1AS and fulfil all functions (g)PTP support, timestamping, Best Master Clock Algorithm (BMCA), rateRatio. The UE, gNB, UPF, NW-TT and DS- TTs are synchronized with the 5G GM (i.e. the 5G internal system clock).

Two synchronization systems can be independent (5GS synchronization used for NG RAN synchronization and TSN domain synchronization for TSN network follows IEEE 802.1AS). To support TSN synchronization, the 5GS calculates and adds the measured residence time between the TTs into the Correction Field (CF) of the synchronization packets of the TSN domain.

The 5G internal system clock shall be made available to all user plane nodes (e.g. UPF, NW-TT) in the 5G system via the underlying PTP compatible transport network.

The 5G internal system clock shall be made available to UE with signaling of time information related to absolute timing of radio frames as described in TS 38.331. The UE makes the 5G internal system clock available to DS-TT.

The NW-TT makes ingress timestamping (TSi) for the gPTP event messages of all domains. UPF forwards the gPTP message to the UE via user plane using only one PDU session per UE per UPF. The UE receives gPTP messages and forwards them all to the DS-TT that receives the original TSN clock timing information and the corresponding TSi via gPTP messages for one or more working domains.

The DS-TT then makes egress timestamping (TSe) for the gPTP event messages for every external TSN working domain. Ingress and egress time stamping is based on the 5G system clock at NW-TT and DS-TT. The difference between TSi and TSe is considered as the calculated residence time spent within the 5G system for this gPTP message.

The DS-TT modifies the gPTP messages it sends towards an end station to include the calculated residence time. The TT output ports at the UPF side (NW-TT) towards the connected TSN networks propagate the 5G clock using the 802.1AS profile.

Co-funded by the Horizon 2020 Framework Programme of the European Union

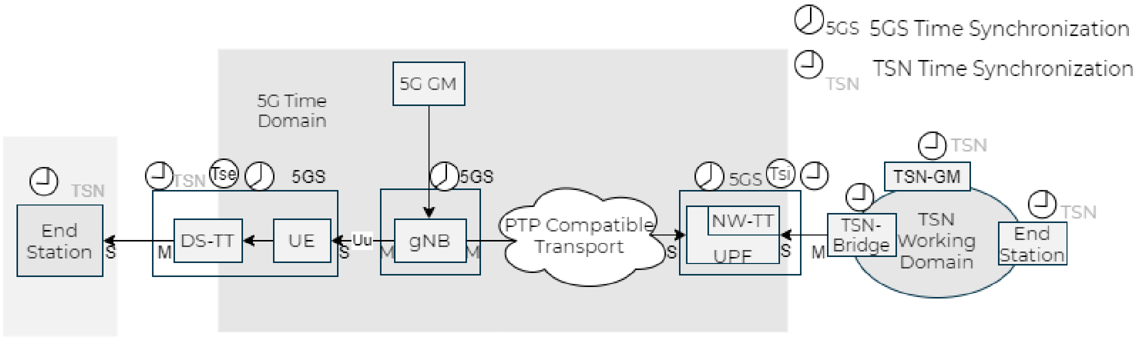


Figure 41: Time synchronization for TSN in 5GS

3.3 Interfaces Between 5GC and SDN

The usage of SDN with the 5GC supports the use cases that require slices to separate traffic with requirements. The SDN allows to route traffic through different parts of the mobile backhaul to ensure complete separation of network resources for different traffic.

The 5G architecture needs to address new requirements to enable network slicing that allow to allocate and separate resources to different devices. 3GPP has defined the logical elements to manage the network slices such as NRF, NSSF, etc. However, 3GPP does not specify how the network slice is physically allocated and managed at the level of the transport networks. The existing networking technologies used in mobile backhaul are suitable for fixed IP networks where fully distributed routing algorithms provide optimal paths based on link costs and react efficiently upon link breaks. IP networking delivers a flat network for best effort traffic management. However, 5G mobile networks aim at new features such as network slicing where the same network provides multiple network overlays each with different traffic requirements.

A network slice in 5G is a set of packet transport links and nodes, set of computing elements and software for the network functions to run a network using the assigned resources. A slice is set up for a particular use case of the 5G network and they can be provisioned in advance to guarantee the QoS requirements for URLLC, mMTC or eMBB communications.

The mobile backhaul networks have fulfilled the traffic requirements based on over-dimensioning and pre-provisioning that ensure enough capacity for best-effort IP based networks. However, pre-provisioning cannot work in 5G networks since the set of assigned resources can be increased and decreased in size based on user needs and policies that change over time. The slices might be created, updated and terminated dynamically based on end-user requirements. The network slices are required to constrain unexpected high peaks of traffic e.g. mMTC under pre-defined set of resources. The slice will contain traffic with lower priority so other critical traffic is not affected and it keeps the resources allocated to their own slice.

In order to integrate the 3GPP network slicing network functions with the physical management of network resources, a Mobile Backhaul Orchestrator (MBO) is defined. The MBO utilize technologies such as SDN on top of basic IP routing to efficiently manage network slices.

Co-funded by the Horizon 2020 Framework Programme of the European Union

3.4 Security Aspects

3.4.1 5G Security State of the Art

Telecommunication networks are evolving rapidly across a broad technological environment which includes virtualization, IoT and Industry 4.0. This is met by an equally broad yet deteriorating cybersecurity environment, giving the rapidly increasing attack surface

Advances in technology, together with the broader development of networks beyond 5G RAN, is having a significant impact on security, such as SDN, network function virtualization (NFV) and edge computing.

In the era of 5G, it’s important that, when we begin to conceptualize security on a system wide level where telecom networks are an important component, we adopt a strong understanding of the following:

- Increased value at stake and decreased risk tolerance
- Cyber-physical dependencies
- Security of standards, products, deployments, operations
- Proactive cybersecurity measures
- Vulnerability management
- Securing the supply chain

From the 5G network point of view, trust in IoT is based on trustworthiness of the device's hardware, software, configuration etc. Hence, trustworthiness is cumulative and will be defined by how well network operators and those who manage IoT devices govern the following:

- Identities and data
- Security and privacy
- Actor compliance with agreed security policies end-to-end

Co-funded by the Horizon 2020 Framework Programme of the European Union

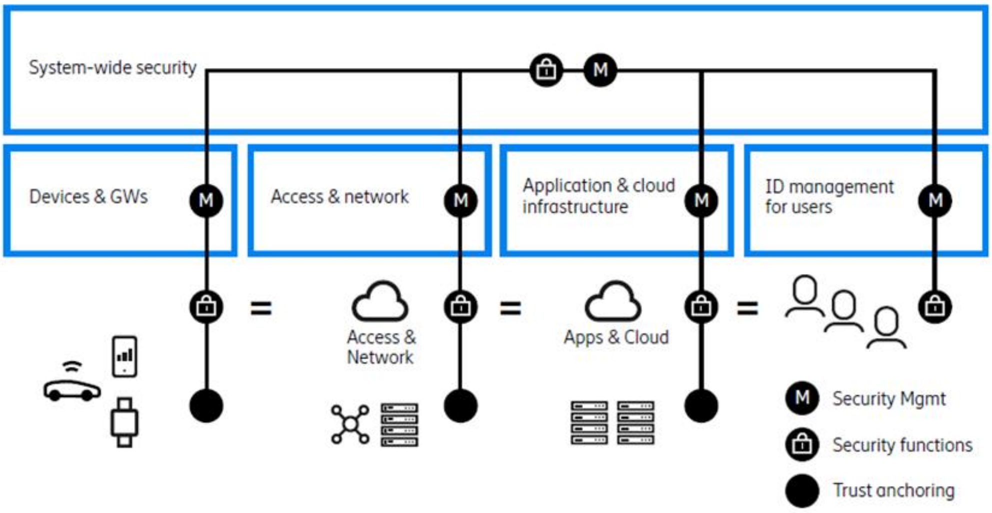


Figure 42: System-wide Security

Interactions between user authentication, traffic encryption, mobility, overload situations, and network resilience aspects need to be considered together to build secure systems.

5G radio network components build further on proven 4G technologies. Those technologies offer today state-of-the art security functions such as support for secure protocols, e.g. TLS and IPsec, on all interfaces, vendor credentials, HW rooted trust anchors for trusted boot, and signed software to ensure that only software provided by trusted and authorized parties can execute on the platforms. Those functions together with others such as access management, logging, and analytic tools constitute a solid foundation for implementing security policies and operating the network securely. Furthermore, specific operational frameworks address operational needs by mandating hardening guidelines and security user guides for all nodes and components.

A fundamental challenge instantiating a VNF is to securely provision it with roots of trust that enable it to become a trustworthy peer in the network that can protect the confidentiality and integrity of data both in transit and at rest. On the other hand, for Physical Network Functions (PNF), i.e. traditional HW/SW deployments, 5G components typically inherit the hardware-rooted security for secure boot and signed software verification established already for 4G/LTE.

5G functionalities are going to provide for many new use cases and more actors interacting with the infrastructure. In this new environment, efficient control of who may interact with whom, and who may do what and where, becomes a central security objective.

To this purpose, specific 5G tools are being developed for efficient and correct policy management, policy distribution, policy verification, and policy enforcement that can enable functionality across tomorrow's networks.

Defence in depth is another very important principle which will become crucial for 5G networks. What cannot be prevented must be detected,

Co-funded by the Horizon 2020 Framework Programme of the European Union

responded to, and recovered from, and telecom networks are uniquely instrumented to monitor performance in general.

These capabilities are being leveraged and augmented, and together with modern analytics technology, drawing upon AI and Machine Learning, specific 5G intrusion detection capabilities are being created.

One specific concern that has received considerable attention is the ability to build a false base station through readily deployable technology and at relatively low cost. In this area, advanced functionalities will be available, which will make it possible to efficiently detect the presence of rogue radio nodes in the networks.

Perhaps one of the most important priorities will be the relationship between usability and serviceability. Security functionality is of little value if it is not used or if it is used in a way that it defeats its own purpose. One very important challenge is going to be making security functionality unobtrusive and the default thing to do.

3.4.2 Policy Analysis and Definition for Identity & Access Management

An analysis of possible security access policies for IoT devices will be carried out. In order to identify such policies, a first step will be to define metrics, temporal ranges, spatial constraints, identities or other possible parameters which can be significant for the purpose of access management and authorization, and to enforce and guarantee access security.

A second step will be to propose suitable sets of rules based on the identified parameters, which will allow to reach security guarantees in line with the needs of the specific application domains of the project.

Then a final step will consist in defining high-level access control policies based on the defined rulesets. Such policies must be enforceable by a specific access control function, where it will be possible to dynamically adjust both the specific policy to enforce, the ruleset to apply and the quantitative parameters of a specific rule. Dynamic rule and parameter adjustments could derive both from automatic feedbacks and from manual intervention and calibration.

3.5 AI/ML for Core Network

The AI/ML integrated with 5GC supports the use cases that require low latency and high availability between radio access and core network where the sensors will deliver data to application servers.

The MBO in addition to SDN integrates Machine Learning (ML) to efficiently manage network slices. The ML techniques are used to estimate the available resources in each link based on different network features and calculations made in the network so that it would be used as an input for the routing algorithms to decide the best route. Those features include link bandwidth usage, end-to-end latency, hop count, packet loss, etc. The MBO uses a centralized SDN based management of resources combined with ML to

Co-funded by the Horizon 2020 Framework Programme of the European Union

effectively allocate network resources based on new requirements for existing or new network slices.

From the different types of machine learning techniques available, we consider the supervised learning to be suitable for managing the network resources. The idea behind supervised learning is that, for some inputs, we want to have certain value as an output. Thus, the supervised ML algorithms run based on the inputs received from the monitoring system until they get output values close enough to the target value which provides optimal usage of the network resources. Therefore, using this technique we evaluate if a given link is congested or not based on the network information such as jitter, bandwidth utilization, packet loss as an input. The ML will trigger some actions to change routing policies in the switches which the SDN controller integrated in the MBO will take care of.

The MBO design includes the modules depicted in Figure 43. i) Network monitoring to check available resources and ii) Machine Learning engine that will apply different traffic engineering to calculate optimal rules to be set in the network switches though SDN controller.

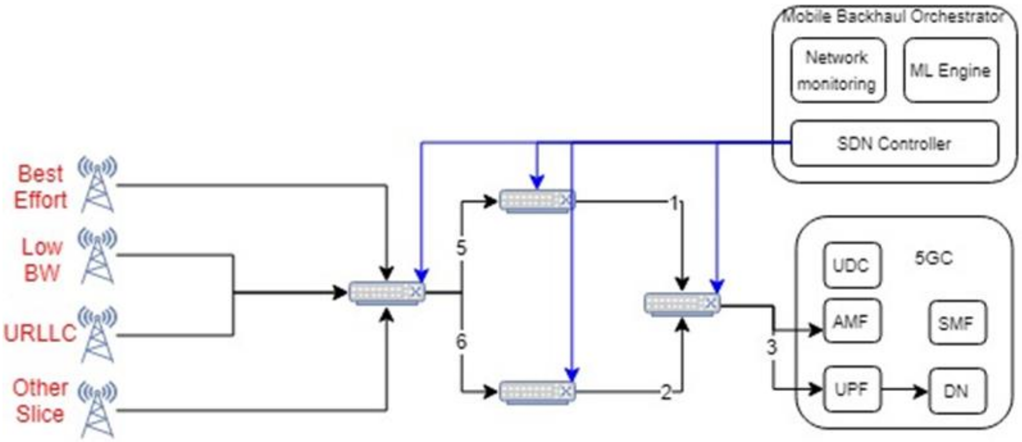


Figure 43: Architecture for MBO

3.6 Relation to UCs and Expected Innovation

This subsection maps the Core Network innovations of this section to the UCs, namely, network slice management in the 5GC and the 5GLAN, TSN functionalities. The specific innovations are given in Table 4, where three categories: i) state-of-the-art, meaning that the related functionality is implemented with state-of-the-art technologies, ii) innovation concept, for those innovative functionalities relevant to the UCs but that will not be demonstrated, iii) innovation demonstrated, for those functionalities that will be implemented and showcased as part of the UC. The mapping provided in Table 4 has to be considered valid at the time of writing, and it may evolve in

Co-funded by the Horizon 2020 Framework Programme of the European Union

the future according to the evolution of UCs implementation and requirements in terms of network slice provisioning and orchestration.

The novel components listed in the table consists of following:

- **Network slice manager**: Network module that creates network slice and Interacts with 5G system to activate the slice and bind It to devices.
- **5GLAN AF**: Application Function to create group of devices to be part of Virtual Network (VN)
- **TSN AF**: Application Function that allows to transfer TSN control messages from fixed network to mobile devices.
- **SDN MBO**: Network module that Interacts with transport network to allocate and manage resources required for the network slices.

In particular, the 5GC includes network slice management as part of high level architecture presented in this section to deliver different slices for various data types. The network slicing implemented within the iNGENIOUS platform will support different UCs that rely on different network resources based on delay and bandwidth requirements. The network slice manager work is at its early design and development stages, at the time of writing some of the functionalities are planned to be fully implemented and demonstrated in the context of the project UCs, while others will be kept at the conceptual level. 5GLAN and TSN integration to 5GC require additional functionality implemented in standard network function like UPF. Moreover, new network functions to create Virtual Networks that integrate mobile devices with fixed industrial devices is required. Currently the design and implementation of those functions are ongoing. Table 4 lists the main 5GLAN and TSN related scenarios that can benefit from this new functionality.

Table 4: Core network functionalities mapped to UCs

Functionality	Factory UC	Transport UC	Port Entrance UC	AGV UC	Ship UC
Network Slice Manager	Innovation demo	Innovation concept	Innovation concept	Innovation concept	SoA
5GLAN AF	Innovation demo		Innovation concept		Innovation concept
TSN AF	Innovation demo				
SDN MBO	Innovation concept		Innovation concept		Innovation concept

Co-funded by the Horizon 2020 Framework Programme of the European Union

4 NG-IoT Network Slice Orchestration

In general, a 5G network infrastructure that provides an end-to-end connectivity service in the form of a network slice to the end users requires proper resource allocation and management. The allocation and the management of these resources becomes critical, especially when the number of UEs starts to increase. To this end, the NG-IoT network slice orchestration supported by an AI/ML platform plays a crucial role, performing semi-automated decisions on resource allocation and management.

In the following sections, a state-of-the-art analysis is performed and a reference architecture is described, Identifying the gaps and thus the possible innovation for NG-IoT Network slice orchestration. Moreover, the requirements, the architecture of the orchestrator, and the interaction with the other components in iNGENIOUS are described and depicted.

4.1 State-of-the-art Reference Architecture

4.1.1 Reference Standards

With 5G, the telecommunication industry is more and more looking at comprehensive MANO solutions to ease the deployment of heterogeneous vertical services and network slices across several technology domains. The concept of network slicing allows to jointly orchestrate resources (network, computing, storage) and network functions (NFs) (virtualized or physical), which are managed and delivered together to instantiate and compose network services over a shared infrastructure. Network slices can be dynamically created and customized according to the requirements of the services that will run on top of them, for example in terms of resource or function isolation and QoS guarantees. This is extremely relevant to iNGENIOUS, where heterogeneous IoT network technologies and devices are required to interoperate with the 5G network to provide smart and innovative supply chain and industrial IoT services.

In this respect, 3GPP has proposed a framework that specifies the functional components required for the management and orchestration of network slices [47], and identifies the three network segments (i.e., radio, access, core) that should be covered in end-to-end network slice deployment scenarios. Furthermore, the network slicing principles are embracing heterogeneous network technologies, including 5G New Radio (5G NR) access technologies, converged fronthaul and backhaul segments with SDN capabilities, as well as hybrid 4G LTE and 5G Core services. According to 3GPP, network slice and slice subnet instances are isolated logical networks and subnets, respectively. They can be realized as a collection and combination of NFs that in virtualized environments become Virtualized Network Functions (VNFs), which are compliant with the ETSI NFV principles. Indeed, 3GPP and ETSI NFV are pretty much aligned in their views and network slicing principles [50], and the NFV network services (i.e. the interconnection of one or more VNFs into a service topology that provide a stand-alone service) are considered the basic logical elements to implement a network slice. In the recursive 3GPP network slice definition, a network slice is mapped to one or more network slice subnets realized as NFV network services. Therefore, what 3GPP defines as NFs are

Co-funded by the Horizon 2020 Framework Programme of the European Union

modelled as VNFs and can be managed by NFV orchestration tools. Following this approach, the 3GPP network slice management functional entities can operate on top of NFV Orchestrators, building end-to-end network slices through multiple NFV network services requested on-demand and across different network and computing domains. This approach can also leverage also on recent evolution of NFV architecture principles towards management and orchestration cloud-native NFs (thus beyond traditional VNFs implemented as Virtual Machines).

4.1.2 Existing Solutions for Slice Orchestration

Several phase-2 and phase-3 5G-PPP projects, like 5G-TRANSFORMER [51], SLICENET [52], 5G-EVE [53] and 5GROWTH [54] among others, propose vertical services and network slicing solutions based on “network slice management” functional entities. NFV Orchestrators are responsible for building end-to-end network slices through multiple NFV network services requested on-demand and across different domains. Vertical services are mapped to network slices according to service constraints and QoS requirements, which are in turn instantiated as a combination of NFV network services managed by the underlying NFV Orchestrators. However, the adoption of such solutions still does not provide truly end-to-end service orchestration as it is mostly limited to access and core network segments.

4.1.3 Gap Analysis

Current MANO framework solutions and existing tools for network slicing and NFV network service orchestration are still implemented as silo-based control and orchestration tools, mostly addressing the coordination of monolithic pipelined services that cannot be easily and transparently adapted to changing network and service conditions. Lack of agility and flexibility in the service and slice lifecycle management is still an evident limitation, thus requiring ad-hoc solutions and customizations for addressing the challenging NG-IoT time sensitive networking and ultra-low latency requirements. Moreover, a full integration of 5G NR, NG-IoT and edge computing technology domains is not yet achieved when it comes to deploying end-to-end network slices. Moreover, the overall capability of such orchestration approaches to fulfil heterogeneous service constraints and requirements is still need to be proved, as it often requires per-service customizations and human-driven adjustments to support end-to-end deployments. In addition, the adoption of AI/ML technologies for cognition-based optimizations, including their interaction across the different technological domains (e.g. network related, edge computing related, cloud computing related) and their tight integration with the service and slice lifecycle management is still at its early stages.

The current MANO coordination functionalities are highly linked to static internal coordination and orchestration logic. The management operations at different levels follow the workflows which MANO is responsible for implementing. This results in lack of flexibility because when either minor adjustments are needed or unplanned events occur, MANO remains strict to its static coordination and orchestration logic. In this context, a tight

Co-funded by the Horizon 2020 Framework Programme of the European Union

integration with AI/ML techniques could address this kind of problem. AI/ML algorithms generally do not follow the if-then approach, but are able to “learn” from past experience and, in some cases, take decisions. To this end, the Experiential Networked Intelligence Industry Specification Group (ENI ISG) is defining a Cognitive Network Management architecture, using Artificial Intelligence (AI) techniques and context-aware policies to assist network management decision-making through closed-loop AI-driven mechanisms. An ENI system can be used to enhance and optimize the network slice management and control operations, leveraging active and passive monitoring services and network status information for anomaly detection (fault, error and unusual behaviour), prevention and fast recovery. These policies are then used to determine when, where and how to place, or adjust, the network slice instance and network functions. This information is used by the to indicate to the network slice management entity when, where and how to place or adjust the network slice instance and network functions (e.g. MANO), as part of the slice and network function lifecycle management.

For these aforementioned described reasons, in the iNGENIOUS project, one key innovation for what concerns the orchestration aspects is the intelligent management and orchestration of network resources for the NG-IoT ecosystem. In this context, since the resource demand could be fluctuating during a time period and at the same time the high-level requirements must be satisfied, a semi-automated decision-based approach comes into place.

In literature, different AI/ML approaches have been extensively investigated for the network slice configuration, optimization, and reconfiguration. In particular, some works focus on where different supervised and unsupervised machine learning approaches and techniques could be applied at different stages of network slice lifecycle [55]. However, most of these techniques use either simulated [56][57] or specific sets [58] of data to assess the accuracy of such AI/ML approaches and very few of them mention the use of network functions and application data together for the data collection.

To address these limitations, the twofold usage of data related to the network functions coming directly from the 5G system and the IoT application data (from the DVL) represents another key innovation in iNGENIOUS. Indeed, using the two data sources together, a robust use-case-oriented AI/ML algorithm could be designed and built to maximise and optimize the supply chain and industrial IoT slices.

4.2 Orchestration Requirements

iNGENIOUS aims to design and implement a comprehensive MANO framework to meeting the requirements described in the previous section. The goal is to enable the management and orchestration of dynamic and flexible network slices capable of supporting the heterogeneous requirements posed by supply chain and industrial IoT services in terms of network performance, time sensitivity, reliability and responsiveness. This has to take into account a tight integration and management over the 5G RAN, edge and core network domains to guarantee seamless end-to-end communications for interconnecting different IoT devices.

Co-funded by the Horizon 2020
Framework Programme of the European Union

Therefore, the iNGENIOUS NG-IoT network slice orchestration, integrated with AI/ML techniques, represents a crucial entity for the management, coordination, and optimization of the network and computational resources available in the 5G network infrastructure. In practice, the network slice orchestration shall support the following functionalities:

- **Slice modelling**: it should support network slice and related network service and resource requirement descriptions, following the 3GPP and ETSI NFV standard information models for Network Slice Templates (NSTs) and Network Service Descriptors (NSDs), as well as for the VNFs (i.e. VNFDs). This allows to model service and resource requirements for the specific slice.

- **Slice onboarding**: it uploads specific per-slice capabilities and requirements in the form of descriptors in the orchestration platform. This drives the lifecycle management of slice instances and network functions, and shall be supported by internal catalogues (that can be exposed for future information retrieval). Two types of slices that are foreseen to be onboarded in iNGENIOUS: industrial IoT and supply chain slices.

- **Automated lifecycle management of network slices supported by AI/ML**: the orchestration platform shall provide interfaces and internal mechanisms to manage the lifecycle of an end-to-end network slice in an automatic way through the ML algorithms that support the decision-making.

- **Interaction with the network and computing infrastructure**: it translates the lifecycle management mechanisms into provisioning actions and configurations to be performed on such infrastructures. The network and computing infrastructures to be managed depend on the specific use case scenario, and can include edge and core computing locations, NG-IoT RAN (i.e. offering different options of non-3GPP, 3GPP and flexible RAN as per section 2.1), SDN-enabled backhaul transport network segments (e.g. in SD-WAN scenarios), considering satellite backhaul links when available.

- **Deployment and orchestration of edge/MEC applications**: it satisfies the ultra-low latency requirements, when required as part of the given slice template by the supply chain and industrial IoT applications.

- **Data collection**: in terms of monitoring the status of network slices and related network functions (i.e. VNFs), as well as of the network and computing resources involved. In addition, data can be collected from the iNGENIOUS DVL as well to augment the network slice monitoring data with NG-IoT application level data. This allows the network operator to use historical data of network slice status, behaviour, resource consumption and utilization combined with NG-IoT application information to maintain the required level of service.

- **AI/ML assisted network slice optimization**: It proactively optimizes, adjusts and adapts the network slices at runtime based on the heterogeneous data collected from the available data sources (as stated above).

Co-funded by the Horizon 2020 Framework Programme of the European Union

4.3 Orchestration Scenarios

4.3.1 Integration with Edge Computing

Edge computing is a distributed computing paradigm whose main aim is to reduce as much as possible the response time of the deployed applications, and guarantee a seamless end-to-end service. For this, it is essential that the iNGENIOUS NG-IoT network slice orchestration framework has a tight integration with edge/MEC computing infrastructures and that it coordinates the deployment of a required network functions and applications at the edge according to ultra-low latency requirements. This can also include the smart deployment of dedicated edge functions and applications that provide specific information on the IoT network and services and can assist and be used by other edge functions and applications to improve the service they implement.

In addition, continuous monitoring of the deployed edge/MEC application(s) and resources is required to facilitate their dynamic and flexible management (e.g. for activation, de-activation or migration assisted by AI/ML algorithms).

4.3.2 Use of ML for Slice Optimization

The end-to-end network slices managed by the NG-IoT network slice orchestration framework are composed by network, compute and storage resources, which are combined to implement (virtual) network functions and satisfy the requirements of the high-level services of the customer/vertical. For instance, such resources could result in the delivery of RAN slices, used by the UEs to access the network and send data with a guaranteed bandwidth or latency, as well as (virtual) network functions instantiated to provide 5G core network functions. All of these resources have to be orchestrated and delivered together on top of a shared end-to-end 5G infrastructure. However, the amount of network and computing resources the 5G infrastructure can offer is limited and is shared by multiple services and slices. For this reason, an optimization of such available resources is twofold: important for the operator to improve their resource usage and for the service to use the assigned resources efficiently.

From an orchestration perspective, the resource optimization is translated into an optimization of the end-to-end network slices: to achieve this goal, the orchestrator platform is supported by AI/ML techniques and algorithms in its decision making processes to drive the actual re-configuration of slices through dedicated interfaces exposed by the various resource controllers sitting on top of the network infrastructure. In literature, different machine learning techniques have been studied not only to optimize the network slices but also to set the more suitable parameters during the end-to-end slice configuration, provisioning, fault detection, forecasting resource utilization, and so on.

In section 4.3.3, , a preliminary study of different machine learning techniques has been made to foresee what are the benefits these techniques bring.

4.3.3 Interaction with other iNGENIOUS Functional Blocks

Figure 44 shows from a high-level perspective the different interactions between the orchestrator MANO supported by the /AI/ML platform and other iNGENIOUS entities. Such interactions are primarily used for the end-to-end network slices configuration, provisioning, management, and monitoring and also include data retrieval to feed the ML algorithm to support MANO during different stages of an end-to-end slice. In particular, within the /AI/ML platform, data acquisition and processing mechanisms are used to feed the AI/ML algorithms for providing output to support MANO decision making for slice management/reconfiguration. Data is received from different sources, such as the DVL, the end-to-end slices deployed and other additional data such as faults, monitoring and so on.

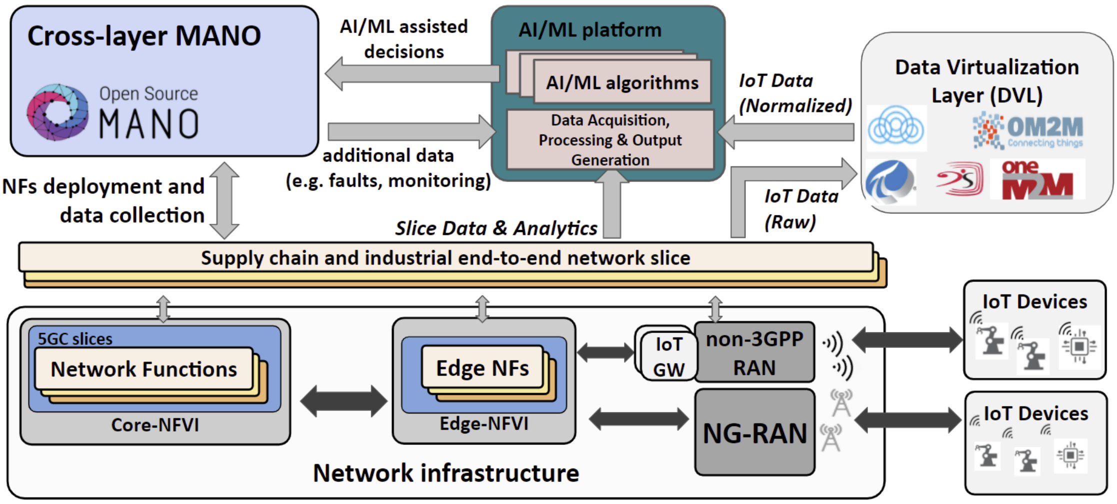


Figure 44: High level MANO interaction with iNGENIOUS entities

To this end, MANO supported by an AI/ML platform interacts with:

- Network infrastructure entities for NFs deployment and data related to NFs and network slices.
- DVL for retrieving IoT data in a normalized format. Using a subscription mechanism, it retrieves the application data from the DVL in order to optimize the supply chain network slice AI/ML.

In particular, MANO supported by an AI/ML platform interacts with the following network infrastructure entities:

- **Edge/MEC (Edge-NFVI in the figure)**: for the deployment and orchestration of edge applications, as well as for data collection to optimize the application placement and lifecycle management. This interaction is realised in the industrial IoT use case.
- **5GC (Core-NFVI in the figure)**: using the so-called Network Exposure Function (NEF) exposed by 5GC, MANO interacts with the 5GC to execute all the network slices and NF provisioning, configuration, monitoring, management, and termination operations. In particular, for what concerns the monitoring, MANO subscribes to specific events and analytics to monitor the slices, NFs, and user equipment status.

Co-funded by the Horizon 2020 Framework Programme of the European Union

- **RAN**: to provision and monitor resources in the heterogeneous iNGENIOUS RAN segment according to the specific slice requirements, interacting with available RAN controllers for non-3GPP and 3GPP technologies (e.g. based on O-RAN for 5G RAN and flexible RAN) using standard interfaces, including where possible and applicable interactions with IoT Gateways.

4.4 High Level Architecture

4.4.1 Main Functional Blocks

From an architectural point of view, the iNGENIOUS cross-layer orchestration is composed of three main management functions as depicted in Figure 45 with the three management and orchestration boxes in green. Moreover, each of these functions carry out a set of specific functionalities within the orchestrator and interact on one side with the AI/ML engine/platform and on the other side with all the resource control and management components.

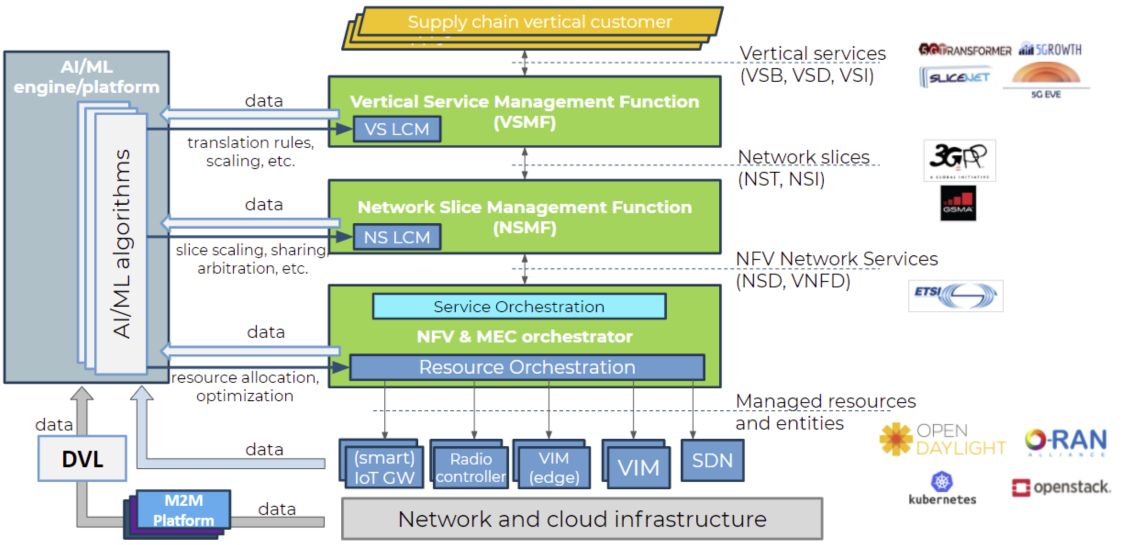


Figure 45: Main functional blocks of the iNGENIOUS MANO layer

The upper functionality of the orchestrator is the Vertical Service Management Function (VSMF) which is in charge of onboarding one or more vertical services and managing their lifecycle. This vertical service offered to the supply chain vertical customer is described from a high-level point of view by the Vertical Service Blueprint (VSB). The VSB includes the parameters of the service, also atomic functional components with their interconnection that compose the VSB itself, while hiding the low-level network (and application) constraints. The VSB offers the verticals the ability to customize their services with specific service constraints. Once the VSB has been properly configured by the vertical specifying service constraints (e.g., bandwidth, latency, maximum number of users and so on), the VSB is translated into a Vertical Service Descriptor (VSD) ready to be instantiated. Once instantiated, the VSMF stores the information of the Vertical Service Instance (VSI) internally and continuously monitors its status for possible optimization, reconfiguration or termination.

Co-funded by the Horizon 2020 Framework Programme of the European Union

The Network Slice Management Function (NSMF) placed under the VSMF, mainly manages the onboarding of NST and Generic network Slice Template (GST) and the lifecycle of NSIs. In a multidomain scenario where a single VSMF interacts with multiple NSMF, more than one NSI can compose a single VSI. An NST contains information about network slice capabilities and requirements, according to the 3GPP Network Resource Model defined in [59], mostly in terms of slice profiles. The slice profile defines the network capabilities expressed in terms of maximum number of UEs supported, the coverage area, the latency, the UEs mobility level and so on. Another way to represent the set of attributes that characterizes the network slice is through the GST [60] defined by the Global System for Mobile Communication Association (GSMA) Standard Organization. Also, in this case, the GST contains a list of attributes that characterizes the network slice. NSTs and GSTs are maintained and exposed to activate NSIs with specific network functions and service performance requirements linked to supply chain and industrial IoT needs. Once activated, the NSMF continuously monitors the status of network slice instance(s) for either possible optimization or (re)configuration or termination.

The lowest layer of the orchestration framework is the NFV & MEC orchestrator. This layer can be further decomposed into two sublayers: the Service Orchestration that interacts with the NSMF and the Resource Orchestration that interacts with all network infrastructure and cloud resource controllers using the offered interfaces. In particular, the Resource Orchestration takes care to coordinate the allocation and configuration of network and cloud resources in different segments and domains (including non-3GPP, 3GPP and flexible RAN, edge and core computing infrastructures to deploy virtualized NFs and applications, SDN backhaul and transport networks) following the ETSI NFV principles and architecture. In particular NFs are modelled as VNFs and the slices requested by the NSMF are mapped to NFV Network Services. This layer of orchestration takes also care to manage network and compute resources at edge locations, where applicable following the MEC architecture.

During the process of network slice instantiation, the instantiation request coming from the NSMF is delivered to the Service Orchestration and then decomposed in one or more Network Service Descriptors (NSDs), a collection of configurations strongly infrastructure-oriented for determining how the Network Service is composed in terms of Virtual/Physical/Container Network Function Descriptors (VNFDs/PNFDs/CNFDs). The NSD and VNFD information model are defined by ETSI NFV in the related specification [61]. Finally, the Resource Orchestration is in charge of interacting with the involved network infrastructure components (non-3GPP and 3GPP Radio controllers, IoT GWs when applicable, VIM at both edge and core compute locations, and SDN controllers) to properly allocate the required network and computing resources.

The resource control layer is placed below the NFV and MEC orchestrator and it represents the interface that makes the communication possible between the slice orchestration functionalities and the network and cloud infrastructure. At the RAN level, the Resource Orchestration within the iNGENIOUS cross-layer MANO is expected to interact with radio controllers (and possibly IoT GWs) available and deployed on top of the radio network

Co-funded by the Horizon 2020 Framework Programme of the European Union

infrastructure to provision non-3GPP as well as 3GPP and flexible RAN resources (in line with the options defined in section 2.1) according to the slice requirements. In this sense, for interaction with 5G and flexible RAN the Resource Orchestration can interact with the O-RAN Non-RT RIC to enforce slice-level policies related to QoS requirements (as stated in 2.1.3). When multiple access technologies are available in the same radio network infrastructure, an interaction with Smart IoT GWs can also be supported to properly select best suited connection technology. For what concerns the allocation of compute resources at edge and core locations to deploy virtualized NFs and applications, the Resource Orchestration coordinates edge and core VIMs based on OpenStack [62] and Kubernetes [63]. The former is a cloud-computing platform for controlling a large amount of compute, storage, and networking resources, while the latter for managing containerized services facilitating the configuration and automation. Indeed, both VIM solutions can be used to deploy the 5G Core NFs (as per section 3.1), and enable the Resource Orchestration to manage and coordinate their configuration and operation in support of the specific network slices to deliver. Similarly, the edge VIM is responsible for the management of computing, storage, and networking resource at the edge, where applicable following the MEC approach for deploying applications and NFs. Moreover, when SDN functionalities are implemented in the network infrastructure, the Resource Orchestration can interact with SDN controllers to dynamically allocate backhaul and transport network connectivity to fulfil the requirements of end-to-end industrial IoT and supply chain slices. This can be done, according to the specific SDN capabilities of the network infrastructure, by leveraging on SD-WAN solutions, as well as on de-facto standard opensource SDN controllers such as OpenDaylight [64]. Whenever the backhaul connectivity has to be provided as a satellite link, dedicated service-oriented interactions with the satellite network operator have to be implemented at the Resource Orchestration level. Indeed, as stated in section 2.3.6, a realistic end-to-end network slice scenario in iNGENIOUS considers two independent management systems to realize and provision end-to-end network slices spanning between terrestrial and satellite segments. On the one hand, the iNGENIOUS cross-layer slice orchestrator takes care to manage 5G RAN and 5G Core resources and NFs, on the other the satellite network orchestrator (owned by a satellite network operator) offers satellite backhaul connectivity as a service in support of end-to-end slices.

Beyond these network slice oriented coordination functionalities, each orchestration layer can be assisted by an AI/ML platform in its decision-making processes to optimize at least part of the lifecycle management procedures. This means for example that resource allocation and network functions placement could be optimized at NFV & MEC orchestration level. Network slice scaling, sharing and arbitration (to trade slice resources across slices) could be adjusted during operation at the Network Slice Management Function level, while vertical service translation rules as well as service scaling policies may be optimized at runtime at the Vertical Service Management Function level.

4.4.2 Workflows for Slicing and Integration with 5GC and NG-IoT

Co-funded by the Horizon 2020 Framework Programme of the European Union

The operations performed on the end-to-end network slices involve all three layers of the orchestrator, the resource control and management layer, and also different network entities. The next paragraph the high-level workflow of an end-to-end slice provisioning is described. The same can be applied to its reconfiguration and termination of a network slice.

Figure 46 illustrates the workflow of the end-to-end industrial IoT slice provisioning. Before a vertical service is instantiated, a vertical should either select an existing VSB or onboard a new one according to its high-level requirements in terms of type of service, bandwidth, latency, and so on. Once the VSB has been properly selected, the by high-level requirements are configured such as the maximum number of users, geographical constraints, and so on. Once configured the instantiation request is performed to the VSMF, i.e. the service is requested. The VSMF translates this vertical service instantiation request into one or more NSTs/GSTs and the related network slice instantiation requests are performed towards the NSMF.

The NSMF once it has received the network slice instantiation request, first performs different checks relate to Service Level Agreement (SLA), constraints, and so on and continues to decompose the network slice instance request into one or more NSDs. Finally, it performs one or more network service instantiation requests towards the NFV/Edge and Resource Orchestrator.

At the NFV/Edge orchestrator level, these instantiation requests are finally translated into the network functions deployment requests on the corresponding network infrastructure entities performed by the different components of resource control and management layer. What network infrastructure entities are involved in this process highly depends on the type of end-to-end slice and the requested vertical in the first place. Usually, the RAN and core part are involved in this process, while the MEC part is not always because it strongly depends on the vertical requirements. Concerning the Core part, the NEF function is used as an entry point to interface with 5GC.

At this point, all the network and compute resources allocated are ready to be used by the high-level service. An asynchronous response process is used to notify the NSMF about the network functions instantiation, and the VSMF about the network slice instantiation and finally the vertical about the end-to-end slice provisioned.

Co-funded by the Horizon 2020
Framework Programme of the European Union

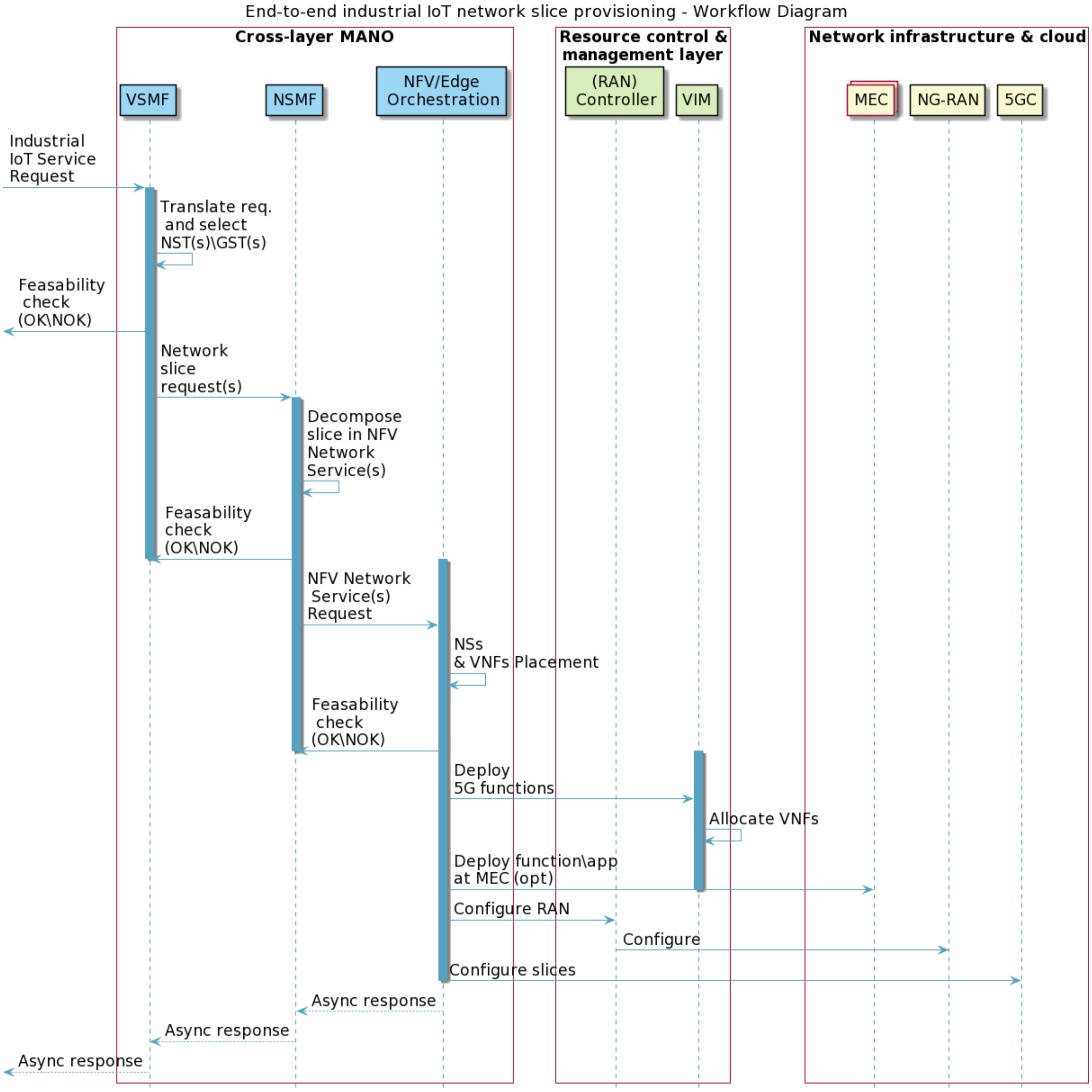


Figure 46: End-to-end industrial IoT network slice provisioning workflow

The interaction between the cross-layer MANO supported by AI/ML and the different entities in iNGENIOUS are primarily carried out for network slice creation, deployment, and management. The following paragraphs, two possible workflows of such interactions are described.

In the high-level workflow diagram depicted in Figure 47, it is shown how MANO supported by an AI/ML platform is able to provision an end-to-end network slice. In particular, it can be noted how MANO finds, thanks to the support of AI/ML platform, the most suitable parameters to build, create and deploy an end-to-end network slice. The AI/ML platform determines the suitable e2e slice parameters using an optimisation model on the received telemetry data. The model can be initially trained from generic network data, and further refined using collected data during operation. Then, after having properly translated the end-to-end network slice into multiple actions, the different segments of network infrastructure are properly configured to provision it.

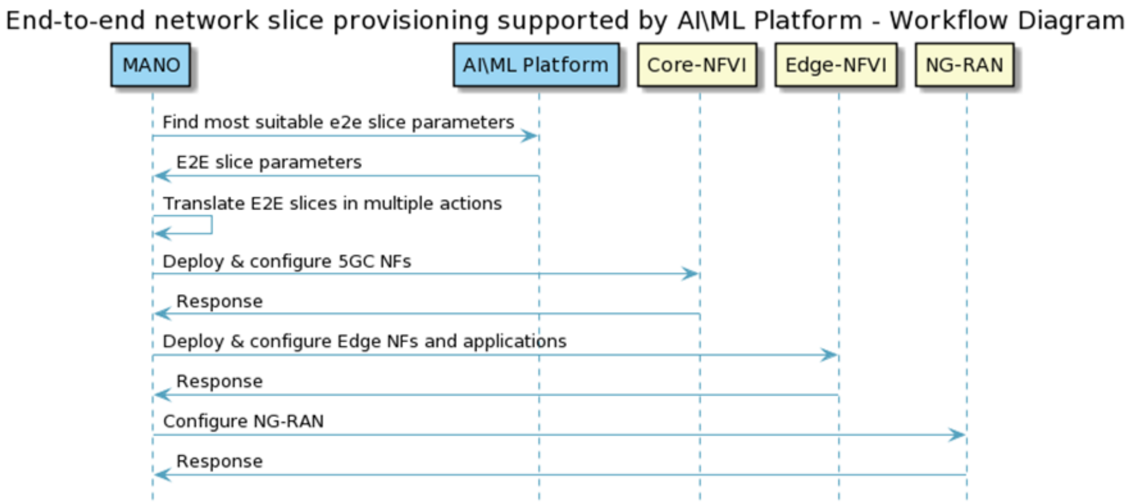


Figure 47: End-to-end network slice provisioning high level workflow

4.4.3 ML Based Network Slice Optimization

The AI/ML platform algorithms can be thought of as a set of flexible rules that have initial conditions from the beginning of operation. As the network data is accumulated, the ML algorithms will update these rules to ensure the best possible efficiency in network decisions. In general, ML algorithms aim to build a model from the available data and predict outputs based on a new valid but unknown input. Thus, machine learning techniques are helpful to obtain analytics from valuable information from raw data and generate insightful pieces of advice and predictions.

Such ML algorithms can be broadly classified into three macro categories:

- **Supervised learning** based on labelled data "learns" or deduces from a training dataset, composed by a set of input and an output, and then predicts given a valid input what would be the output based on what it has learned.
- **Unsupervised learning** based on unlabelled or unstructured data "learns" or deduces the input patterns even though no explicit output or feedback is provided.
- **Reinforcement learning**, which performs iterative and continuous learning based on the reward/punishment mechanism after having taken a particular action. The main aim of reinforcement learning is to collect as much reward as possible based on the state of the system it is monitoring and the consequent actions it takes.

Once the end-to-end slice has been correctly provisioned, it is possible to obtain its information and related statistics for possible optimization and/or reconfiguration. This process can be achieved by performing a subscription to the NetWork Data Analytics Function (NWDAF). In Figure 48, it is shown how the subscription mechanism performed by MANO and supported by AI/ML allows the system to initially collect data and subsequently make decisions based on the AI/ML algorithm. The actions the platform performs highly depend on the data received and the final aim of the AI/ML algorithm. In the

Co-funded by the Horizon 2020 Framework Programme of the European Union

Figure 48, it is assumed that the final aim is to optimize the end-to-end slice usage.

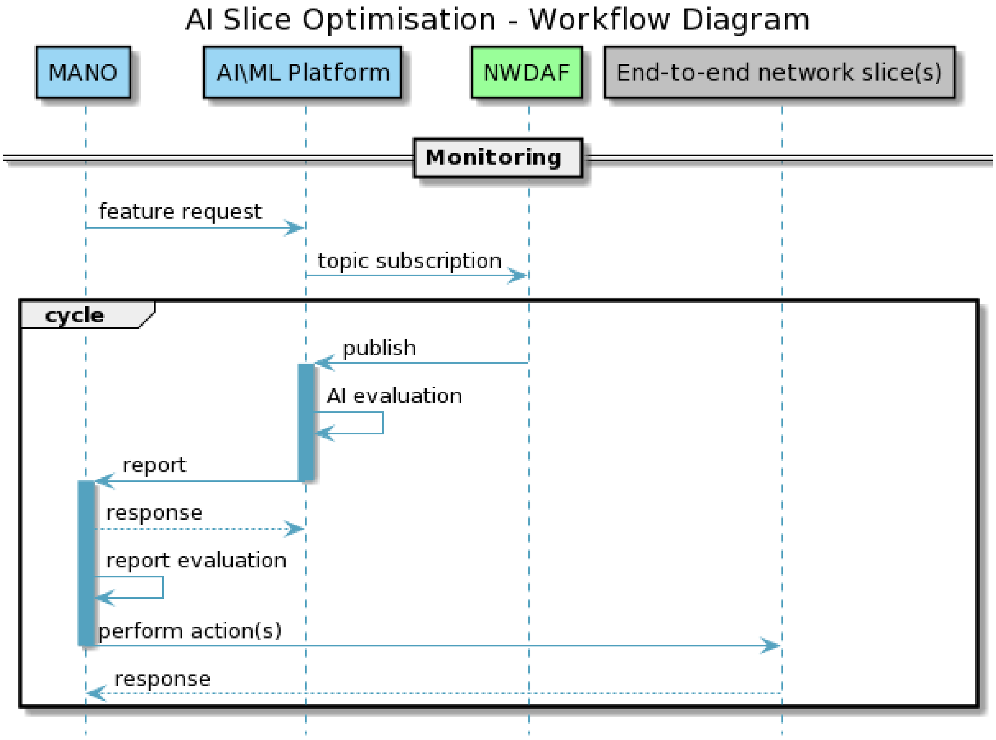


Figure 48: AI/ML slice optimization high level workflow

Each option outlined above contains a list of machine learning techniques that can be used to improve the slice configuration, provisioning, and management. The following machine learning techniques that can improve MANO in the decision-making process have been identified:

- **Support Vector Machine (SVM)**: a supervised learning-based approach for classification and regression analysis. After the training process with a multidimensional training dataset, the SVM classifies new unknown data in one of the two categories, i.e. two regions of the multidimensional space used to perform the classification. In the case of the MANO platform, the SVM can likely help in the configuration, provisioning, monitoring, and fault detection of an end-to-end network slice.
- **Gradient Boosted decision tree**: a supervised learning-based approach for classification. This machine learning technique, similarly to the SVM, could assist MANO to the configuration, provisioning, and fault detection of an end-to-end network slice. The difference between this approach and SVM are currently under investigation.
- **K-mean clustering**: an unsupervised learning-based approach to identify k clusters of data objects in a dataset. In general, this technique classifies the available heterogeneous data in those clusters which could help MANO to identify some data pattern for the end-to-end network slice management and monitoring and result in proactive action being taken.

Co-funded by the Horizon 2020 Framework Programme of the European Union

- **All the reinforcement learning techniques**: an agent using the trial-and-error approach, "learns" or deduces what are the results, expressed in terms of reward or punishment, based on the actions it has performed. In this way, it helps MANO in the decision concerning the network slice management and (re)configurations.

All the above learning techniques are used to train a model which would be subsequently used (and dynamically updated) in the AI/ML layer. In particular, for the reinforcement learning techniques, the trial and error phase that would result in poor user experience will not be experienced by a user, as with most learning techniques, this phase is done by an operator

4.5Relation to UCs and Expected Innovation

The NG-IoT network slice orchestration high level architecture presented in this chapter provides an initial description of the expected cross-layer MANO functionalities to be supported and implemented within the iNGENIOUS platform in support of the various industrial IoT and supply chain UCs. In particular, while the NG-IoT network slice orchestration work is at its early design and development stages, at the time of writing some of the functionalities are planned to be fully implemented and demonstrated in the context of the project UCs, while others will be kept at the conceptual level.

In particular, the cross-layer MANO functionalities are planned to be developed, integrated and demonstrated mostly in the context of the Automated Robots with Heterogeneous Networks, Situational Understanding and Predictive Models in Smart Logistics Scenarios and Supply Chain Ecosystem Integration UCs. Other functionalities relevant for other UCs are planned, at the time of writing, to be considered as design-level features only. However, in both cases, detailed architecture and functionalities specification will be provided with deliverable D4.4. Table 5 below lists the main NG-IoT network slice orchestration functionalities introduced in the previous sections, and for each of them how they map with the various relevant iNGENIOUS UCs in terms of innovation maturity. For this mapping, three options are available: i) state-of-the-art, meaning that the related functionality is supported with state-of-the-art technologies, ii) innovation concept, for those innovative functionalities relevant to the UCs but that will not be demonstrated, iii) innovation demonstrated, for those functionalities that will be implemented and showcased as part of the UC. The mapping provided in this Table 5 has to be considered valid at the time of writing, and it may evolve in the future according to the evolution of UCs implementation and requirements in terms of network slice provisioning and orchestration. For the sake of readability of the table, the following list provides a brief summary of each network slice orchestration functionality:

- **End-to-end network slice provisioning**: it is the capability of the cross-layer slice orchestrator to provide the automated creation of network slices and the deployment of the required NFs in support of industrial IoT and supply chain services.

Co-funded by the Horizon 2020 Framework Programme of the European Union

- **Orchestration and slicing of Flexible RAN**: it refers to the capability of the Resource Orchestration to manage network resources when the RAN includes flexible PHY/MAC for 5G-NR functionalities.
- **Orchestration and slicing of 5G RAN**: it refers to the capability of the Resource Orchestrator to configure and manage slice resources and QoS policies in the 5G RAN, e.g., leveraging on the functionalities offered by the O-RAN Near-RT RIC.
- **Orchestration of satellite backhaul**: it is the capability of the cross-layer slice orchestrator to automatically provision satellite backhaul connectivity services, and stitch them with 5G RAN and 5G Core NFs and slice resources to deliver end-to-end slices spanning across terrestrial and satellite segments.
- **Orchestration and data collection from 5G Core**: it refers to the capability of the Resource Orchestration to automatically deploy, configure and operate 5G Core NFs as part of the network slices. It includes the interaction with the NEF for slice configuration and the NWDAF for slice, NFs and UEs related data collection.
- **Orchestration of applications and NFs at edge/MEC**: it is the capability of the cross-layer MANO to support deployment, configuration and operation of NFs at the edge locations in support of specific latency or real-time requirements of URLLC network slices.
- **AI-assisted (IoT/DVL-data based) slice optimization**: it refers to the capability of the various slice orchestration layers and management functions (namely VSMF, NSMF and NFV & MEC orchestrator) to be assisted in their lifecycle management decision making by external ML algorithms that process IoT domain related data collected from the DVL.
- **AI-assisted (network-data based) slice optimization**: it refers to the capability of the various slice orchestration layers and management functions (namely VSMF, NSMF and NFV & MEC orchestrator) to be assisted in their lifecycle management decision making by external ML algorithms that process network slice related data collected from the 5G Core NWDAF (and possibly other network or compute infrastructure related sources).

Table 5: network slice orchestration functionalities mapped to UCs

Functionality	Factory UC	Transport UC	Port Entrance UC	AGV UC	Ship UC	DVL UC
End-to-end network slice provisioning	innovation demo	innovation concept	innovation concept	innovation concept	innovation concept	
Orchestration and slicing of Flexible RAN	innovation demo					
Orchestration and slicing of 5G RAN	innovation demo			state of the art		

Co-funded by the Horizon 2020 Framework Programme of the European Union

Orchestration of satellite backhaul		innovation concept			innovation concept	
Orchestration and data collection from 5G Core	innovation demo					innovation concept
Orchestration of applications and NFs at edge/MEC	innovation demo	state of the art		state of the art		
AI-assisted (IoT/DVL-data based) slice optimization			innovation concept			innovation demo
AI-assisted (network-data based) slice optimization	innovation demo					innovation concept

Co-funded by the Horizon 2020
Framework Programme of the European Union

5 Conclusion

This document has described the benchmarking and planned innovations of iNGENIOUS with respect to IoT networks beyond 5G. In particular, this deliverable has detailed how the planned innovation related to WP4 will be employed in the use cases (UCs). The document has been divided in three parts, namely, Radio Access Network (RAN), core network, and network slice orchestration.

With respect to the RAN, this document has described several technologies that enable heterogeneous connectivity. In particular, 3GPP and non-3GPP RAN will be considered in iNGENIOUS, such as NB-IoT/LTE-M, 5G, LoRaWAN, Sigfox, flexible RAN, etc. Satellite and ML for network optimization were also investigated. Moreover, the innovations of iNGENIOUS related to RAN include Flexible PHY/MAC, AI/ML for RAN, mmWave deployment and the Smart IoT Gateway. Among these innovations, the mmWave deployment and Smart IoT Gateway are planned to be demonstrated in the UCs.

Regarding the Core Network, this document covered the functionality provided by the mobile packet core to deliver data to either local edge computing, cloud services or fixed local area networks. Additional functions for time sensitive networking (TSN) were considered as well as the required interfaces between 5G Core and SDN components to deliver the networking solutions for IoT applications. The Core Network innovations described in this document Include the network slice manager, 5GLAN application function (AF), TSN AF and SDN MBO. Among these functionalities, the first three are planned to be demonstrated in the UCs.

In relation to network slice orchestration, the document has shown how the Management and Orchestrations (MANO) architectures will be extended in iNGENIOUS. Furthermore, the workflows for slicing and integration with 5GC and NG-IoT and ML slice optimization strategies were described. This document has shown a list of innovation contributions of iNGENIOUS related to the network slice orchestration, namely, end-to-end network slice provisioning, orchestration and slicing of flexible and 5G RAN, orchestration of satellite backhaul, orchestration and data collection from 5G Core, orchestration of applications and NFs at edge/MEC and AI-assisted (IoT/DVL-data based) slice optimization. Except for the orchestration of satellite backhaul functionality, all other Innovations are planned to be demonstrated In the UCs.

Co-funded by the Horizon 2020 Framework Programme of the European Union

References

[1] iNGENIOUS Deliverable 2.1 Use Cases, KPIs and Requirements. [Online]. Available at: D2.1
[2] Verizon News Center, Press Release, "What are Radio Access Networks and 5G RAN?", February 2020.
[3] Ericsson White Paper, Mobile radio access networks and 5G evolution, "5 keyfacts about 5G radio access networks", 2020.
[4] Mikko Säily, Carlos Barjau, Eds., 5G-Xcast Deliverable D3.3, " RAN Logical Architecture and Interfaces for 5G-Xcast", February 2019.
[5] 3GPP TS 38.300 V16.4.0 (2020-12): "NR; NR and NG-RAN Overall Description; Stage 2 (Release 16)", December 2020.
[6] 3GPP TS 38.401 V16.4.0 (2021-01): "NG-RAN; Architecture description (Release 16)", January 2021.
[7] 3GPP TS 36.300 V16.4.0 (2020-12): "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 16)", December 2020.
[8] 3GPP TS 38.410 V16.3.0 (2020-09): " NG-RAN; NG general aspects and principles (Release 16)", September 2020.
[9] 3GPP TS 38.420 V16.0.0 (2020-07): " NG-RAN; Xn general aspects and principles (Release 16)", July 2020.
[10] A. Lavric, A. I. Petrariu and V. Popa, "Long Range SigFox Communication Protocol Scalability Analysis Under Large-Scale, High-Density Conditions," in IEEE Access, vol. 7, pp. 35816-35825, 2019.
[11] Sigfox Technical Overview. [Online - accessed in 16.04.2021]. Available: PDF available.

[12] ITU-R M.[IMT-2020.TECH PERF REQ]—Minimum Requirements Related to Technical Performance for IMT–2020 Radio Interface(s), document ITU-R M.2410-0, International Telecommunication Union—Recommendations, Nov. 2017.

[13] A. Nimr, D. Zhang, A. B. Martinez, and G. P. Fettweis, "A study on the physical layer performance of GFDM for high throughput wireless communication," in EUSIPCO. IEEE, 2017, pp. 638–642.
[14] C. Tai, B. Su, and P. Chen. Optimal filter design for GFDM that minimizes PAPR under performance constraints. In 2018 IEEE Wireless Communications and Networking Conference (WCNC), pages 1–6, April 2018.
[15] R. Bomfin, D. Zhang, M. Matthé, and G. Fettweis, "A theoretical framework for optimizing multicarrier systems under time and/or frequency-selective channels," in IEEE Communications Letters, 22(11), 2394-2397.
[16] A. Nimr, M. Chafii, and G. P. Fettweis, "Precoded-OFDM within GFDM Framework," in VTC Spring. IEEE, 2019, pp. 1–5.
[17] Z. Li, A. Nimr, and G. Fettweis, "Implementation and Performance Measurement of Flexible Radix-2 GFDM Modem," in 2019 IEEE 2nd 5G World Forum (5GWF), 2019, pp. 130–134.
[18] H2020 ORCA project, D3.5: Final operational real-time SDR platforms, https://www.orca-project.eu.
[19] A. Nimr, M. Chafii, and G. P. Fettweis, "Unified Low Complexity Radix-2 Architectures for Time and Frequency-Domain GFDM Modem," IEEE Circuits and Systems Magazine, vol. 18, no. 4, pp. 18–31, Fourthquarter 2018.

Co-funded by the Horizon 2020 Framework Programme of the European Union

[20] A. Nimr, M. Chafii, M. Matthe, and G. P. Fettweis, “Extended GFDM framework: OTFS and GFDM comparison,” in GLOBECOM. IEEE, 2018, pp. 1–6.

[21] de Looper, Christian; "What is 5G? The next-generation network explained", Digital Trends. April 25, 2020.

[22] Stagg, Matt; “BT - Exploring the virtual future of sport broadcasting”, online available at https://business.bt.com/insights/the-future-is-now/future-of-sport-broadcasting.

[23] S. R. Pokhrel, J. Ding, J. Park, O. -S. Park and J. Choi, "Towards Enabling Critical mMTC: A Review of URLLC Within mMTC," in IEEE Access, vol. 8, pp. 131796-131813, 2020.

[24] H. Zhang, N. Liu, X. Chu, K. Long, A. -H. Aghvami and V. C. M. Leung, "Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges," in IEEE Communications Magazine, vol. 55, no. 8, pp. 138-145, Aug. 2017.

[25] P. Popovski, K. F. Trillingsgaard, O. Simeone and G. Durisi, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," in IEEE Access, vol. 6, pp. 55765-55779, 2018.

[26] M. Ballerini, T. Polonelli, D. Brunelli, M. Magno and L. Benini, "NB-IoT Versus LoRaWAN: An Experimental Evaluation for Industrial Applications," in IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7802-7811, Dec. 2020.

[27] D. Feng et al., "Toward Ultrareliable Low-Latency Communications: Typical Scenarios, Possible Solutions, and Open Issues," in IEEE Vehicular Technology Magazine, vol. 14, no. 2, pp. 94-102, June 2019.

[28] C. She, C. Yang and T. Q. S. Quek, "Cross-Layer Optimization for Ultra-Reliable and Low-Latency Radio Access Networks," in IEEE Transactions on Wireless Communications, vol. 17, no. 1, pp. 127-141, Jan. 2018.

[29] C. She et al., "A Tutorial on Ultrareliable and Low-Latency Communications in 6G: Integrating Domain Knowledge Into Deep Learning," in Proceedings of the IEEE, vol. 109, no. 3, pp. 204-246, March 2021.

[30] R. Dong, C. She, W. Hardjawana, Y. Li and B. Vucetic, "Deep Learning for Radio Resource Allocation with Diverse Quality-of-Service Requirements in 5G," in IEEE Transactions on Wireless Communications.

[31] 5G/NR-FR/Operating Bandwidth, [Online - accessed in 16.04.2021] Available: http://www.sharetechnote.com/html/5G/5G_FR_Bandwidth.html.

[32] Hoffman, Chris; "What is 5G, and how fast will it be?", How-To Geek LLC. Jan. 24, 2019.

[33] Horwitz, Jeremy; "The definitive guide to 5G low, mid, and high band speeds", VentureBeat online magazine. April 23, 2020.

[34] ITU's IMT-2020 document, International telecommunications Union, 2020.

[35] Open5GCore. The next mobile core network testbed platform. [Online

Co-funded by the Horizon 2020 Framework Programme of the European Union

- accessed in 16.04.2021]. Available: https://www.open5gcore.org/.

[36] Openstack. [Online - accessed in 16.04.2021]. Available: https://www.openstack.org/.

[37] SatCube Ku. [Online - accessed in 16.04.2021]. Available: http://www.satcube.com/products/satcube-ku/.

[38] H2020 5G PPP Phase III project 5G-VINNI (5G Verticals Innovation Infrastructure), [Online - accessed in 16.04.2021]. Available: https://www.5g-vinni.eu/.

[39] SaT5G Consortium, "SaT5G: Satellite and Terrestrial Network for 5G", July 2017. [Online - accessed in 16.04.2021]. Available: http://sat5g-project.eu.

[40] SA Tis5 Consortium, "SATis5 - Demonstrator for Satellite-Terrestrial Integration in the 5G Context", ESA ARTES, October 2017. [Online]. Available: https://satis5.eurescom.eu/.

[41] B. Tiomela Jou, O. Vidal, F. Arnal, J.-M. Houssin, K. Liolis, J. Cahill, H. Khalili, P. Sayyad Khodashenas, M. Boutin, D.-K. Chau, S. Sendra Diaz, "Architecture Options for Satellite Integration into 5G Networks", 27th European Conference on Networks and Communications (EuCNC 2018), Ljubljana, Slovenia, June 2018.

[42] M. Corici, K. Liolis, et al., "SATis5 Solution: A Comprehensive Practical Validation of the Satellite Use Cases in 5G", 24th Ka and Broadband Communications Conference and 36th International Communications Satellite Systems Conference (ICSSC), Niagara Falls, Ontario, Canada, 2018.

[43] 3G PP TS 23.501: System Architecture for the 5G System.

[44] ETSI TR 103 611: Satellite Earth Stations and Systems (SES); Seamless integration of satellite and/or HAPS (High Altitude Platform Station) systems into 5G system and related architecture options.

[45] ETSI MEC Deployment towards 5G. [Online - accessed in 16.04.2021]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp24_MEC_deployment_in_4G_5G_FINAL.pdf.

[46] ETSI MEC Deployment towards 5G. [Online - accessed in 16.04.2021]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf.

[47] 3GPP TS 23.502.

[48] 3GPP TS 23.501.

[49] 3GPP TR 28.801.

[50] ETSI GR NFV-EVE 012 v3.1.1, Dec 2017.

[51] 5G-TRANSFORMER. [Online - accessed in 16.04.2021]. Available: http://5g-transformer.eu/.

[52] SLICENET. [Online - accessed in 16.04.2021]. Available: https://slicenet.eu/.

Co-funded by the Horizon 2020 Framework Programme of the European Union

[53] 5G-EVE. [Online - accessed in 16.04.2021]. Available: https://5g-eve.eu/.

[54] 5GGROWTH. [Online - accessed in 16.04.2021]. Available: https://5growth.eu/.

[55] V. P. Kafle, Y. Fukushima, P. Martinez-Julia and T. Miyazawa, "Consideration On Automation of 5G Network Slicing with Machine Learning," 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), Santa Fe, Argentina, 2018, pp. 1-8.

[56] A. Thantharate, R. Paropkari, V. Walunj and C. Beard, "DeepSlice: A Deep Learning Approach towards an Efficient and Reliable Network Slicing in 5G Networks," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2019, pp. 0762-0767.

[57] S. Sevgican, M. Turan, K. Gökarslan, H. B. Yilmaz and T. Tugcu, "Intelligent network data analytics function in 5G cellular networks using machine learning," in Journal of Communications and Networks, vol. 22, no. 3, pp. 269-280, June 2020.

[58] L. Le, B. P. Lin, L. Tung and D. Sinh, "SDN/NFV, Machine Learning, and Big Data Driven Network Slicing for 5G," 2018 IEEE 5G World Forum (5GWF), Silicon Valley, CA, USA, 2018, pp. 20-25.

[59] ETSI TS 128 541 [Online - accessed in 16.04.2021]. Available: https://www.etsi.org/deliver/etsi_ts/128500_128599/128541/15.02.00_60/ts_128541v150200p.pdf.

[60] Generic Network Slice Template [Online - accessed in 16.04.2021]. Available: https://www.gsma.com/newsroom/wp-content/uploads//NG.116-v3.0.pdf.

[61] ETSI GS NFV-IFA 014 [Online - accessed in 16.04.2021]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/014/02.03.01_60/gs_NFV-IFA014v020301p.pdf.

[62] OpenStack. [Online - accessed in 16.04.2021]. Available: https://www.openstack.org/.

[63] Kubernetes [Online - accessed in 16.04.2021]. Available: https://kubernetes.io/.

[64] ODL. [Online - accessed in 16.04.2021]. Available: https://www.opendaylight.org/.

Co-funded by the Horizon 2020 Framework Programme of the European Union