Enhancing Competitiveness, Resilience and Sustainability of Remote Farming, Forestry and Rural Areas through Holistic Assessment of Smart XG, Last-mile and Edge Solutions' Gains

D3.1: Technology and edge processing enablers, and service processing optimisation solutions (1st version)

Lead Author: Kostas Chartsias (ICOM)

Funded by
the European Union

Grant Agreement No.	101060294		
Project Acronym	XGain		
Project Title	Enhancing Competitiveness, Resilience and Sustainability of Remote Farming, Forestry and Rural Areas through Holistic Assessment of Smart XG, Last-mile and Edge Solutions' Gains		
Type of action	HORIZON Research and Innovation Actions		
Call Topic	HORIZON-CL6-2021-COMMUNITIES-01-03		
Project Start Date	September 1st, 2022	**Project End Date**	August 31st, 2025
Project URL	xgain-project.eu		
Work Package	WP3 | XGain Connectivity and Edge Technology Enablers		
WP Lead Beneficiary	Intracom S.A. Telecom Solutions (ICOM)		
Deliverable type[1] | Dissemination level[2]	R | PU		
Contractual due date	30 June 2023	**Actual submission date**	29 June 2023
Lead Author (s)	Kostas Chartsias (ICOM)		
Contributors	Kostas Chartsias (ICOM), Tanel Järvet (CAFA), Marinos Charalambides (ICCS), Anargyros J. Roumeliotis (ICCS), August Betzler (i2CAT), Adrián Pino (i2CAT), Dimitrios Kritharidis (ICOM), Konstantinos Stamatis (ICOM), Ioannis Neokosmidis (INC), Leonardo Andreasi (ASTRO), Aravindh Raman (TID)		
Internal reviewer(s)	Marinos Charalambides (ICCS), Anargyros J. Roumeliotis (ICCS), August Betzler (i2CAT)		

Disclaimer

Funded by the European Union. Views and opinions expresses are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA) (granting authority). Neither the European Union nor the granting authority can be held responsible for them.

Copyright message

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

[1] *R: Document, report; DEM: Demonstrator, pilot, prototype, plan designs; DEC: Websites, patents filing, press & media actions, videos, etc.; DATA: Data sets, microdata, etc; DMP: Data management plan; ETHICS: Deliverables related to ethics issues; SECURITY: Deliverables related to security issues; OTHER: Software, technical diagram, algorithms, models, etc.*

[2] *PU – Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page); SEN – Sensitive, limited under the conditions of the Grant Agreement; Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444; Classified C-UE/EU-C - EU CONFIDENTIAL under the Commission Decision No2015/444; Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444*

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1	8/3/23	15	Initial deliverable structure	Kostas Chartsias (ICOM)
V0.2	31/03/23	15	QA review	Panayiotis Klitou (eBOS)
V0.3	28/4/23	35	1st merge of T3.1 and T3.2 content, T3.1 amendments	Kostas Chartsias (ICOM), Tanel Järvet (CAFA), Marinos Charalambides (ICCS)
V0.4	17/5/23	80	T3.1 and T3.2 further amendments, T3.3 content	Kostas Chartsias (ICOM), Tanel Järvet (CAFA), Marinos Charalambides (ICCS), Anargyros J. Roumeliotis (ICCS)
V0.5	26/5/23	90	T3.1, T3.2, T3.3 amendments	Kostas Chartsias (ICOM), Marinos Charalambides (ICCS), Anargyros J. Roumeliotis (ICCS)
V0.6	30/5/23	95	Introduction, Executive summary, Conclusions, Editorial work	Kostas Chartsias (ICOM)
V0.7	2/6/23	97	T3.3 amendments, T3.1 amendments, Editorial work	Anargyros J. Roumeliotis (ICCS), Kostas Chartsias (ICOM)
V0.8	16/6/23	100	1st document after the review	Kostas Chartsias (ICOM)
V0.9	21/6/23	100	QA review	Panayiotis Klitou (eBOS)
V1.0	28/6/23	100	Final review	Marinos Charalambides (ICCS), Anargyros J. Roumeliotis (ICCS)

XGain Consortium			
#	Participant Organisation Name	Short name	Country
1	INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	ICCS	EL
2	BENCO BALTIC DOO ZA SAVJETOVANJE IUSLUGE	BENCO	HR
3	CAFA TECH OU	CAFA	EE
4	IKE KAPTEN KOOUTS	CAPT	EL
5	EBOS TECHNOLOGIES LIMITED	eBOS	CY
6	EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW - EN VISSERIJONDERZOEK	EVILVO	BE
7	FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA	I2CAT	ES
8	INCITES CONSULTING SA	INC	LU
9	INTRACOM SA TELECOM SOLUTIONS	ICOM	EL
10	RAINNO IDIOTIKI KEFALAIOUCHIKI ETAIREIA	RAIN	EL
11	REGIONAL DEVELOPMENT FUND OF CENTRAL MACEDONIA	RDFCM	EL
12	STICHTING WAGENINGEN RESEARCH	WR	NL
13	UAB ART21	ART	LT
14	ZENTRUM FUR SOZIALE INNOVATION GMBH	ZSI	AT
15	TELEFONICA INVESTIGACION Y DESARROLLO SA	TID	ES
16	Astrocast SA	ASTRO	CH
17	JONATHAN MICHAEL SMITH	JS	UK

Table of Contents

1 Executive Summary 14
2 Introduction 15
2.1 Mapping XGain Outputs 16
2.2 Deliverable Overview and Report Structure 17
3 Characterisation of Connectivity Technologies 18
3.1 Access Technologies Overview 19
3.2 Transport Technologies Overview 22
3.3 Connectivity Technologies in Depth 23
3.3.1 Access Networks 23
3.3.2 Transport Networks 37
3.4 Signal Propagation Constraints 43
4 Edge Processing Enablers 48
4.1 Characterisation of Edge Processing Technologies 48
4.1.1 Compute Continuum Overview 48
4.1.2 Generic Technological Characteristics of Edge Processing Enablers 49
4.1.3 Virtualisation Solutions for Edge Computing 56
4.2 Edge Processing Technologies in Depth 56
4.2.1 Extreme Edge Processing Enablers 56
4.2.2 Far and Near Edge Processing Enablers 59
5 Service Processing Optimisation and Green ICT 63
5.1 Technology Selection Algorithm 63
5.2 Optimisation Problem 68
5.3 Indicative Examples of the Algorithm 69
5.3.1 Toy Example (Video Analytics) 70
5.3.2 Toy Example (Low Data Processing) 73
5.3.3 Drone-based Video Analysis 75
6 Conclusions 79
7 References 80

List of Figures

Figure 1: Mobile and fixed network architecture for rural and remote areas 18

Figure 2: Analysis of contributions from the 2018-2021 study period by Q 5/1 vice-rapporteur - access technologies 20

Figure 3: Analysis of contributions from the 2018-2021 study period by Q 5/1 vice-rapporteur - transport technologies 22

Figure 4: Simplified overview of a mobile network 24

Figure 5: Simplified overview of a FWA deployment 25

Figure 6: Simplified overview of a Wi-Fi deployment 26

Figure 7: Simplified overview of a LPWAN deployment....28

Figure 8: Starlink dish power consumption....30

Figure 9: Satellite IoT access configuration 31

Figure 10: Satellite orbits used for IoT services 31

Figure 11: Passive optical network architecture 34

Figure 12: Active ethernet network 34

Figure 13: Microwave Long-Haul connectivity for remote areas....38

Figure 14: Satellite IoT Backhaul configuration 40

Figure 15: Illustration of satellite backhaul link 42

Figure 16: Oneweb constellation and satellite connected to a gateway in the UK 42

Figure 17: Attenuation of signals in our atmosphere by $H2O$ and $O2$ 44

Figure 18: The intensity of precipitation can affect atmospheric attenuation 44

Figure 19: Signal propagation through foliage 45

Figure 20: Propagation paths in a rural environment 46

Figure 21: Satellite constellations and blockage from objects....46

Figure 22: Different types of compute continuum deployments....49

Figure 23: Different types of SBCs....57

Figure 24: Nvidia Jetson modules....58

Figure 25: Indicative mini PCs 59

Figure 26: Indicative tower servers 60

Figure 27: Comparison between CPUs and GPUs based on GFLOPS/W 61

Figure 28: Filter matrix and preference vector 65

Figure 29: Optimisation problem 68

Figure 30: Solution of optimisation problem no1 72

Figure 31: Solution of optimisation problem no2 75

Figure 32: Solution of optimisation problem no3 78

List of Tables

Table 1: Adherence to XGain GA deliverable & tasks description....16
Table 2: Comparison of access network technologies....20
Table 3: Comparison of transport network technologies....23
Table 4: 5G NR by spectrum and transceivers....24
Table 5: Wi-Fi specifications....26
Table 6: Satellite constellations and characteristics....29
Table 7: Technical characteristics of some of the players in satellite IoT....32
Table 8 : Characteristics of passive and optical networks....35
Table 9: xDSL characteristics....35
Table 10: Prices of FTTx components....36
Table 11: Power consumption per “group” of FTTx components....37
Table 12: Terrain constraints and communication technologies....47
Table 13: General characteristics of different compute continuum zones....54
Table 14: Extreme processing enablers comparison....58
Table 15: Comparison of different types of GPUs....61
Table 16: Filter matrix....63

Glossary of terms and abbreviations used

Abbreviation / Term	Description
2G	2nd Generation mobile network
3G	3rd Generation mobile network
3GPP	3rd Generation Partnership Project
4G	4th Generation mobile network
5G	5th Generation mobile network
AAU	Active Antenna Unit
AI	Artificial Intelligence
API	Application Programming Interface
ASIC	Application Specific Integrated Circuit
BBU	Baseband Unit
BPON	Broadband Passive Optical Networks
BS	Base Station
CA	Carrier Aggregation
CO	Central Office
COTS	Commercial off-the-shelf
CPE	Customer Premises Equipment
CPU	Central Processing Unit
DNN	Deep Neural Networks
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
EPON	Ethernet Passive Optical Networks
ESON	Ethernet Switched Optical Network

FDD	Frequency-Division Duplexing
FLOPS	Floating-point Operations per Second
FOV	Field of View
FSPL	Free Space Path Loss
FTTB	Fibre to the Building
FTTC	Fibre to the Curb
FTTCab	Fibre to the Cabinet
FTTH	Fibre to the Home
FTTN	Fibre to the Node
FWA	Fixed Wireless Access
GEO	Geostationary orbit
GPON	Gigabit Passive Optical Networks
GPU	Graphics Processing Unit
GW	Gateway
HDMI	High-Definition Multimedia Interface
I/O	Input/Output
ICT	Information and Communication Technology
IDU	Indoor Unit
IoT	Internet of Things
IP/MPLS	Internet Protocol/Multi-Protocol Label Switching
ISL	Inter-Satellite Links
ITU	International Telecommunication Union
KFT	Knowledge Facilitation Tool
KPI	Key Performance indicator

LEO	Low Earth Orbit
LH	Long Haul
LIDAR	Light Detection and Ranging
LoS	Line-of-Sight
LPWAN	Low Power Wide Area Networks
M2M	Machine-to-Machine
MEO	Medium Earth orbit
ML	Machine Learning
MW	Microwave
NB-IoT	Narrowband Internet of Things
NFV	Network Functions Virtualisation
NOC	Network Operation Centre
NPU	Neural Processing Unit
NR	New Radio
NTE	Network Termination Equipment
ODU	Outdoor Units
OLT	Optical Line Termination
ONT	Optical Network Termination
ONU	Optical Network Unit
OTN	Optical Transport Networks
P2M	Point to Multi Point
P2P	Point to Point
PC	Personal Computer
PON	Passive Optical Network

PoP	Point of Presence
QoS	Quality of Service
RAM	Random-Access Memory
RAN	Radio Access Network
RF	Radio Frequency
RRH	Remote Radio Head
RRU	Remote Radio Unit
SBC	Single-Board Computers
SDK	Software Development Kit
SDN	Software-Defined Networking
SIMD	Single-Instruction Multiple-Data
SoC	System on a Chip
SPI	Serial Peripheral Interface
STM	Synchronous Transfer Mode
TDD	Time-Division Duplexing
TDM	Time-Division Multiplexing
UART	Universal Asynchronous Receiver/Transmitter
UAV	Unmanned Aerial Vehicle
UC	Use Case
UE	User Equipment
UGV	Unmanned Ground Vehicle
UHF	Ultra High Frequency
UPF	User Plane Function
USB	Universal Serial Bus

VHF	Very High Frequency
VM	Virtual Machines
WDM	Wavelength Division Multiplexing
WDM PON	Wavelength Division Multiplexing Passive Optical Network
Wi-Fi	Wireless Fidelity
WiGig	Wireless Gigabit Alliance
WiMax	Worldwide Interoperability for Microwave Access
XPIC	Cross-Polarisation Interference Cancelling

1 Executive Summary

The XGain project aims to promote sustainable and inclusive development in rural, coastal, and sub-urban areas by providing access to a comprehensive inventory of smart XG, last-mile connectivity, and edge computing solutions for relevant stakeholders such as municipalities, policymakers, farmers, foresters and their associations. The project's overall objective is to deliver a Knowledge Facilitation Tool (KFT) to aid in business model development and decision-making in selecting technology ecosystems that increase systemic resilience and energy efficiency, contribute to climate mitigation, and reduce digital divides among different groups.

The purpose of this report is to provide an in-depth assessment of the available remote connectivity and edge computing technology portfolio, stemming both from the consortium expertise as well as the current state-of-the-art of the available Commercial off-the-shelf (COTS) solutions. The target of the deliverable is to collect rich information that will serve as a database for the KFT, regarding performance characteristics, deployment specificities and limitations, power consumption as well as indicative prices of the relevant ICT infrastructure.

Subsequently, the gathering of communication technologies as well as edge processing solutions can be combined, in a holistic approach, considering the rural Use Cases' (UCs) requirements in order to develop end-to-end integrated and optimised technological propositions, as the first step of the KFT's decision support engine. Each technology can be the best fit for certain scenarios, while not being appropriate for other scenarios. To this end, in this deliverable, algorithmic mechanisms for the optimisation of computation and communication resources are being designed and evaluated.

2 Introduction

For those living in rural and remote areas, especially young people, the Internet is a road out of isolation, a pathway to education, understanding, wider opportunities and a richer life. A good example to understand where we are standing is the fact that only 65% of Europe's rural areas enjoy connectivity of more than 30 Mbps [1]. This creates a real obstacle for the use of all the emerging technologies that could potentially benefit greatly from the development of rural areas. Alongside this, are the more immediately felt improvements to our day-to-day quality of life. Good connectivity can be the difference between young people and families putting down roots in rural areas or leaving them to seek better education and work opportunities; between farmers and agricultural businesses, thriving or struggling; between being able to age well at home, in the heart of their community, or having to move away to find the medical care and daily support they need. For the sustainability of rural communities themselves, reliable high-speed broadband communication networks can make all the difference.

In conjunction with the high-speed connectivity solutions, services and applications can be made available to the rural locations and isolated areas by deploying a plethora of edge computing solutions. Edge computing provides solutions that cater for different segments in the rural areas such as agriculture, where farm productivity can be achieved by farmers. In-site analysis with regard to soil pH, moisture and meteorological conditions monitoring, can lead to real-time decision making for the optimisation of production, such as optimised water management and monitoring of the production quality. Edge computing can also be applied to emergency situations such as forestry/firefighting where drones/UAVs can be used to obtain georeferenced footages and images which are sent to the local edge computing platform of the security forces, allowing better response times, better control and optimised decision-making in face of the overall situation. As a last example, edge computing solutions can help to realise the "One Health idea" [2] which states that the health of people is connected to the health of animals, plants, and our shared environment. Precision farming and decreased usage of chemicals and fertilisers in agriculture, livestock and aquaculture monitoring, as well as digital health and wellbeing devices, are some applications that can leverage edge computing hardware/software capabilities and can lead to optimised health of people, animals and the environment.

The XGain project is in line with the ambitions of the EU's Green Deal, the Digital Age and an Economy that works for people, ensuring that "No one is left behind". It aims to strengthen the capacities of farmers, foresters, and rural communities through digital connectivity gains. As a first step, in this deliverable, a plethora of communication and edge computing solutions that can be deployed in rural areas will be compared with each other, and optimised end-to-end deployment propositions will be suggested to the relevant stakeholders.

2.1 Mapping XGain Outputs

Table 1: Adherence to XGain GA deliverable & tasks description

XGain GA Component Title	XGain GA Component Outline	Respective Document Chapter(s)	Justification
DELIVERABLE			
D3.1 Technology and edge processing enablers, and service processing optimisation solutions (Version 1) Initial report on technology and edge processing enablers, and optimisation solutions.			
TASKS			
Task 3.1	Task 3.1 will perform an exhaustive investigation of the available connectivity solutions for rural connectivity environments. The investigation will build on the expertise of technology providers available in the consortium, and their technology offerings, but will further extend to state-of-the-art and Commercial-off-the-self (COTS) solutions.	Section 3	This section provides a thorough overview and comparison of different access and transport communication technologies.
Task 3.2	Task 3.2 will investigate and develop a portfolio of edge computing solutions to address the different set of requirements and challenges considered part of the data processing enablers for the edge location (as well as, where applicable, the related cloud-based processing).	Section 4	This section provides a thorough overview and comparison of different edge computing solutions.
Task 3.3	Task 3.3 will delve into the details of end-to-end system design and eventual service provisioning, taking into consideration technological capabilities and limitations of both connectivity (T3.1) and edge computing (T3.2) enablers, along with	Section 5	This section provides algorithmic mechanisms for the optimisation of computation and communications resources according to the requirements of rural Use Cases.

	service/performance (non-)functional requirements (T2.3).		

2.2 Deliverable Overview and Report Structure

This deliverable presents the first outcomes from 3 tasks, namely T3.1 “Technology Enablers for Connectivity in Remote Areas”, T3.2 “Edge Processing Enablers” and T3.3 “Service Processing Optimisation and Green ICT”. This deliverable serves as an initial report of D3.1 on Month 10 and will be complemented and enhanced by a second deliverable (Version 2) which will be reported in Month 26.

More specifically the document is divided logically into three main sections:

1. Classification of state-of-the art communication solutions into access and transport and the collection of relevant KPIs and specifications.
2. Classification of state-of-the art edge computing solutions into Extreme, Far, Near edge, as well as the collection of relevant KPIs and specifications.
3. Algorithmic techniques for the optimisation of both computation and communications resources, based on rural Use Case requirements and the information gathered in the aforementioned sections.

3 Characterisation of Connectivity Technologies

The increase of digitalisation across all industries and sectors (i.e., health, education, transportation, farming, aquaculture etc.), also exacerbated by the Covid-19 pandemic, highlighted and amplified the already known digital demand. Digital accessibility and connectivity for rural communities and businesses still present a great challenge to many rural and remote areas. As next generation telecommunication networks emerge, it is vitally important that rural areas are not left behind.

Telecommunication networks are normally divided into three parts, as outlined in Figure 1: access, transport/backhaul and core, where the transport segment routes traffic from access network sites into the core network. More specifically the main segments of the telecommunication infrastructure are as follows [3]:

- **Last-mile or access network:** Refers to the final leg of the telecommunications networks that delivers telecommunication services to end-users' premises.
- **Middle-mile network,** or **backhaul:** This is the distribution network that connects the national backbone to a point in an outer locality/geographic area for broader distribution out to the last-mile network.
- **National backbone** (or core) network: This connects international Internet traffic (through undersea or terrestrial fibre-optic cables) via submarine cable landing stations (or terrestrial gateways for land borders) to the national high-speed, high-capacity backbone network.

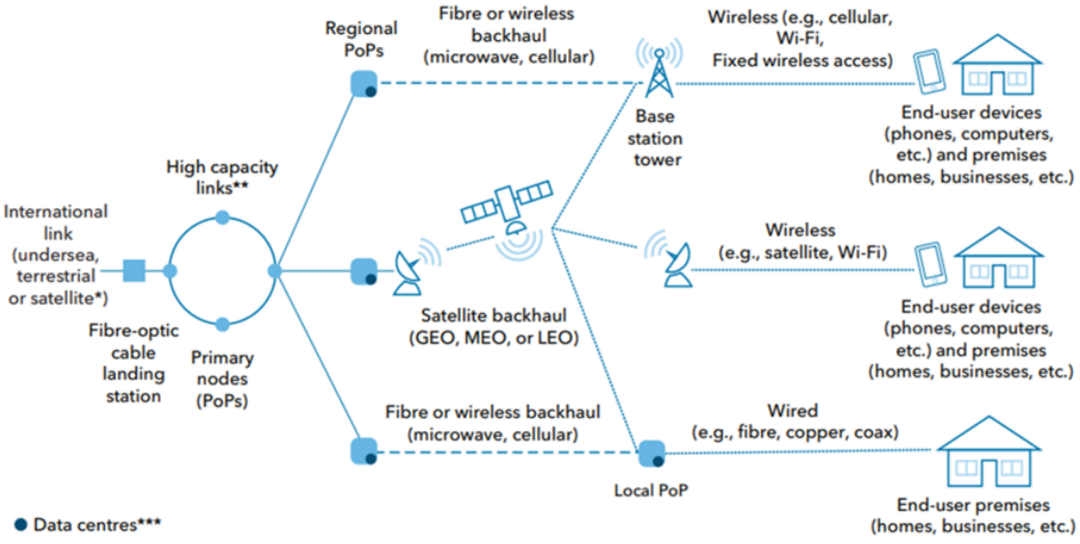


Figure 1: Mobile and fixed network architecture for rural and remote areas [3]

Both wireless and wired technologies can be used in the access and backhaul segments. Since the invention of the optical cable, its use for transport has become the standard design pattern for high-capacity national networks. On the other hand, for the access segment, the dispersed nature of the area involved makes wireless technologies equally effective as wired ones. However, different types of rural regions require different solutions, as there is no single technology that can meet the requirements for any rural setup. In fact, adopting a given solution would depend, for example, on geographic/terrain characteristics and distance to the nearest gateway/exchange point, among other parameters.

The explanation and definition of all the terms included in the tables of the following two sub-sections are provided next, in the order they appear in the tables.

- **Throughput** and **Range**: In any communication system, two of the most important and tightly interrelated specifications are throughput and range. Throughput indicates how much data can be transferred from a source to a destination at any given time and range is a measurement of the maximum distance between two telecommunication nodes.
- **Latency**: The latency of a communication network is defined as the time needed to transport information from a sender to a receiver.
- **Mobility**: Mobility refers to a network node changing its point of attachment to the network while its communication to the network infrastructure remains uninterrupted.
- **Required Infrastructure**: In addition to the telecommunication systems and their accessories, this term also includes essential components that are required for their correct functionality, such as ducts/pipes for underground cables or poles for the mounting of wireless equipment.
- **Spectrum Licensing**: For wireless telecommunications, unlicensed spectrum means that is free and anyone can broadcast in the respective spectrum block. On the other hand, licensed spectrum is reserved by spectrum management authorities and requires the payment of a fee in order to have the right to broadcast a signal.

3.1 Access Technologies Overview

Access or last-mile connectivity refers to the final leg of a telecommunication network that connects end-users, such as homes, businesses, or mobile devices, to the wider network infrastructure. Last-mile connectivity is often the most challenging and expensive part of building a telecommunication network. It can be impacted by factors such as distance, terrain, population density, and the availability of infrastructure, which make it particularly challenging for rural and coastal regions. It is, though, critical for delivering reliable and high-speed data services to the end-users, to enable a wide range of applications, for example in agriculture, aquaculture and farming, which are just some of the Use Cases that XGain is addressing.

There are several access network technologies that can be used to provide Internet access in these areas, including Wi-Fi, WiGig, Mobile Cellular, Fixed Wireless Access (both 5G-based and proprietary), satellite (broadband and Internet of Things (IoT)), Low Power Wide Area Networks (LPWAN) such as LoRaWan or Sigfox, fibre, copper and coaxial cable. Each technology has its own strengths and weaknesses, and as it has already been stated, the best option for a particular area will depend on factors such as population density, topography, and the availability of existing infrastructure. Based on contributions submitted under Output Report on ”ITU-D Question 5/1 Telecommunications/ICTs for rural and remote areas, Study period 2018-2021” [4], a breakdown of the main access technologies used for connecting rural and remote areas is given in Figure 2.

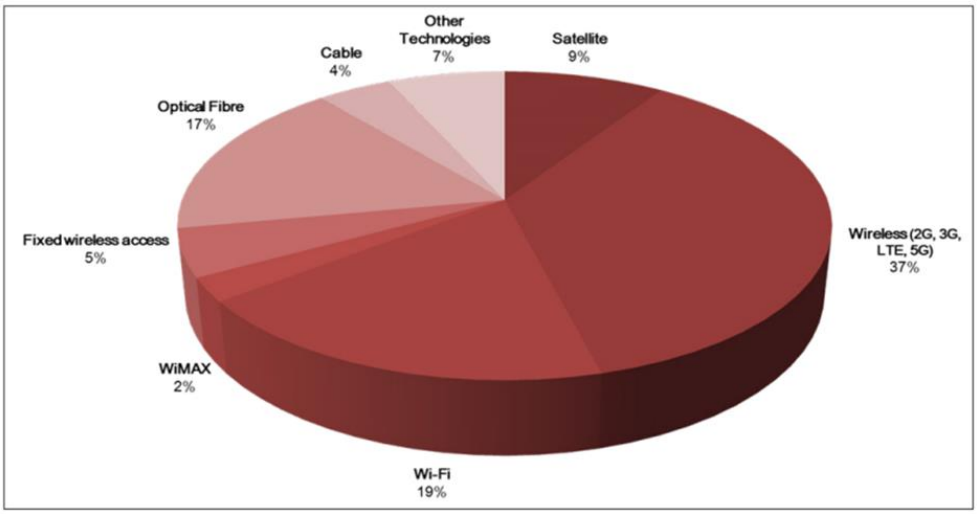


Figure 2: Analysis of contributions from the 2018-2021 study period by Q 5/1 vice-rapporteur - access technologies [4]

Table 2, compares various wireless and wired connectivity technologies which are used in last-mile access network deployment in the light of a range of factors, including throughput, range, latency, infrastructure requirements, spectrum usage, mobility etc. The aforementioned values are theoretical, but can provide a rough but comprehensive comparison of the different technological options available.

Table 2: Comparison of access network technologies

Access Technology	Throughput	Range	Latency	Mobility	Infrastructure Required	Spectrum Licensing	Access Device Type
Wi-Fi	2 Mbps (IEEE 802.11a) to 10 Gbps (IEEE 802.11ax)	100s of m	Low	Limited	Towers and radio equipment	No	Wi-Fi enabled devices/ CPEs, phones, tablets
WiGiG	0.4 to 8 Gbps (IEEE 802.11ad)	A few 100s of m	Low	No	Towers and radio equipment	No	Modems to Ethernet or Wi-Fi
Mobile cellular (2G, 3G, 4G, 5G)	0.1 – 1000 Mbps	5 to 15 Km	Low	Yes	Towers and radio equipment	Yes	Cellular mobile phones, laptops, devices via dongles)

Fixed Wireless Access (4G/5G)	20 – 1000 Mbps	Up to 10 Km	Low	No	Towers and radio equipment	Yes	Modems to Ethernet or Wi-Fi
Fixed Wireless Access (proprietary)	Up to 1000 Mbps	Up to 11 Km	Low	No	Towers and radio equipment	Yes/No	Modems to Ethernet or Wi-Fi
Satellite (broadband)	5 – 150 Mbps	1000s of Km	High	Yes	Earth station, Satellite, Terminal	Yes	Modems to Ethernet or Wi-Fi
Satellite (IoT)	In the orders of bps	1000s of Km	High	Yes	Earth station, Satellite, Terminal	Yes/No	Modems (UART TTL levels)
LPWAN (LoRaWAN , Sigfox)	0.3-50 Kbps Up to 100 bps	Up to 30 Km Up to 40 Km	High High	Limited	Gateway, IoT devices	No	IoT Devices
Fibre	100 – 1 000 Mbps	100s of Km	Low	No	Overhead cabling: Tower, poles, cabinets, active network equipment Below ground: Subterranean duct work, cabinets, active network equipment	No	Modem to Ethernet or Wi-Fi

Copper	0 to 24 Mbps (for ADSL, ADSL 2, ADSL 2+); 100 Mbps (for VDSL, VDSL2, Vectoring); 1 Gbps (G.Fast)	0.1 to 5 Km	Low	No	Tower, poles, cabinets, active network equipment	No	Modem to Ethernet or Wi-Fi

3.2 Transport Technologies Overview

Transport or backhaul technologies are the connections between the last-mile access nodes and the core network. A range of different technologies that are commonly used in access networks can also be used for longer distances in backhaul links, however, microwave, optical and satellite backhaul are the most prevalent solutions. Copper-based technologies continue to be phased out and other wireless technologies (such as WiMAX) are also currently used in limited situations. Once again, the best choice of a transport solution for a specific area can be impacted by factors such as distance, terrain specificities and the availability of nearby infrastructure.

Based on contributions submitted under Output Report on ”ITU-D Question 5/1 Telecommunications/ICTs for rural and remote areas, Study period 2018-2021” [4], a breakdown of the main transport technologies used for connecting rural and remote areas is given in Figure 3.

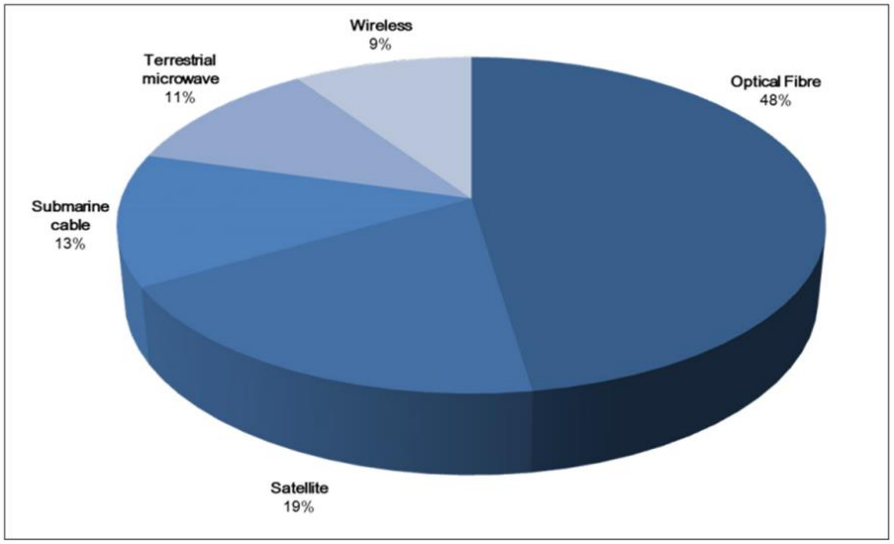


Figure 3: Analysis of contributions from the 2018-2021 study period by Q 5/1 vice-rapporteur - transport technologies [4]

The following table, Table 3, compares various wireless and wired connectivity technologies which are used in transport network deployment in the light of a range of factors, including throughput, technology range and infrastructure requirements and spectrum usage.

Table 3: Comparison of transport network technologies

Backhaul Technology	Throughput	Range	Latency	Mobility	Infrastructure required	Spectrum Licensing
Microwave	5 – 1000+ Mbps	Up to 100 Km	Low	No	Radio equipment, towers/poles	Yes
Satellite broadband	1 – 1600 Mbps	1000s of Km	High	Yes	Satellites, hub earth stations, remote earth stations	Yes
Satellite IoT	100 Kbps to 1000 Kbps	1000s of Km	High	Yes	Satellite, Ground Stations	Yes
Fibre	100 – 1000 Mbps	100s of Km	Low	No	Fibre-optic cable installed in-ground or overhead via poles	Not applicable

3.3 Connectivity Technologies in Depth

3.3.1 Access Networks

3.3.1.1 Mobile Cellular

Mobile networks have matured over the last two decades from a data speed perspective, but they continue to evolve to enable many new Use Cases. The first generation of mobile networks is now obsolete, but we still have 2G, 3G, 4G and 5G networks active in most parts of the world. Regardless of the generation, a mobile network divides a region into cells, as shown in Figure 4, with each cell covered by a radio base station. These base stations are typically mounted on towers or tall buildings, with the sites owned by tower companies and the equipment owned by telecoms in most developed countries. Mobile devices (smartphones, tablets, etc.) within each cell communicate with the nearest base stations via radio waves, where the signal is then transmitted to the core network either via cables, or high-frequency radio links, to routers connected to the core network.

A radio base station can be functionally separated into two sub-systems:

- BBU (Baseband Unit, sometimes also referred to as a digital unit), which generates and processes a digitised baseband signal.
- RRH (Remote Radio Head, aka RRU, Remote Radio Unit or AAU, Active Antenna Unit), which creates the analogue transmit RF signal from the baseband signal and sources it to the antenna and also digitises the RF received signal.

In brief, the BBU possesses the “digital” function, while the RRH possesses mostly the analogue functions.

Another critical element of the cellular system is the mobile core which is a central part of the overall mobile network and allows subscribers to get access to the services that they are entitled to use. It is responsible for critical functions such as subscriber profile information, location, service authentication, etc. Another relevant characteristic of the core is that it allows for local breakouts towards applications to happen close to the users, in order to minimise latency between the users and their service. More specifically, in 5G networks, the User Plane Functions (UPFs) are placed normally next to the radio base stations, so that traffic for the services offered at the edge to which the UPF connects to, does not need to travel all the way to the (central) 5G core. This reduces end-to-end latencies and also the amount of traffic that needs to be carried over the transport (backhaul) network.

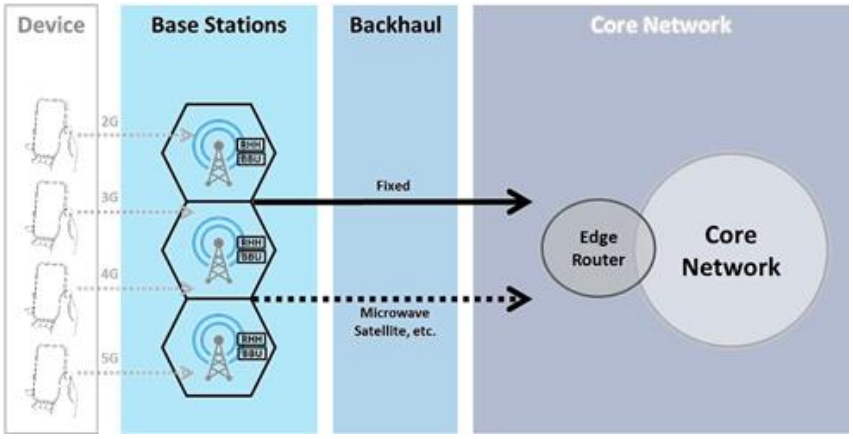


Figure 4: Simplified overview of a mobile network [5]

Table 4: 5G NR by spectrum and transceivers [6]

Frequency Range	Spectrum Tier	Duplex Mode	Transceiver Configuration	Coverage	Peak Speed (BW)
Sub – 1GHz	Low-Band	FDD	4T4R	>10 Km	~ 0.4 Gbps (20 MHz)
2 GHz	Mid-Band	FDD	8T8R/32T32R	~ 5 Km	~ 0.8 Gbps (20 MHz)
2.5 to 6 GHz	Mid-Band	TDD	32T32R/64T64R	1 to 5 Km	~ 6 Gbps (100 MHz)
6 GHz	Mid-Band	TDD	128T/128R	1 to 5 Km	~ 20 Gbps (400 MHz)
mmWave	High-Band	TDD	2T2R/4T4R	< 1 Km	~ 2 Gbps (4 x 100 MHz)

5G is the latest cellular generation and there are three spectrum classifications depicted in Table 4: low-band, mid-band and high-band [7].

Low-band spectrum, which includes frequencies below 1 GHz provides exceptional wide-area coverage qualities and capacity to offer both deep inside and outside-in coverage. The typical deployed channel bandwidths in this spectral range are 5 MHz, 10 MHz, and 20 MHz. Many current cellular networks are based on low band spectrum, which will continue to support these bands in the future.

Mid-band spectrum is usually considered as spectrum above 1 GHz and below 6 GHz but in some cases can address spectrum all the way up to 8GHz. There are a lot of capabilities unlocked in this frequency range with licensed, lightly-licensed (3.55 GHz to 3.7 GHz) and unlicensed spectrum (e.g. 5 GHz) deployment models. Mid-band spectrum is typically optimised for macro cell deployments and/or small cell deployments between 3 GHz and 6 GHz. Collectively bands from 450 MHz to 6 GHz are known as 5G Frequency Range 1 (FR-1).

Bi-directional multi-gigabit speeds are made possible by high-band spectrum. 5G band in this region is known as 5G Frequency Range 2 (FR-2) band, which spans the frequencies of 24 GHz to 52 GHz. Channel bandwidths at these frequencies are commonly 50 MHz or 100 MHz, and it is possible to combine several contiguous channels to enable bi-directional multi-gigabit services. Additionally, the antenna components working at these frequencies have a form factor that is very tiny, allowing for dense arrays and high system capacities. Operating systems in these high-band frequencies can have various drawbacks, such as attenuation caused by vegetation or meteorological conditions. These restrict the adoption of high-band systems for widespread coverage.

Although 5G is getting a lot of attention for its benefits in performance, there are also worries about how much power it consumes [8]. A single 5G station consumes electricity at a rate that is 2.5 to 3.5 times greater than a single 4G station. The AAU which uses a lot of power, is the primary cause of this rise in 5G power consumption. In 5G, the maximum power consumption of a 64T64R AAU at 3.5 GHz is highly dependent on user traffic, from an average of 650 W at zero load per sector, to 900 W at 50% load and to 1100 W at 100% load. Assuming that the BBU power consumption is estimated at 300 W and a 3-sector setup (3 AAUs and 1 BBU), we anticipate that the power consumption can vary from 2250 W up to 3600 W at peak load [9]. Regarding the 5G equipment prices, there is a huge gap between the price of buying a single device and the price of the one (collection) purchased by an operator. For a typical combination of one BBU and 3 AAUs, an average price of 50K€ can be estimated for this combination.

3.3.1.2 FWA

Fixed Wireless Access (FWA), is a type of technology that uses radio waves to send high-speed signals that offer data transfer to and from consumer devices. FWA systems typically consist of a base station connected to a fixed network and a number of subscriber units (Customer Premises Equipment (CPE)) spread out over a wide area. The base station then uses radio waves to communicate with the subscriber units, making it possible for consumers to connect to the fixed network and access high-speed data services. These transmitters are strategically attached to stationary structures such as poles, buildings (as depicted in Figure 5) or towers.

There is a plethora of FWA solutions, whether purpose-built (IEEE 802.11-based or proprietary) or 3GPP-based (LTE/5G). The suitability of one type of FWA solution being better depends on many factors, such as market segments (throughput-distance requirements), consumer demand, availability of spectrum, etc., which are outside of the pure technical evaluation of 3GPP vs purpose-built FWA technology.

Figure 5: Simplified overview of a FWA deployment [10]

It is expected that the total number of FWA connections will double in the next 5 years and 5G-based FWA will make up most of these connections. While 5G FWA potentially supports all frequencies that are also offered by conventional 5G radio access networks, two frequency ranges stand out above other options: i) the millimetre range (e.g. 26-28 GHz) and ii) the sub 6 GHz range, in particular at 3.5 - 4 GHz. The millimetre range offers very high throughput at short – medium distances, with the limitation of being very directive and requiring line-of-sight, whereas the sub 6 GHz range find a compromise between capacity and coverage, not strictly requiring line-of-sight. Lower or higher frequency ranges are limited in their throughput or range, respectively. The peak data rates can vary vastly in FWA technology and depend on the chosen frequency and bandwidth. Using up to several hundred of MHz of spectrum in the mmWave high-band, the achievable peak data rates are much higher than with generic 5G New Radio (NR) access. Speeds of above 1 Gbps are achievable in real physical deployments, with room to improve (at least on paper). Regarding the power consumption of FWA sub-systems, while the CPEs consume around 20 W, the consumption of the base station can vary a lot depending on the actual configuration (array size, number of sectorial panels, etc.). A typical 5G FWA macro base station could consume around 5 KW [11]. Also, the cost of a single sector (4x4) can be expected at around 25K€.

3.3.1.3 Wi-Fi

Wi-Fi works in unlicensed frequencies and therefore does not require a licence to be used. Licence-exempt spectrum allows users to operate any technology for any recreational or business purpose in the band. The technology is supported in a variety of unlicensed, or licence-exempt, spectrum bands: 2.4, 5 and 6 GHz. The frequencies involved give Wi-Fi a potential range of hundreds of metres for local area access points, to tens of kilometres for point-to-point and point-to-multipoint radios – based on, and subject to, regulatory exceptions and the density of Wi-Fi and other devices in the same unlicensed band. The quality of a Wi-Fi connection can thus vary greatly, as it is prone to interference. The range is also very limited due to regulatory constraints: the range of a single average Wi-Fi router is several orders of magnitude lower than that of a cellular station. A simplified overview of a Wi-Fi deployment is depicted in Figure 6.

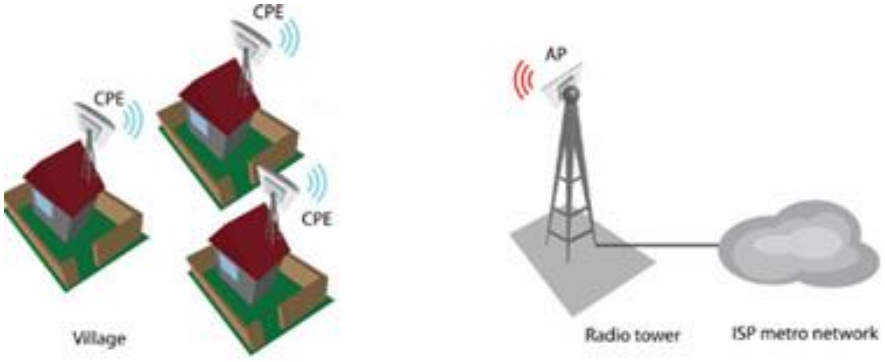


Figure 6: Simplified overview of a Wi-Fi deployment [12]

Since Wi-Fi was first introduced in 1997, it has been constantly improved in terms of capacities and features, reaching the currently already commercialised 6th generation of Wi-Fi (IEEE 802.11 ax) and with Wi-Fi 7 on the horizon. The latest generations are backwards compatible with older generation User Equipment (UE), which makes it easy to upgrade infrastructures without having to replace the rest of devices, unless the newest features are to be supported.

Table 5: Wi-Fi specifications [13]

Frequency Range	Wi-Fi 4	Wi-Fi 5	Wi-Fi 5	Wi-Fi 6E	Wi-Fi 7 (expected)
Launch date	2007	2013	2019	2021	2024
IEEE Standard	802.11n	802.11ac	802.11ax		802.11be
Max Data Rates	1.2 Gbps	3.5 Gbps	9.6 Gbps		46 Gbps
Bands	2.4 GHz and 5 GHz	5GHz	2.4 GHz and 5 GHz	6 GHz	1 – 7.25 GHz (including 2.4 GHz, 5 GHz, 6 GHz bands)
Channel Size	20, 40 MHz	20, 40, 80, 80+80, 160 MHz	20, 40, 80, 80+80, 160 MHz	20, 40, 80, 80+80, 160 MHz	Up to 320 MHz
Modulation	64-QAM OFDM	256-QAM OFDM	1024-QAM OFDM		4096 WAM OFDMA
MIMO	4x4 MIMO	4x4 MIMO, DL MU-MIMO	8x8 UL/DL MU-MIMO		16x16 MU-MIMO

Wi-Fi 6, the current state-of-the art Wi-Fi version, supports the 2.4 GHz and 5 GHz bands, adding the 6 GHz band with Wi-Fi 6E. A plenitude of channel size configurations is supported, as shown in Table 5. While the specifications define a maximum throughput of 9.2 Gbps, a real 2x2 Wi-Fi 6 network can reach up to 1.2 Gbps of real throughput in ideal conditions (no interference, less than 10 m of line of sight between AP and UE). However, under normal conditions, several hundreds of Mbps are realistic data rates for a Wi-Fi 6 network in an indoor network setup. Depending on the antenna chosen for the access point and user equipment, Wi-Fi can reach tens to hundreds of meters of coverage (directive vs. omnidirectional). With everyday devices, like a home router and smartphones, Wi-Fi is considered a short-range Radio Access Network (RAN) technology that covers tens of meters. Note that the highest modulation coding schemes are only reached when being less than 10 m away from an access point. Investigating various websites dedicated to listing and comparing power consumption of wireless routers of newest and older generations, an average power consumption of 10 W can be assumed for a domestically used router. Using the website https://www.tpcdb.com, we obtain an average peak consumption of 8.7 W when considering over 25 of the latest router models. The price of Wi-Fi routers can vary considerably, depending on the quantity of antennas, the supported technologies, and expected coverage. Checking the first 12 results on www.amazon.es, an average cost of 142€ is determined. The list of devices includes a variety of devices, all the way from the most basic routers to high-end gaming routers, most of them supporting the latest Wi-Fi standard.

3.3.1.4 LPWAN

Low Power Wide Area Networks (LPWAN) are terrestrial wireless networks for long-range communication and data transmission at ultra-low bit rates. LPWAN devices use low complexity and low-cost hardware to perform simple task such as sensing environmental parameters, often powered by batteries and can therefore operate in the field up to years. LPWAN technologies are primarily operating in unlicensed spectrum and typically in sub-GHz bands where signal propagation is beneficial to achieve longer range while still using relatively little power. Available bands vary across countries and regions. Using unlicensed spectrum is more prone to interference compared to technologies working in licensed spectrum bands. Limiting the duty cycles (times during which the radio is active) per device, reduces potential interference. LPWAN providers either provide network services where the leading example is Sigfox [14], or technology such as chipsets and designs, with LoRa [15] being the prime example. In Europe, both LoRa and Sigfox operate in the 868 MHz frequency band. A simplified overview of a LPWAN LoRa deployment is depicted in Figure 7, where a star topology is employed in which all end devices are directly connected via one hop links to a gateway/base station.

Figure 7: Simplified overview of a LPWAN deployment

Sigfox is an ultra-narrowband service designed for very low throughput and small packets of up to 12 bytes for uplink messages and 8 bytes for the downlink, respectively. The daily number of packets per device is limited to 140 over the uplink and 4 over the downlink, per day. It offers a maximum rate of 100 bps and its range is up to 40 Km in open field and 10 Km in urban environments [16]. Concerning equipment cost, end devices typically cost less than 2€, while base stations cost over 4000€ (and 6000€ for installation). Since the spectrum is unlicenced, there's no associated cost. The yearly operating cost (electricity) of end devices is roughly 1€ and for base stations is 1000€ [17][19].

The technical characteristics of LoRa include data rates of 0.3 Kbps-50 Kbps, with maximum packet size of 256 bytes (unlimited messages per day), and a range of up to 30 Km in rural areas (5 Km in urban environments) [16]. Concerning CAPEX, end devices typically cost 3-5€ while base stations over 1000€ (and 2000€ for installation) [17][19]. The yearly operating cost (electricity) of end devices is roughly 1€ and for base stations is 100€ [17][19].

The power consumption of end devices for both technologies is extremely low and depends on device type and activity. According to [18], a Sigfox device powered by a 2400 mAh battery, can achieve a theoretical lifetime of 1.5 or 2.5 years while sending one message every 10 minutes at 100 bps or 600 bps, respectively, and an asymptotic lifetime of 14.6 years as the message transmission rate decreases. On average a 10 KWh consumption per device per year can be considered typical, which roughly corresponds to 1 W per device. On the base station side, the power consumption is in the range between 6 and 15 W, with 10 W being a measure of good performance [20].

3.3.1.5 Satellite Broadband

Satellite broadband is a type of high-speed Internet service that uses a satellite in Geostationary Orbit (GEO) or Low Earth Orbit (LEO) or Medium Earth Orbit (MEO) to provide connectivity to users, as shown in Table 6. Instead of relying on traditional land-based infrastructure such as cables or Digital Subscriber Lines (DSL) lines, satellite broadband delivers Internet connectivity directly to a satellite dish located on the user's property.

Satellite broadband works by transmitting data between the user's satellite dish and a satellite. Usually, satellite broadband is considered as a useful option for those in rural or remote areas who have limited access to other types of Internet service.

The components of Satellite Access network typically include:

1. User terminal: The user terminal, commonly known as a satellite dish, is the equipment located on the user's premises that communicates with the satellite. It receives and transmits signals to and from the satellite.
2. Satellite: The satellite is a man-made object placed in orbit around the earth that receives and retransmits signals from the user terminal to the ground station, following a bent-pipe architecture. The satellite operates in a specific frequency band, such as Ku-band or Ka-band.
3. Ground Station/Gateway: The ground station or a gateway is the facility on the ground that communicates with the satellite. It receives signals from the satellite, processes them, and sends them to the Internet backbone network.
4. Internet Backbone Network: The Internet backbone network is the high-speed data network that connects different regions and countries around the world. The network is composed of fibre optic cables and other communication links.

Apart from these elements, the satellite operator also maintains a Network Operation Centre (NOC) to monitor and manage the satellite access network. The facility oversees the performance of the network, troubleshoots issues, and coordinates with other network operators to ensure smooth operation.

Table 6: Satellite constellations and characteristics

Satellite Orbit	LEO	MEO	GEO
Altitude	160 – 2.000 Km	2.000 – 35.786 Km	35.786
Orbital Period	90 to 120 minutes	2 – 24 hours	24 hours
Orbit Shape	Circular	Elliptical	Circular
Coverage Area	Limited	Regional	Global
Latency	Low (20 – 50 ms)	Moderate (20 – 120 ms)	High (around 600 ms)
Cost	High	Medium	High
Propagation Delay	Low	Moderate	High

Satellite Orbit	LEO	MEO	GEO
Number of Satellites	Large constellations	Small constellations	Single satellite or small constellations
Operating Band	Ku-band, Ka-band, C-band, and L-band	Ku-band, Ka-band, X-band, and V-band	Ka-band and Ku-band(18.3–30 GHz)

Figure 8: Starlink dish power consumption

On average, a LEO satellite terminal, such as Starlink, consumes 50W [21]. Figure 8 shows the consumption of dish during the morning hours of a day. Weather conditions, especially rainy periods, impact the power consumption. For example, Figure 8 shows a higher power consumption between 6:45 to 9:10, due to rain. There are different Starlink flavours, such as the Residential and Business flavours. Residential Starlink has a download speed of 50-200 Mbps, upload 10-20 Mbps and the cost of satellite dish is about 500€. Starlink Business has 24/7 service, a download speed of 100-350 Mbps, upload speed 20-40 Mbps and the cost of satellite dish is about 2700€.

3.3.1.6 Satellite IoT

The components of Satellite IoT Access network typically include:

1. User terminal: composed of a satellite modem and an antenna (form factor will depend on the specific transmit, receive frequency of the modem). The terminal is used to transmit or receive data to the user endpoint via the satellite constellation.
2. Satellite: The satellite is a man-made object placed in orbit around the earth that receives and retransmits signals from the user terminal to the ground station, following a bent-pipe architecture. The satellite Machine-to-Machine (M2M) payload operates in a specific frequency band, such as UHF, VHF, L-band, S-band, etc.
3. Ground Station/Gateway: The ground station or a gateway is the facility on the ground that communicates with the satellite. It receives signals from the satellite, processes them, and sends them to the backend of the company providing the connectivity solution.
4. Cloud platform: The user can retrieve or upload data to or from the terminal thanks to a web portal or Application Programming Interfaces (APIs)

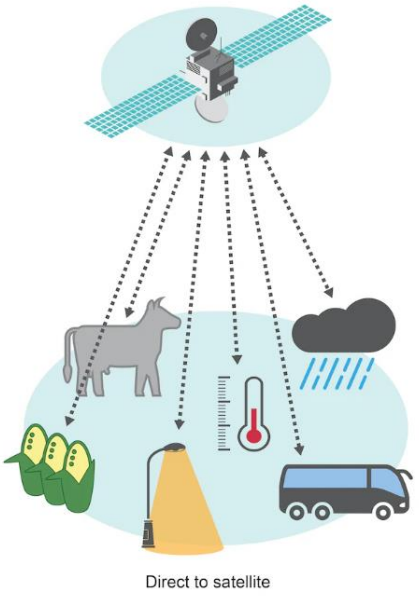


Figure 9: Satellite IoT access configuration [22]

The devices themselves can range from simple sensors to complex machines, and can be used in a variety of industries, including agriculture, transportation, energy, and environmental monitoring.

One of the key advantages of satellite IoT is its ability to provide ubiquitous coverage, which means that devices can be connected anywhere in the world. This can enable a wide range of applications, such as tracking wildlife, monitoring weather patterns, and managing global supply chains.

Satellite IoT uses mainly LEO orbits as they are well suited for low power applications. A shorter distance between the terminal and the satellite means lower propagation losses (estimated at approximately 25 dB less than with GEO), which reduce the user equipment's power requirements and make it ideal for communication with low-power IoT devices as shown in Figure 9.

Today, most LEO satellites for IoT are CubeSats, which allows companies to mass-produce them using of the shelf components. It drastically reduces the cost and time to design and develop satellites.

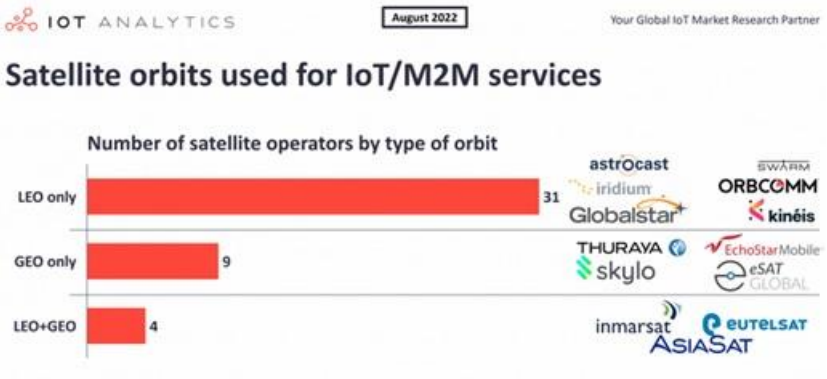


Figure 10: Satellite orbits used for IoT services [23]

Satellite IoT Access latency depends on different factors such as:

- Satellite constellation:

  - Number of orbital planes and number of satellites per orbital plane: the more we have, the less we have to wait/search for a satellite.
  - ISL (Inter Satellite Link): a mesh network can be established between the satellites of the constellation. The satellite that is in view of a ground station will relay the data of the other satellites. ISL can be between satellites in LEO-LEO, LEO-GEO, etc.
- Ground segment: the denser is the network, the less the satellites in the constellation will have to wait before offloading the data that have been stored in their memory.
- Density and locations of the terminals: higher density of terminals in the region will lead to collisions which will trigger multiple retransmissions of the messages.
- Type of antenna used: will define the Field of View (FOV) of the terminal. The larger, the more it will see satellites, the lower is the latency.
- Environmental conditions: anything that will reduce the FOV of the antenna will of course trigger multiple retransmissions of the data which will have an impact on the latency.

Typically, for satellite IoT, access latency (terminal to satellite) can go from the order of seconds to the order of hours. A simplified overview of satellite orbits used for IoT services is depicted in Figure 10. The following table summarises the main characteristics of some satellite IoT providers:

Table 7: Technical characteristics of some of the players in satellite IoT [24],[25],[26],[27][28],[29]

Name	Module	RF Band	Payload Data (Bytes)	Constellation	Power (mW)
Astrocast	Astronode S	L – Band	168	LEO	Tx=251 Rx=158
Swarm	M138	VHF	400	LEO	Tx=3300 Rx=132
Kinesis	KIM1	UHF	30	LEO	Tx=250 - 1000 Rx=N/A
EchoStarMobile	EM2050	S – Band	50	GEO	Tx=1056 Rx=N/A
Iridium	Iridium 9602	L – Band	340	LEO	Tx=7500 Rx=975
Myriota	M2-24	UHF	20	LEO	Tx=1881 Rx=109

These IoT modules are usually integrated with 3rd party processors and microcontrollers and they accommodate different communication interfaces such as Universal Asynchronous Receiver/Transmitter (UART) and Serial Peripheral Interface (SPI). The power consumption of the IoT modules can range from a couple of mWatts in

sleep mode up to a couple of Watts when they transmit or receiver and their costs can also range from 50 up to 200€.

3.3.1.7 Fixed Wireline Access

In the past, various fixed wireline access network architectures have been proposed. Fixed wireline access technology refers to the infrastructure and systems used to provide high-speed Internet and communication services through physical wired connections. It is a reliable and established method of delivering connectivity to homes, businesses, and other premises. In fixed wireline access, data and voice signals are transmitted using various types of cables, such as copper or fibre-optic, which are installed and maintained by telecommunications companies. These cables connect the user's premises to a central network hub, typically located at an Internet Service Provider (ISP) facility or a telephone exchange. Nowadays, fibre is the common technology used.

These architectures can be categorised based on the termination point of the optical fibre. The more general term Fibre-To-The-x (FTTx) is used to describe these architectures, which denotes the replacement of part or all of the copper cabling with optical fibre cables. Therefore, depending on the degree of penetration of the optical fibre, we can distinguish the following architectures: Fibre To The Home – FTTH (the optical fibre terminates inside the user's home or workplace), Fibre to the Building – FTTB (the fibre is terminated outside of end user's home or workplace), Fibre To The Node - FTTN / Cabinet - FTTCab / Curb – FTTC (the path from the main point of the service provider to the intermediate node (outside cabinet) that serves an entire neighbourhood consists exclusively of optical fibre).

FTTH technologies can be divided into two main categories, those implemented using active devices (devices that require power supply) and those implemented using passive devices [30],[31]. The former is usually also known as point-to-point technologies (Point to Point, P2P), while the latter as passive optical networks (Passive Optical Networks, PONs) or networks from point to multiple points (Point to Multi Point, P2M). The choice of each technology depends on the type of service to be transmitted, the cost of the infrastructure, the existing infrastructure and future plans to move to new technologies.

Passive Technologies

Passive optical networks are Point-To-Multipoint (P2MP) networks [32]. Each customer is connected to the optical network through a passive optical splitter, so no active electronic devices are used in the distribution network and the bandwidth is shared from the feeder to the edge branch. The PON architecture consists of the Central Office (CO) in which the Optical Line Termination (OLT) is located, the Optical Distribution Network (ODN) which contains the passive splitters and finally the Optical Network Unit (ONU) in combination with the Network Termination Equipment (NTE) or otherwise Optical Network Termination (ONT) in the case that the termination equipment is identical to the ONU. A typical architecture of a passive optical network is shown in Figure 11.

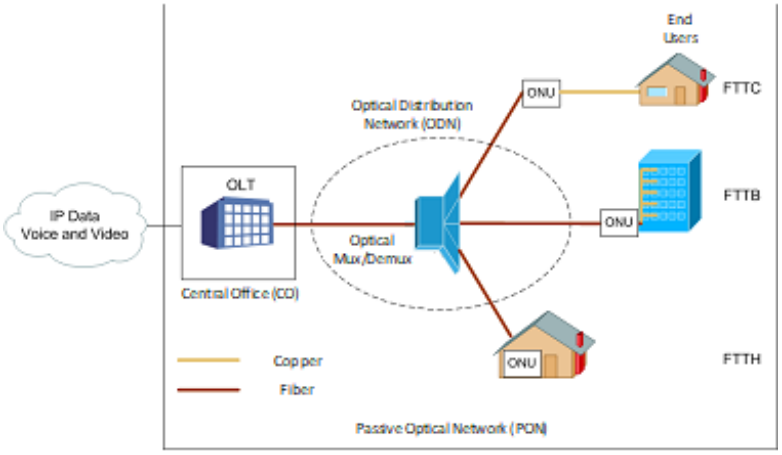


Figure 11: Passive optical network architecture

The number of ONUs that can be connected to the power splitter is limited by the power losses introduced by the splitter and the links between the OLTs and ONUs. The ITU G.983.1 standard recommends sharing the signal to a maximum of 32 users.

The standards used to develop FTTH passive optical networks are ATM Passive Optical Networks (APON) which were renamed to Broadband Passive Optical Networks (BPON), Gigabit PON (GPON), Ethernet PON (EPON) and Wavelength Division Multiplexing PON (WDM PON). GPON and WDM PON standards which are widely used, are analysed below.

Active Technologies

Active Ethernet also known as Ethernet Switched Optical Network (ESON) or P2P network architecture, provides a dedicated fibre from the Central Office (CO) to each end user. Point-to-point architecture is essentially the simplest network design.

The core switch, the aggregation switch and the ONT are the main building blocks of a P2P network. The core switch is a high-capacity Ethernet switch that communicates with the aggregation node using standard GbE optical signals. The aggregation node transmits these data flows to multiple gateways, which are called ONTs. Each ONT receives a 100 Mbps (or 1 Gbps) signal in a standard 100 BaseFX (or 1000 BaseLX) format for optical fibre. A core switch interconnects multiple content and service providers over a metropolitan or local area network to deliver data, video and voice services to access network users. A typical connection diagram using active technology in the access network is depicted in Figure 12.

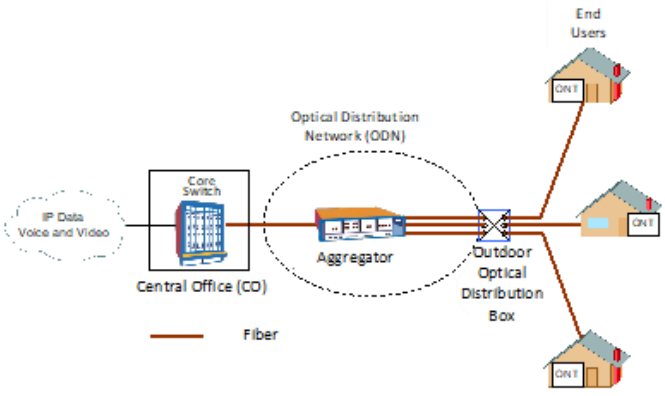


Figure 12: Active ethernet network

The characteristics of the passive and active networks are presented below (Table 8). WDM PON is not considered because it is a solution that will be widely deployed in the near future.

Table 8 : Characteristics of passive and optical networks

	BPON	GPON	EPON	P2P
Standard	ITU-T G.983	ITU-T G.984	IEEE 802.3ah	IEEE 802.3ah/ae
Transmission rate (down)	155.52, 622.08, or 1244 Mbps	2488 Mbps	1244 Mbps	100 Mbps to 1 Gbps per user
Transmission rate (up)	155.52 Mbps or 622.08 Mbps	1244 Mbps or 2488 Mbps	1244 Mbps	100 Mbps to 1Gbps per user
Splitting ratio	Up to 1:32	Up to 1:128	Up to 1:32 or 1:64 using FEC	1:1
Maximum physical distance **OLT-ONU**	10 Km / 20 Km	10 Km / 20 Km	10 Km / 20 Km	5 Km / 70 Km (from the OLT to the active device) 10 Km / 20 Km (from the active device to ONU)

In FTTN/FTTC architecture, a switch / DSLAM which is installed in the outdoor cabin, is connected to the main point of the service provider with a fibre or a pair of fibres and carries the traffic collected from the neighbourhood usually via Gigabit Ethernet. The connection of each end user to the switch in the outdoor cabin is carried out by any physical means other than optical fibre. Usually, the area covered by such an architecture has a radius of up to 1500 meters. In the event that the radius of the service area is less than 300 meters, then the architecture is called Fibre to the Curb. There are different xDSL technologies that can be used in this architecture. Depending on the xDSL technology and the distance, the user enjoys symmetric or asymmetric access data rates that can reach up to 100Mbps. The main characteristics of xDSL technologies are listed below (Table 9).

Table 9*:* xDSL characteristics

	ADSL2+	VDSL	VDSL2
Max rate (downlink)	24 Mbps (<1 Km)	52 Mbps (300m) 26 Mbps (symmetric)	>200 Mbps (<300m)
Max rate (uplink)	3.5 Mbps	16 Mbps 26 Mbps (symmetric)	>200 Mbps (<300m)

Performance vs. Distance (downlink)	18 Mbps (2 Km) 3 Mbps (3 Km) 4 Mbps (4Km) 2 Mbps (5Km)	25 Mbps (800m) 13 Mbps (1.3 Km) 10 Mbps (2 Km)	100 Mbps (500m) 50 Mbps (1 Km) ADSL2+ (>1.6 Km)

VDSL Vectoring is an alternative copper-based technology that is deployed in Fibre to the Cabinet (FTTC) subscriber access lines using VDSL technology, significantly increasing transfer rates (speeds up to 200Mbps) by employing the coordination of line signals for reduction of crosstalk levels (noise cancellation). This technology allows further exploitation of the existing copper network, given that high speeds can be achieved, whereas the cost is relatively lower than deploying a Fibre to the Home (FTTH) network based on Gigabit-capable Passive Optical Networks (GPON) technology.

Indicative prices and power consumption values for the abovementioned components are shown in Table 10 and Table 11 respectively.

Table 10: Prices of FTTx components

Components	Prices (€)
ONT	18
FTTH Cabinet	2000
OLT Rack (10 cards)	1800
Card (192 ports)	600
ODF for racks	1800
ODF (96 fibres)	4500
Splitter: 1:4, 1:8, 1:16, 1:32	8, 10, 20, 30
VDSL DSLAM	880
VDSL Vectoring DSLAM	1400
VDSL & VDSL Vectoring Splitter	250
DSLAM VDSL & VDSL Vectoring line cards/ports	800

DSLAM POTS line cards/ports	800
DSLAM network-facing ports	30

Table 11: Power consumption per "group" of FTTx components

Components Group	Power Consumption (W)
VDSL CPE	10
VDSL Vectoring per street cabinet (48, 96, 144, 192, >192ports)	195, 268, 277, 254, 291
VDSL Central Office per customer	0.96
FTTH P2P per subscriber (PoP size: 50, 100, 250, 500, 750, 1000)	2.3, 2.6, 3.2, 4.5, 7.2, 10
FTTH GPON per subscriber (PoP size: 512, 1024)	1.15, 0.68
FTTH ONT	3.5
FTTH Router	10

3.3.2 Transport Networks

3.3.2.1 Microwave

Microwave (MW) Point-to-Point (PtP) transmission systems that operate in licensed frequency bands within the 6-42 GHz spectrum have been widely used in 3G / 4G backhaul networks, enabling high speed, carrier Ethernet connections over long ranges. Modern MW Internet Protocol (IP) P2P solutions are also becoming part of the 5G ecosystem by offering multi-Gigabit IP capacity, smart network control and provisioning intelligence with the injection of IP/ Multi-Protocol Label Switching (MPLS). A specific flavour of the microwave P2P solutions are the microwave LH (Long Haul) (Figure 13) which can accommodate better the requirements of suburban and rural areas. LH term identifies the MW P2P links having the following characteristics:

- Operate at lower frequency bands: 6L, 6U, 7, 8, 11 GHz, since the bands below 10GHz are essential for long-range transmission due to their superior propagation characteristics.

- Enable reliable wireless connection at longer ranges, usually 25 to 50 Km or more incorporating large parabolic dish antennas, 1.8 to 4.6m.
- Utilise a multitude of RF carriers (carrier aggregation) for increased capacity and enhanced protection. MW LH solutions are widely used around the world and their importance is well recognised.

Figure 13: Microwave Long-Haul connectivity for remote areas

At the LH frequency bands, RF carriers with wider channel sizes of 112 MHz are not available, so operators need to use narrower channels like 28 / 40 /56 MHz. Since the reduction of the channel size reduces the capacity, the solution to increase the capacity is to combine many RF carriers together in one link. This can be achieved by employing carrier aggregation techniques to consolidate in the physical layer multiple (N) carrier transmissions over the same or different frequency band(s) with same or different channel sizes. The RF carriers may have the same polarisation or utilise both the Vertical (V) and Horizontal (H) polarisations on the same carrier along with XPIC functionality.

Regarding the components of the MW transport solutions, legacy MW LH solutions were based on an all-indoor architecture, where both baseband and RF components were installed indoors. These systems were optimised for TDM/STM traffic and are not appropriate for full-packet MW LH. Today, the most capable MW LH architecture is based on split-mount systems consisting of Indoor Units (IDUs) and Outdoor Units (ODUs) that are connected to the IDU modem cards using Intermediate Frequency (IF) cable. The key building blocks of a novel split-mount IP MW LH system for 5G networks along with their advantages are outlined as:

- LH IDUs based on available modular IDUs and equipped with new modem cards with embedded Carrier Aggregation (CA) capability over one IF interface, offer high capacity and advanced functionality to ensure robust RF performance.
- LH compact High Power ODUs supporting multiple carriers per unit enable extended link ranges while saving space on the antenna pole. In addition, a modular composition comprising a band-free base unit and an advanced duplexer greatly simplifies operator's logistics.

3.3.2.2 Fibre

Optical Transport Networks (OTNs) are high-speed communication networks that use optical fibre as the transmission medium. These networks are designed to provide high bandwidth, low latency, and reliable connectivity for various applications, such as data centres, cloud computing, and telecommunication services. The main features of optical transport networks are:

- Multiplexing: OTNs use various multiplexing techniques to combine multiple channels of data into a single optical signal. This allows for efficient use of the available bandwidth and increased data

throughput, making efficient use of any deployed fibre. One technique in particular is Wavelength Division Multiplexing (WDM). WDM is a key technology used in OTNs, which allows multiple channels of data to be transmitted over a single optical fibre by using different wavelengths of light.

- Optical Amplification: OTNs use optical amplifiers to boost the signal strength of optical transmissions over long distances, which helps to overcome the signal attenuation that occurs in fibre optic cables. Amplifiers are as such key elements that need to be considered in optical networks.
- Protection and Restoration: OTNs provide protection and restoration mechanisms to ensure high availability and reliability of the network. These mechanisms include redundancy in the network, automatic switch-over to backup routes, and fast fault detection and recovery.
- Quality of Service (QoS): OTNs support different QoS levels to prioritise traffic and ensure that critical data is transmitted with minimum delay and maximum reliability.
- Carrier-Grade: OTNs are designed to meet carrier-grade standards, which means that they are highly reliable, scalable, and secure. They are built to withstand harsh environmental conditions and provide uninterrupted service. When compared to wireless technologies, the reliability and guarantees OTNs can provide, stands out.

One of the most significant advantages of OTNs is their high bandwidth capacity, which can support multiple terabits per second of data transmission. This makes them ideal for high-capacity applications such as video streaming, cloud computing, and big data analytics. Further, optical fibres can transmit signals over long distances with minimal signal degradation, making them suitable for use in large metropolitan areas, to cover large distances across the land or even across continents. Also, OTNs are highly secure, as optical signals are difficult to intercept or tap. This makes them ideal for applications that require secure data transmission, such as financial transactions and government communications. This feature might be less relevant for rural environments, but it stands out nonetheless.

Overall, it can be said that optical transport networks are critical infrastructure that plays a crucial role in modern communication networks, providing high-speed, high-bandwidth, and reliable connectivity for a wide range of applications. However, it should be noted that this requires a complex infrastructure and a complex maintenance in case of failures. There are other shortcomings too that need to be highlighted.

- Installation Costs: While OTNs offer numerous advantages, the initial cost of installing them can be high, particularly in remote or hard-to-reach areas.
- Vulnerability to Physical Damage: Optical fibres are vulnerable to physical damage from environmental factors such as storms or construction activities. This can result in signal degradation or complete signal loss.
- Limited Reach: While optical fibres can transmit signals over long distances, their reach is still limited. For very long distances, signal regeneration equipment is needed, which can increase the overall cost of the network.
- Complexity: OTNs are complex and require specialised knowledge and skills to design, build, and maintain. This can make it challenging for smaller organisations or non-experts to implement them effectively.

3.3.2.3 Satellite IoT

Satellite IoT backhaul configuration is a connectivity solution composed by a gateway which collects data from the customer's sensors and send them to the satellite via an integrated IoT terminal (Figure 14).

The components of Satellite IoT Backhaul networks include:

1. User terminal: composed of a satellite modem and an antenna (form factor will depend on the specific transmit/receive frequency of the modem). The terminal is used to transmit or receive data to the user endpoint via the satellite constellation. In the backhaul configuration the terminal is normally integrated in the gateway.
2. Gateway: typically LoRa or NB-IoT base station which retrieves data from the user sensors and sends them to the satellite through the integrated IoT terminal. A pre-processing step might be included before passing the data to the terminal, especially in the case of narrow-band communication where only a small amount of data can be sent.
3. Satellite: The satellite is a man-made object placed in orbit around the earth that receives and retransmits signals from the gateway to the ground station, following a bent-pipe architecture. The satellite M2M payload operates in a specific frequency band, such as UHF, VHF, L-band, S-band, etc.
4. Ground Station/Gateway: The ground station or a gateway is the facility on the ground that communicates with the satellite. It receives signals from the satellite, processes them, and sends them to the backend of the company providing the connectivity solution.
5. Cloud platform: The user can retrieve or upload data to or from the terminal thanks to a web portal or APIs.

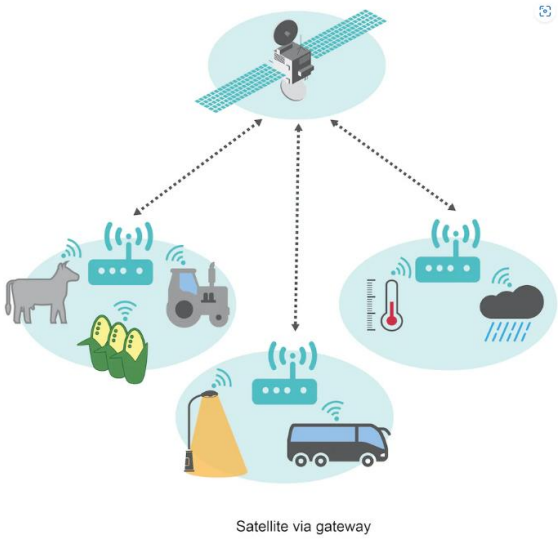


Figure 14: Satellite IoT Backhaul configuration [22]

Satellite IoT Backhaul latency (from terminal to customer portal) will depend on the same factors reported before in the satellite IoT access system, as from the satellite network perspective there is no difference in using the IoT terminal as access or backhaul point. Therefore, there is no difference between access and backhaul latency (from terminal to customer portal).

In a backhaul configuration, the IoT terminal has to transmit data collected through the gateway from multiple sensors rather than one (as in the access configuration), therefore the quantity of data to be transmitted is higher. As a consequence, data will remain a higher amount of time within the queue of the terminal. This waiting time depends on the overall amount of data to be transmitted and the maximum message size of the IoT terminal. This is an important factor to be considered when deciding to select a backhaul configuration. A way to overcome the waiting time in the terminal queue is to increase the amount of IoT terminals in the gateway, increasing the data rate. However, increasing the number of devices reduces the network capacity, therefore this solution must be carefully evaluated together with an expert in the Satellite IoT sector.

Power consumption of the IoT terminal in the backhaul configuration is typically higher than the one in the access configuration. Indeed, as the amount of data to be transmitted is higher the terminal will have to transmit more often than in the direct access configuration which will increase its own power consumption.

The advantages of this configuration are the following:

- Reducing the number of IoT terminals needed, indeed a single IoT terminal might serve multiple sensors thanks to the gateway.
- The possibility to connect sensors which do not have direct visibility with the satellite, e.g. a farmer that has an indoor sensor cannot communicate with the satellite due to the building impeding the transmission of data to the satellite.

In satellite IoT access configuration the user is normally forced to have the terminal close to the asset, which is not always the best configuration in terms of transmission conditions. Moving the terminal to a better location (i.e on top of a mountain/hill) might help the transmission of data.

3.3.2.4 Satellite Broadband

Satellite broadband backhaul refers to the use of satellite links to connect remote or hard-to-reach areas to the Internet or other communication networks, as illustrated in Figure 15. In many cases, these areas are not serviced by traditional terrestrial networks such as fibre-optic or copper cables, and satellite broadband backhaul can provide a cost-effective solution for providing connectivity. The link serves as an intermediate between the main network and the small networks used for distribution to other smaller channels. While satellite access is focused on providing end-user connectivity, satellite broadband backhaul is focused on connecting different parts of the network together.

Satellite broadband backhaul typically involves using a LEO/GEO/MEO satellite, which transmits data to and from the remote area. The satellite is equipped with a transponder, which receives the data signal from a ground station, amplifies it, and then retransmits it to the remote area. For example, OneWeb provides a mobile backhaul allowing operators to expand LTE and 5G networks into underserved or unconnected communities. Figure 16 shows OneWeb providing backhaul coverage to regions in Europe with the ground station located in the UK.

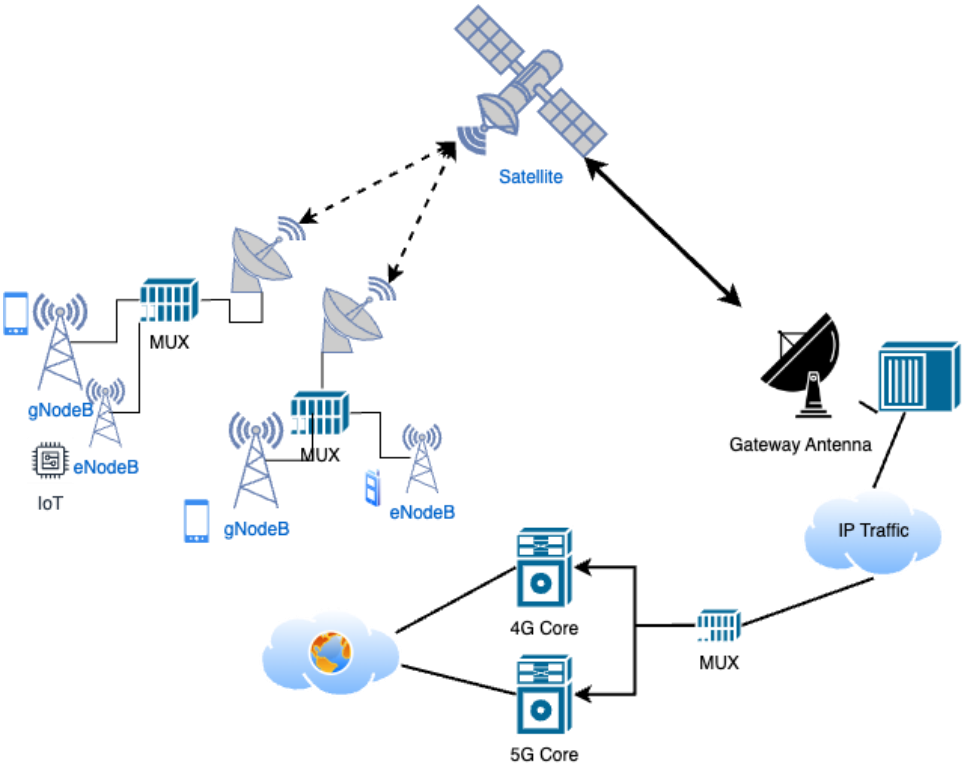


Figure 15: Illustration of satellite backhaul link

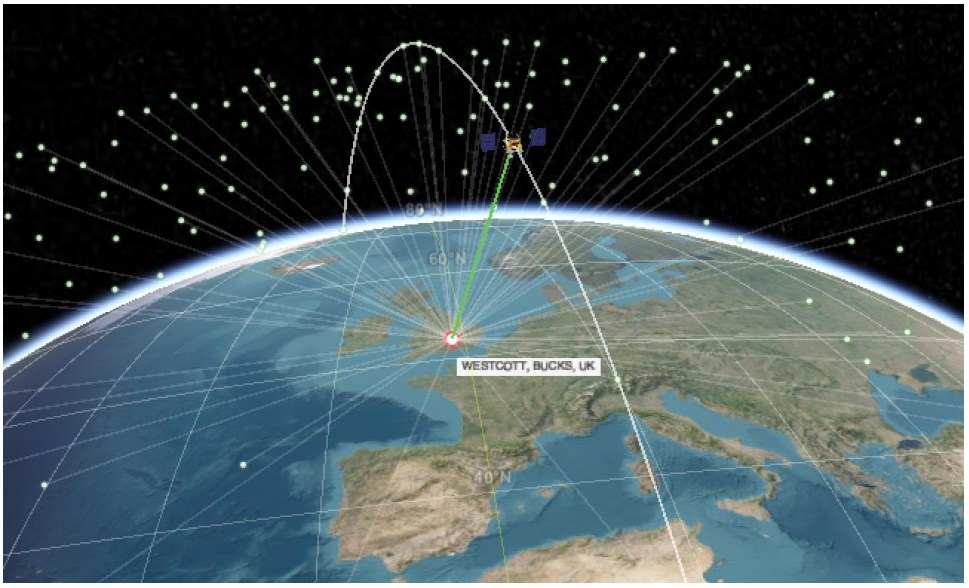


Figure 16: Oneweb constellation and satellite connected to a gateway in the UK

3.4 Signal Propagation Constraints

Signal propagation constraints are essential considerations in the field of communications because they directly impact the quality and reliability of transmitting and receiving signals over various media. Understanding and managing these constraints is crucial for designing efficient communication systems and ensuring effective signal transmission. Signal propagation constraints arise from the physical properties and characteristics of the transmission medium, such as cables, airwaves, or optical fibres, through which signals travel. These constraints can affect signal strength, clarity, speed, and coverage, and they must be carefully addressed to achieve reliable and high-quality communication.

In wired communication systems, such as those utilising copper or fibre-optic cables, signal propagation constraints primarily involve issues like signal attenuation (weakening of the signal over distance), dispersion (spreading of the signal), and noise interference (unwanted signals or disturbances). These constraints become more pronounced with longer distances and higher data rates. Wireless communication systems, on the other hand, face additional challenges due to signal propagation through the atmosphere. Factors such as free space path loss (attenuation of signal strength with distance), interference from other wireless devices, multi-path fading (signal reflections and interference from obstacles), and environmental conditions (such as weather) can significantly impact signal quality and coverage.

Additionally, propagation constraints influence the design and deployment of communication networks. They determine the spacing and location of network infrastructure, such as cell towers, base stations, or relay stations, to ensure adequate signal coverage and minimise signal degradation. These considerations are vital for providing reliable connectivity to users across various environments and geographical areas. For the XGain project, taking these limitations into account is important, since the KFT needs to perform the task of finding the proper dimensions of a deployment and the signal propagation constraints might be relevant.

From a technical perspective, in telecommunications, the link budget is an accounting of all of the power gains and losses that a communication signal experiences, from a transmitter, through a communication medium such as air, cable, waveguide, or optical fibre, to the receiver. Link budget is a design aid, calculated during the design of a communication system to determine the received power, to ensure that the information is received intelligibly with an adequate signal-to-noise ratio.

A simple **link budget equation** looks like this:

$$\text{Received power (dBm)} = \text{transmitted power (dBm)} + \text{gains (dB)} - \text{losses (dB)}$$

While each communication medium (wireless or wireline) imposes its own loss mechanisms on the signals, we will focus on the wireless systems, since the loss factors that come into play are dependent on external factors such as the weather conditions of the respective location, which is not the case in wireline one. In wireless systems, transmission loss is accounted principally by the free-space loss. Free-space transmission assumes clear weather, no absorption or reflection of energy by nearby objects, and the respective equation of Free Space Path Loss (FSPL) is calculated as follows:

$$\text{FSPL dB} = 92.4 + 20 \log f + 20 \log R$$

Where f is the frequency in Gigahertz (GHz) and R is the Line-of-Sight (LoS) range between the antennas in kilometres.

Other types of losses are as follows:

- **Atmospheric Gaseous Losses**

Transmission losses occur when wireless signals that are traveling through the atmosphere are absorbed by molecules of oxygen, water vapour, and other gaseous atmospheric constituents. These losses are greater at certain frequencies, coinciding with the mechanical resonant frequencies of the gas molecules. For systems

operating below 10 GHz, the atmospheric absorption and precipitation play a secondary role, and with many systems these issues are neglected. Figure 17 gives qualitative data on gaseous losses. It shows several peaks that occur due to absorption of the radio signal by water vapour (H_2O) and oxygen (O_2). For current technologies, the important absorption peaks occur at 24 and 60 GHz. It is worth mentioning that other gases display resonant lines as well, but because of their low density in the atmosphere, they have negligible effect on propagation.

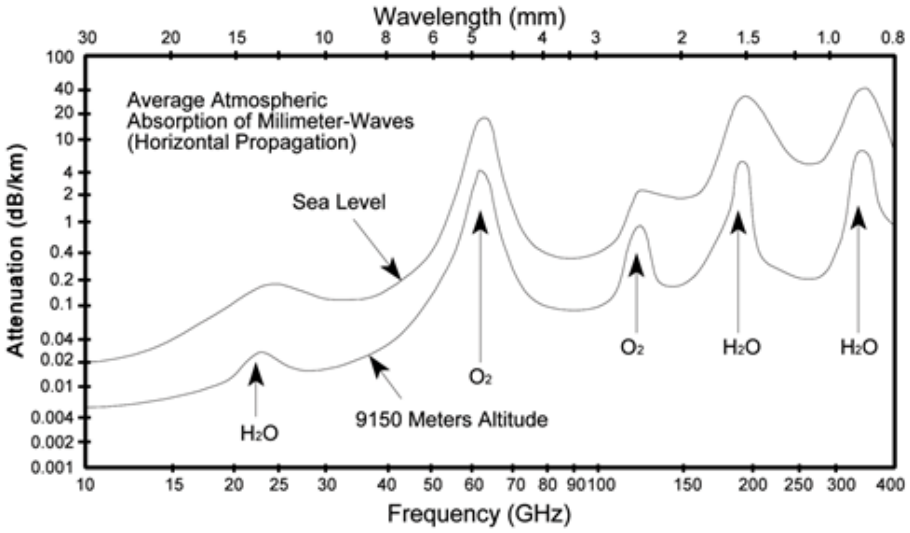


Figure 17: Attenuation of signals in our atmosphere by H_2O and O_2 [33]

- **Rain Losses/Fade**

Rain fade refers primarily to the absorption of a microwave RF signal by atmospheric rain, snow, or ice, and this kind of losses is especially prevalent at frequencies above 10 GHz. It does not need to be raining at a location for it to be affected by rain fade, since the signal may pass through precipitation many miles away. Specific rain attenuation model by the International Telecommunication Union (ITU) recommendation P.838-3 [34] provides a method to calculate the specific attenuation due to rain. Based on this model, Figure 18 shows the attenuation per kilometre as a function of different rain rates.

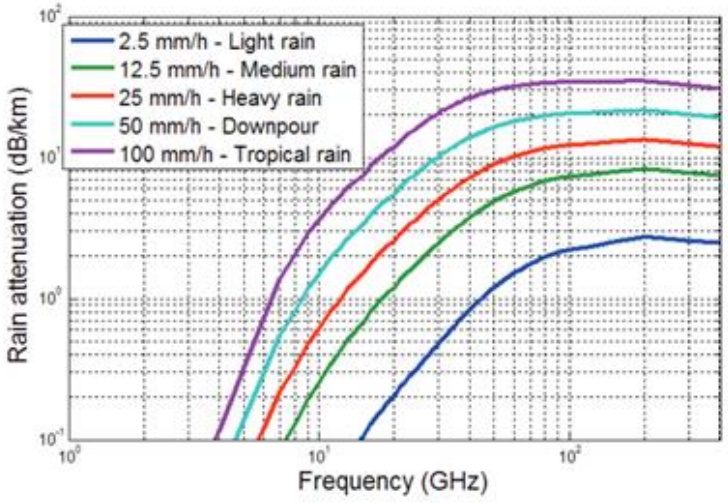


Figure 18: The intensity of precipitation can affect atmospheric attenuation [35]

- **Foliage Losses**

Radio waves that travel within foliated environment such as forest medium, tree line etc., experience power losses in several ways such as reflection, diffraction and multipath dispersion as depicted in Figure 19. However, the actual foliage attenuation varies depending on various factors such as the foliage thickness and humidity. The signal attenuation also depends on a range of factors such as tree types, whether trees are in-leaf or out-of-leaf, whether trees are dry or wet, size of leaves, the distribution of leaves, leaf density, branches and trunks, wind speed, dielectric constant, height of the tree relative to the antenna heights, frequency and path length through foliage.

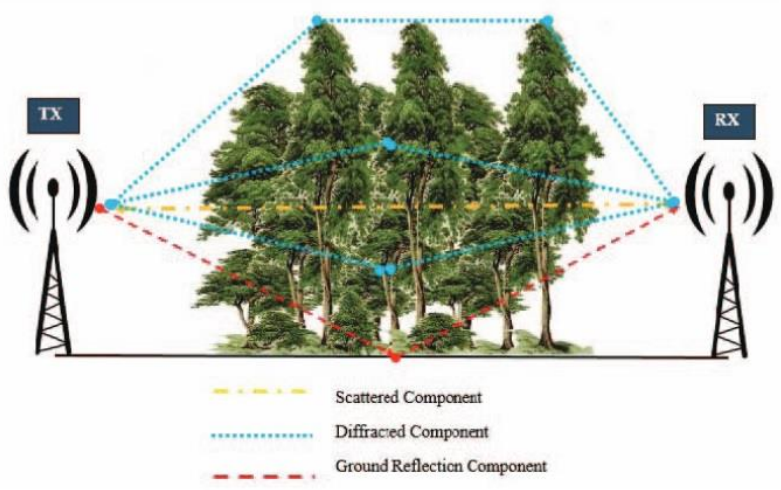


Figure 19: Signal propagation through foliage [36]

Many studies have been carried out to characterise and model the effects of vegetation experimentally. They have been reviewed and summarised into several well-known through-vegetation loss models, such as Weissberger's modified exponential decay model [37] COST235 model [38], ITU Recommendation (ITU-R) [39] etc. Regardless of the approach, it is evident that high frequencies and high depth of vegetation result to greater attenuation of electromagnetic waves.

- **Terrain constraints**

Irregular rural terrain (i.e., hills, rocks, mountains) has also an effect on the propagation of wireless signals, creating different paths (direct, ground reflected, terrain diffracted and scattered, troposphere scattered signals) as depicted in Figure 20. If there is no LoS path between the transmitter and the receiver, the signal may still reach the receiver via reflections from objects in proximity to the receiver or via diffraction or bending.

Normally, for non-LoS paths, the greatest contribution at the receiver is reflected power. Reflections and the associated amount of signal diffusion are strongly dependent on the reflectivity of these reflecting materials. It is also worth mentioning that higher frequencies cause higher diffusion of the signal and less specular reflection.

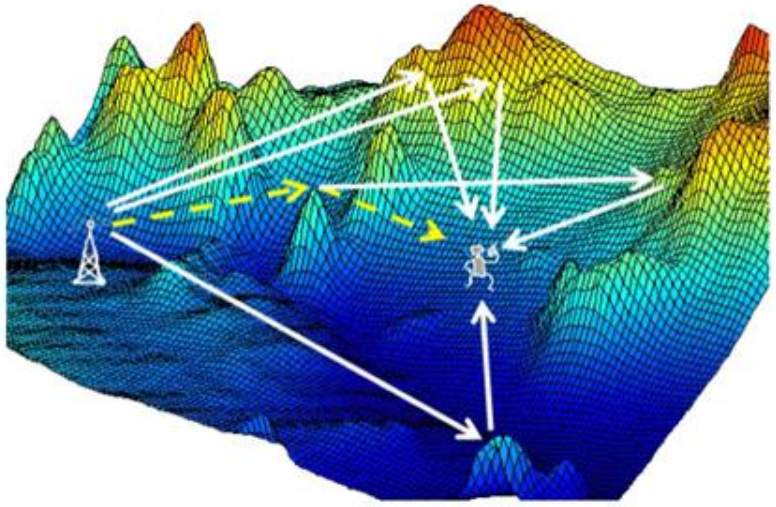


Figure 20: Propagation paths in a rural environment [40]

As it has already been stated, technology options are determined by population size, geographic area and topographical features. Small areas with high population density may be served by localised technologies such as Wi-Fi, while on the other hand, wide areas may need cellular network coverage or direct satellite connectivity. Similarly, mountainous terrain with limited line-of-sight options between radios may require more complex network design (circumvent the irregular terrain employing multiple links of LoS systems) and/or the use of non-LoS technologies. As an example, terrain poses a problem for the terminal to satellite transmission when the terminal field of view is blocked by an object, as depicted in Figure 21. This is especially valid for LEO constellation, where the communication between terminal and satellite is established during a satellite pass. As soon as the satellite to terminal visibility link is blocked by an object, the communication is blocked and the terminal cannot transmit more data.

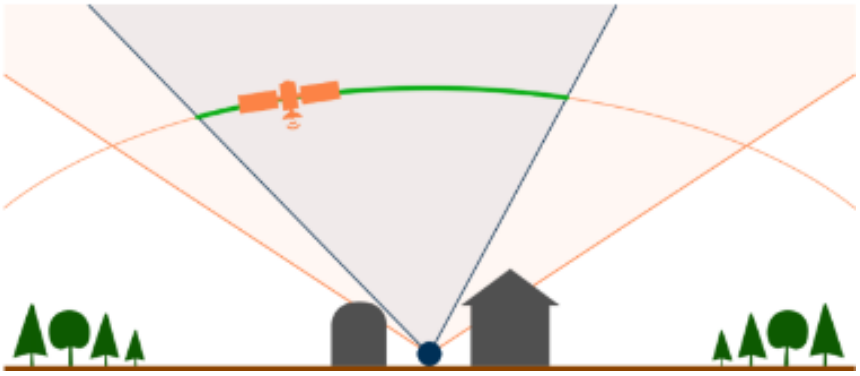


Figure 21: Satellite constellations and blockage from objects

Table 12 provides a brief summary of indicative technology options for given combinations of population size and topography. It is evident that areas with flat terrain are the most versatile in terms of the technology solutions available for deployment, given the reliance of some technologies on LoS. Wireline technologies have also been taken into account, though not technically constrained by terrain peculiarities, but financially.

Table 12: Terrain constraints and communication technologies

Terrain Constraints	Small geographic area, flat terrain	Small geographic area, mountainous terrain	Large geographic area, flat terrain	Large geographic area, mountainous terrain
Relative Thresholds	< 10 square Km, line of sight possible across most of the terrain	< 10 square Km, non-line of sight across most of the terrain	> 10 square Km, line of sight possible across most of the terrain	> 10 square Km, non-line of sight across most of the terrain
Potential Technologies	Wi-Fi, Cellular, Fibre, Copper, Satellite, Microwave transport	Cellular, Wi-Fi, Satellite, Fibre, Copper	Cellular, Satellite, Fibre, Microwave transport	Cellular, Satellite, Fibre

4 Edge Processing Enablers

Computing capacities play a crucial role in applications deployed in rural environments due to their relevance in overcoming unique challenges and unlocking opportunities. Robust computing capabilities enable the processing and analysis of data collected from various sources, empowering decision-making, resource management, and service provision tailored to rural contexts. They facilitate the deployment of advanced technologies like precision agriculture or telemedicine, which can enhance productivity, improve healthcare access, optimise resource allocation, and bridge the digital divide between rural and urban areas. By leveraging computing capacities, rural communities can harness the benefits of digital innovation and drive sustainable development in their regions. In this section we provide an overview of different computing and in particular edge processing enablers, which serves as a basis for the assessment done by the KFT when proposing technology mixes for rural services.

4.1 Characterisation of Edge Processing Technologies

The aim of edge computing is to lower latency and enhance efficiency by handling data near the source of data, instead of depending on centralised data centres or cloud computing solutions [41]. It leverages technology that facilitates data analysis, processing, and storage close to the edge devices, such as IoT devices, cameras, and other sensors. Edge computing is particularly useful for real-time applications, such as self-driving cars, intelligent homes, and industrial automation, where quick responses are crucial. Edge intelligence also enables offline processing, reducing the need for remote data centres and offering greater privacy and security for sensitive data.

4.1.1 Compute Continuum Overview

The term compute continuum refers to the idea that computing resources are adaptable and can be allocated effortlessly to meet the requirements of various tasks, extending from edge devices such as embedded systems and IoT to cloud computing and data centres. The objective of the compute continuum is to provide a smooth and uninterrupted supply of computing power, regardless of the location or device being used, to fulfil the needs of users and applications in real-time [42][43]. The compute continuum can be defined as a continuum of edge zones including the following ones: Extreme Edge, Far Edge, Near Edge and Cloud (Figure 22) [44].

Cloud: This zone comprises the central cloud servers that store and process vast amounts of data. Cloud computing is typically used for high-performance computing tasks such as machine learning and big data analytics. The cloud zone is well-suited for applications that require massive computing power and storage capacity.

Near Edge: This zone consists of regional servers and data centres that are geographically distributed and located closer to the source of data generation. Near Edge computing could be used to support real-time applications that require low latency and high bandwidth. Near Edge servers could be used for all applications that cloud servers are currently used for.

Far Edge: The Far Edge zone comprises Telecom base stations and enterprise premises, which are located closer to the data source than Near Edge servers. Far Edge computing could be used for applications that require even lower latency and higher bandwidth, such as autonomous vehicles, augmented reality, and virtual reality. Far Edge servers could be used for applications such as video streaming, asset tracking, environmental monitoring, and smart cities.

Extreme Edge: The Extreme Edge zone comprises unmanned vehicles (UGVs, UAVs) and manned vehicles and wearable devices such as smartphones, tablets and sensors. These devices are the closest to the source of data generation, but have limited computing power and storage capacity, which they will make up for in greater numbers. A big difference from the previous zones is that the hardware utilised in the Extreme Edge will mostly be independently owned and as such, the compute power will have to be leased from them. Extreme Edge computing could be used for applications such as personal health monitoring, smart homes, and mobile gaming.

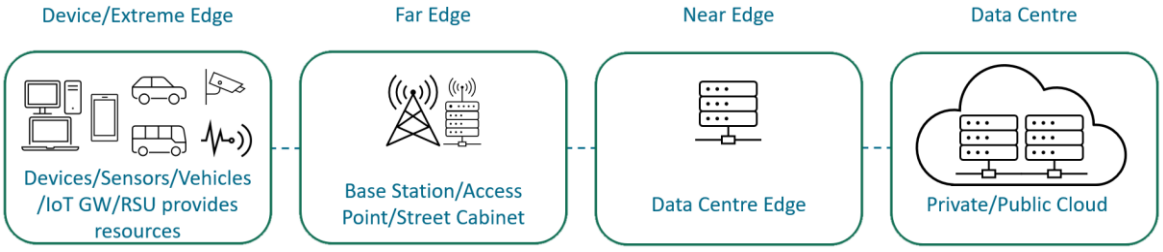


Figure 22: Different types of compute continuum deployments

The main technologies that are used in compute continuum deployments are as follows:

Server types: Compute continuum solutions can use a variety of server types, including traditional physical servers, Virtual Machines (VMs), and containers. Physical servers provide the highest performance and can be customised for specific workloads, but they are also the most expensive and difficult to manage. VMs and containers are more lightweight and easier to manage, but they also offer less performance, since you're splitting the performance of a single server between multiple containers or VMs.

Accelerators: Compute continuum solutions often use hardware accelerators to speed up specific workloads. These can include Graphics Processing Units (GPUs), Field-Programmable Gate Arrays (FPGAs), and Application-Specific Integrated Circuits (ASICs). Accelerators can improve the performance of AI/ML workloads, video encoding/decoding, and other computationally intensive tasks.

Device types: Compute continuum solutions can run on a variety of devices, including smartphones, tablets, laptops, and IoT devices. Each device type has its own strengths and weaknesses, and the choice of device will depend on the specific requirements of the application.

Virtualisation and container solutions: Virtualisation and container solutions can be used to create isolated environments for running applications. This can help to improve security and manageability and can also make it easier to move applications between different edge computing nodes. Popular virtualisation solutions include VMware ESXi and Microsoft Hyper-V, while popular container solutions include Docker and Kubernetes.

Software-Defined Networking (SDN) and Network Functions Virtualisation (NFV): SDN and NFV are two technologies that can be used to improve the flexibility and scalability of the compute continuum. SDN separates the network control plane from the data plane, making it easier to configure and manage the network. NFV replaces traditional network hardware with software-based virtual appliances, making it easier to add new network functions and services.

4.1.2 Generic Technological Characteristics of Edge Processing Enablers

Several technologies enable edge processing, including:

1. Edge devices, such as sensors, actuators, and gateways that collect and process data at the edge.

2. Edge computing platforms, such as edge servers and edge gateways, that provide computing resources and software infrastructure for edge applications.
3. Low-latency networking technologies, such as 5G and Wi-Fi 6, that enable high-speed and reliable communication between edge devices and the cloud.
4. Machine Learning (ML) and Artificial Intelligence (AI) technologies, such as edge-based inference engines, that allow real-time analytics and decision-making at the edge.
5. Security and privacy technologies, such as blockchain and secure enclaves that protect sensitive data and prevent unauthorised access or tampering.

The generic technological characteristics of edge processing enablers are:

1. **Low-latency and high-bandwidth communication**: Edge processing enablers rely on fast and reliable communication networks that can support real-time data processing and analysis.
2. **Distributed computing**: Edge processing enablers are designed to support distributed computing architectures that can perform computations across multiple devices and platforms.
3. **Scalability**: Edge processing enablers are designed to scale up or down as needed to meet changing demands for processing power and storage.
4. **Heterogeneity**: Edge processing enablers support a wide range of devices, platforms, and operating systems, allowing for seamless integration of different technologies.
5. **Security and privacy**: Edge processing enablers employ robust security and privacy mechanisms to protect data and prevent unauthorised access or tampering.
6. **Energy efficiency**: Edge processing enablers are designed to be energy-efficient, using low-power devices and optimised algorithms to minimise power consumption and maximise battery life.

The technical characteristics of different zones of compute continuum vary based on their distance from the source of data, their computational capabilities, and their connectivity.

- Extreme edge devices typically have low processing power, limited memory, and low energy consumption. They may use specialised hardware or software designed to process a limited amount of data locally, or perform simple data filtering, aggregation, or pre-processing. Extreme edge devices may use wireless communication protocols with limited bandwidth and range or may be directly connected to the device they are monitoring or controlling.
- Far edge devices have higher processing power, more memory, and more connectivity options than extreme edge devices. They may use specialised hardware or software, such as FPGAs or GPUs, to perform more complex computations, such as machine learning. Far edge devices may use wired or wireless communication protocols with higher bandwidth and range, and may be connected to multiple sensors or devices.
- Near edge devices have even higher processing power, more memory, and more connectivity options than far edge devices. They may be equipped with high-performance processors, such as multicore Central Processing Units (CPUs) or GPUs, and may use specialised software frameworks, such as TensorFlow or PyTorch, to perform real-time data processing, analysis, and decision-making. Near edge devices may use wired or wireless communication protocols with low latency and high reliability, and may be connected to high-speed local networks.

- Cloud computing infrastructure typically has the highest processing power, memory, and connectivity options of all the edge computing points. Cloud computing can scale horizontally to process large amounts of data in parallel, and can provide access to a vast array of software tools and services, such as big data analytics, artificial intelligence, and high-performance computing. Cloud computing may use wired or wireless communication protocols with high bandwidth and global reach, and may be accessed through the Internet or private networks.

In summary, the differences in technical characteristics between the different zones relate to their processing power, memory, connectivity, and specialised hardware and software. Extreme edge devices are typically low-powered and limited, while cloud computing infrastructure is highly scalable and flexible, with high processing power and connectivity. Near and far edge devices fall in between, with varying levels of processing power, memory, and connectivity, depending on their specific Use Case and application. The generic technological characteristics of edge processing enablers are further analysed in the following paragraphs.

Security

The different zones of the compute continuum (Extreme Edge, Far Edge, Near Edge, and Cloud) can differ significantly in terms of security considerations. Here are some of the keyways in which they may differ:

- **Extreme edge devices may have limited security measures** due to their low processing power and memory. They may lack the resources to implement robust security protocols or encryption, leaving them vulnerable to attacks such as data tampering or theft.
- **Far edge devices may have more security measures in place** but may still be vulnerable to attacks due to their proximity to the edge of the network. They may be exposed to a range of potential threats, including malware, phishing, or social engineering attacks.
- **Near edge devices may have more sophisticated security measures in place** due to their increased processing power and memory. They may be able to implement more advanced encryption and authentication protocols and may be able to detect and respond to potential threats in real-time.
- **Cloud computing infrastructure may have the most advanced security measures in place** due to the large number of potential threats it faces. Cloud providers typically use a range of security measures, such as firewalls, intrusion detection and prevention systems, and data encryption, to protect their infrastructure and their customers' data.

Overall, the security methods are device-specific according to their computing capabilities. Data processing and storage can be done at the origin, minimising the sensitive data sent over the Internet. In addition, the security considerations for edge computing will also depend on the specific Use Case, the type of data being processed and stored, and the level of risk associated with potential security breaches. Regardless of the point of edge computing being used, it is important to consider security at every stage of the process, from the initial data collection to the final analysis and decision-making. Appropriate security measures should be implemented at each stage, and regular testing and auditing should be carried out to ensure that the system is secure and resilient.

Scalability

Scalability is another important consideration for edge computing. The edge processing must consider the diversity of connected devices. The types, capabilities and number of the devices can be chosen according to the

needs of the specific task. Moreover, the various zones of edge computing can differ significantly in their ability to scale. Here are some ways in which they may differ:

- Extreme edge devices typically have very limited processing power and memory, which can make it difficult to scale up their capabilities. As a result, extreme edge computing is often used for simple, low-level data processing tasks that can be easily handled by these devices.
- Far edge devices may have more processing power and memory than extreme edge devices, which can make them more scalable. However, they may still face limitations in terms of their ability to handle large volumes of data or complex processing tasks.
- Near edge devices typically have the most processing power and memory of any edge computing devices, which makes them the most scalable. They can handle more complex tasks and larger volumes of data than extreme or far edge devices, and can be used for a wider range of applications.
- Cloud computing infrastructure is typically the most scalable of any edge computing point, as it can provide virtually unlimited processing power and storage capacity. Cloud providers can easily add more resources to their infrastructure as needed, making it an ideal choice for applications that require high scalability.

Overall, the scalability of edge computing will depend on the specific Use Case, the amount of data being processed, and the level of processing power and memory required. When planning an edge computing system, it is important to consider scalability from the outset and to choose the most appropriate point of edge computing for the specific Use Case. This can help ensure that the system can handle growing volumes of data and processing requirements over time.

Reliability

Reliability is another important consideration for the compute continuum, and the different zones can differ significantly in their reliability characteristics. Here are some ways in which they may differ:

- **Extreme edge devices may be less reliable** than other zones of compute continuum due to their limited processing power and memory, which can make them more prone to errors or failures. Additionally, extreme edge devices may be subject to harsh environmental conditions or physical damage, which can further reduce their reliability.
- **Far edge devices may be more reliable than extreme edge devices** due to their increased processing power and memory, which can make them better equipped to handle a wider range of tasks. However, they may still be subject to environmental factors or physical damage that can impact their reliability.
- **Near edge devices are typically the most reliable of any edge computing point** due to their advanced processing power and memory, which makes them better able to handle complex tasks and larger volumes of data. They may also be designed with more robust physical and environmental protections, which can further enhance their reliability.
- **Cloud computing infrastructure is typically highly reliable**, as cloud providers are able to distribute data and processing tasks across multiple servers and data centres. This helps to ensure that the system can continue to operate even in the event of hardware failures or other issues. Cloud providers may also implement redundancy measures such as backup power supplies or data backups to further enhance their reliability.

Overall, the reliability of an edge computing system will depend on a range of factors, including the specific Use Case, the hardware and software being used, and the environmental conditions in which the system operates. When planning an edge computing system, it is important to consider reliability from the outset and to choose the most appropriate point of edge computing for the specific Use Case. This can help ensure that the system is reliable and resilient, and can continue to operate even in the face of potential failures or disruptions.

Computing Speed

Computing speed (how fast a processing task can be completed) is a crucial consideration for compute continuum and the different zones. Here are some ways in which the different points of edge computing can differ in terms of speed:

- **Extreme edge devices may be limited in their processing speed**, due to their lower processing power and memory. However, because they are located closest to the source of data, they may be able to provide faster response times for certain applications. For example, an extreme edge device may be able to quickly detect a specific event or anomaly in a data stream and trigger an immediate action.
- **Far edge devices may be faster than extreme edge devices due to their increased processing power and memory**. They may be able to handle more complex data processing tasks and provide faster response times for certain applications. For example, a far edge device may be able to perform real-time analysis of video data, from a surveillance camera and alert security personnel, to a potential threat.
- **Near edge devices are typically the fastest of any edge computing point due to their advanced processing power and memory**. They may be able to handle even more complex data processing tasks and provide even faster response times than far edge devices. For example, a near edge device may be able to perform real-time analysis of data from multiple sensors and make decisions based on that data in a matter of milliseconds.
- **Cloud computing infrastructure may have slower response times than edge computing points due to the additional network latency involved in sending data to and from the cloud.** However, cloud providers may be able to make up for this by using advanced caching and data distribution techniques, and by providing high-speed connectivity to their data centres. Additionally, cloud computing may be able to provide faster response times for certain applications that do not require immediate action or processing.

Overall, the speed of an edge computing system will depend on a range of factors, including the specific Use Case, the hardware and software being used, and the network connectivity available. When planning an edge computing system, it is important to consider speed from the outset and to choose the most appropriate compute continuum zone for the specific Use Case. This can help to ensure that the system is able to process data and make decisions in real-time, and provide the fastest possible response times for the application.

Latency

The latency (the time it takes for the computing task to be sent and the result to be received) of each point of edge computing can vary widely depending on a number of factors, such as the specific hardware and software being used, the quality and speed of the network connection, and the complexity of the data processing tasks being performed.

As a general rule, extreme edge devices are likely to have the lowest latency, with far edge and near edge devices having slightly higher latency, and cloud computing infrastructure having the highest latency.

However, it is difficult to provide specific latency figures for each compute continuum zone, as the actual latency will depend on many factors that are unique to each system and Use Case.

That being said, some estimates suggest that extreme edge devices can have latencies which are as low as a few microseconds, while far edge and near edge devices may have latencies in the range of a few milliseconds to tens of milliseconds. Cloud computing infrastructure can have latencies in the range of tens of milliseconds to hundreds of milliseconds, or even higher.

It's important to note that these are only rough estimates, and actual latency figures can vary widely depending on the specific system and Use Case. It's important to carefully evaluate the latency requirements of a given application and choose the most appropriate compute continuum zone based on those requirements.

Efficiency

The different compute continuum zones can differ in terms of efficiency, particularly in terms of their energy consumption and processing power.

- **Extreme edge devices are typically the most efficient in terms of energy consumption**, as they are designed to perform basic data processing tasks with minimal power consumption.
- **Far edge devices may consume more power than extreme edge devices**, as they may need to perform more complex data processing tasks. However, they still typically use low-power hardware and software architectures, and may be designed to operate for long periods of time without human intervention.
- **Near edge devices may consume more power than far edge devices**, as they are designed to perform more demanding data processing tasks. They may use more powerful processors and more memory than far edge devices, which can increase their power consumption.
- **Cloud computing infrastructure typically consumes the most power of any edge computing point**, as it requires large data centres with extensive cooling and power distribution systems. However, cloud providers may be able to achieve high levels of efficiency through the use of advanced power management techniques and the optimisation of their data centre designs.

In terms of processing power, extreme edge devices may be limited in their ability to handle complex data processing tasks, due to their limited processing power and memory. Far and near edge devices may be able to handle more demanding tasks but may still be limited by the available processing power and memory. Cloud computing infrastructure typically offers the most processing power of any edge computing point, due to the large-scale computing resources available in the data centre.

Overall, the efficiency of an edge computing system will depend on many factors, including the specific hardware and software being used, the workload being processed, and the overall design of the system. When designing an edge computing system, it's important to carefully consider the efficiency of each compute continuum zone and choose the most appropriate zone based on the specific requirements of the Use Case.

Table 13 summarises the main general characteristics of different compute continuum zone.

Table 13: General characteristics of different compute continuum zones

Edge zone	Security	Scalability	Reliability	End-to-end Latency	Efficiency
Extreme edge	Limited control over the edge devices, it is not suitable for	The number and type of the devices can be quickly chosen	Availability depends on consumers. Decentralised architecture.	< 10 ms	The specific processing can be done using devices already in

	sensitive data processing.	according to the demands of the task.			circulation, saving network bandwidth and data centre computing resources. Increased energy efficiency – Low power consumption.
Far edge	Standardised and controllable security requirements for the third-party network apps deployed near the base station.	The processing capability (hardware and software) can be upgraded in the cell base station. The stations can be added.	In case of the malfunction of the cell base station the data traffic can be diverted to the second nearest cell with the same functionality.	< 25 ms	The cloud computing experience is brought close to the client. More power consumption than extreme edge but with still increased efficiency.
Near edge	Similar to the cloud data centre	Similar to the cloud data centre	Similar to the cloud data centre	< 50 ms	The regional data centres serve the local data processing needs. More power consumption that far edge accompanied with less energy efficiency.
Cloud Data Centre	Centrally managed security	Built with the scalability in mind	The workload of the data centre can be diverted to the mirrored centre in case of failure. Regular hot swapping of the hardware and updates of the software	< 100 ms	Efficiency achieved by optimising the hardware and software for the specific functions. High level of energy efficiency due to advanced power management techniques

4.1.3 Virtualisation Solutions for Edge Computing

Virtualisation is an important technology to consider in Edge scenarios, as it enables multiple virtual machines/containers to run on a single physical device, providing the flexibility and scalability needed to support the demands of many edge Use Cases.

The following are different virtualisation techniques available:

- Container virtualisation: Container virtualisation solutions such as Docker [45] and Kubernetes [46] allow multiple containers to run on a single host, providing a lightweight and efficient virtualisation solution. Kubernetes is the de-facto orchestration solution.
- Hypervisor-based virtualisation: Hypervisor-based virtualisation solutions such as VMware vSphere [47], Microsoft Hyper-V, and OpenStack [48] enable multiple Virtual Machines (VM) to run on a single physical host.
- Unikernels: Unikernels [49] are a new type of virtualisation technology that enable a single application to run in its own virtual machine, providing a highly efficient and secure virtualisation solution.

Containers are considered to be the best solution for Edge environments. Why are they more suitable than VMs and unikernels? The answer lies in their combination of lightweight resource utilisation, ease of deployment, and compatibility with existing applications and infrastructure.

Compared to VMs, containers offer a smaller footprint, faster launch times, and better resource utilisation, making them ideal for edge computing Use Cases where latency and efficiency are critical [50]. Additionally, containers can be easily deployed and managed on edge devices, and they can run existing applications and software components without modification.

Compared to unikernels, containers offer a more familiar and widely-used development and deployment model, as well as compatibility with existing application and infrastructure components [51]. Unikernels, while offering improved security and resource utilisation, require significant changes to existing applications and can be more difficult to deploy and manage.

4.2 Edge Processing Technologies in Depth

In this section we will dive into the technical details and specific examples of each of the three edge zones identified in the previous section. The details provided include information about the computing capacities, the power consumption and expected buying costs for different models of hardware in each zone.

4.2.1 Extreme Edge Processing Enablers

IoT devices and equipment very close to where data is generated in general, should not just be about data collection, but they should bring intelligence to the extreme edge of the network to derive value from the data for quicker decision support [52]. In addition, and according to [53], communication is about 5 times more expensive in terms of energy compared to compute operations. Lightweight processing of data at the source including simple data filtering can thus contribute to reducing energy consumption as a result of less data being communicated over the network infrastructure. In this part of the compute continuum, we are focusing on processing platforms that can be mounted on small devices, such as drones.

A Single-Board Computer (SBC) is a complete computer built on a single circuit board, with microprocessor(s), memory, Input/Output (I/O) and other features required of a functional computer and can serve as an extreme edge processing solution. Indicative single-board computers are depicted in Figure 23.

Raspberry Pi is a well-known series of small low-cost SBCs developed by the Raspberry Pi Foundation in association with Broadcom. The original model became more popular than anticipated, selling outside its target market for uses such as robotics. It is widely used in many areas, such as for weather monitoring, because of its low cost, modularity, and open design. It is typically used by computer and electronic hobbyists, due to its adoption of the HDMI and USB standards.

There are three series of Raspberry Pi, and several generations of each have been released [54]. Raspberry Pi SBCs features a Broadcom System on a Chip (SoC) with an integrated ARM-compatible CPU and on-chip GPU, while Raspberry Pi Pico has a RP2040 system on chip with an integrated ARM-compatible CPU. The high-end model at time of writing this deliverable is the Raspberry Pi 4 B, which runs at 1.5 GHz and is equipped with Broadcom VideoCore VI GPU. It has network interfaces for Gigabit Ethernet and 2.4GHz/5GHz Wi-Fi, while the price varies according to the memory size. A lighter model is the Raspberry Pi Zero 2 W, which also runs a 64-bit quad-core CPU, but at 1 GHz. It is equipped with Broadcom VideoCore VI GPU and communicates over 2.4 GHz Wi-Fi. Another ARM-powered processing platform is the Radxa Rock 4 [55], which combines the powerful Cortex-A72 and the energy efficient Cortex-A53 cores on the same chip. The Rock Pi 4 Model C+ is equipped with two of the former CPUs and four of the latter, it has a Mali T860MP4 GPU and communicates over Ethernet and 2.4GHz/5GHz Wi-Fi. Most of the ASUS Tinker series models [56] are also based on ARM processors, such as the Tinker Board R2.0 with four cores operating at 1.8GHz. This model has the Mali-T764 MP4 GPU and communicates over Ethernet and 2.4GHz/5GHz Wi-Fi. A combination of ARM processors on the same chip is used by Hardkernel's Odroid-N2L (six cores) [57]. It incorporates a Mali-G52 GPU, but Ethernet/Wi-Fi connectivity can only be realised through a USB adapter. It can be connected to 4G/5G network via modem (Quectel or similar).

Nvidia Jetson (Figure 24) is also a series of embedded computing boards from Nvidia [58]. The Jetson TK1, TX1 and TX2 models all carry a Tegra processor (or SoC) from Nvidia that integrates an ARM architecture CPU. Jetson is a low-power system and is designed for accelerating machine learning applications. NVIDIA® Jetson™ is the platform for autonomous machines and other embedded applications. It includes Jetson modules, which are small form-factor, high-performance computers, the NVIDIA JetPack™ SDK for accelerating software, and an ecosystem with sensors, SDKs, services, and products to speed up development. Jetson is compatible with the same AI software and cloud-native workflows used across other NVIDIA platforms and delivers the power-efficient performance customers need to build software-defined autonomous machines. Adding the Jetson on-board to the drone, robot or camera intended for the computer vision tasks enables to execute computationally demanding video analytics and AI operations, getting the results with very low latency and being able to quickly proceed to the following actions using these results. Sending only the calculated result to the cloud saves valuable network bandwidth. A thorough comparison of different SBCs is presented in Table 14.

Figure 23: Different types of SBCs

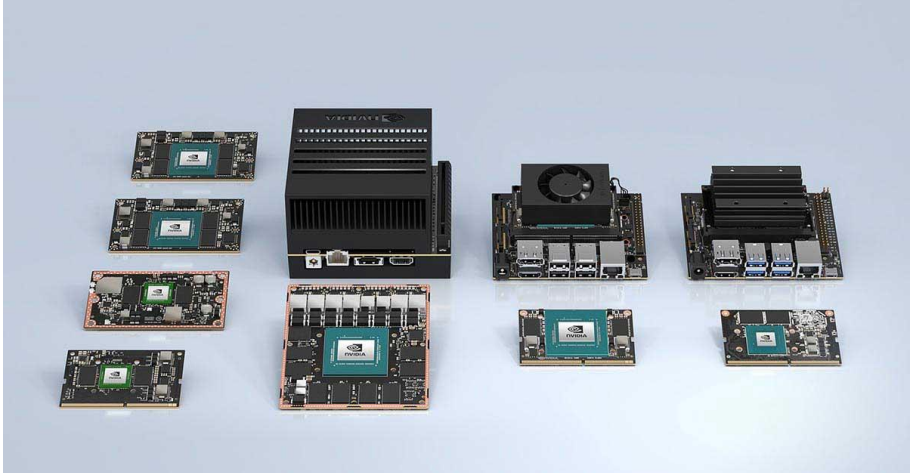


Figure 24: Nvidia Jetson modules

Table 14: Extreme processing enablers' comparison

Platform	FP32 (Single)	Avg. power consumption	Memory size	Cost (€)
Raspberry PI 4	48 GFLOPS	5.1 W	Up to 8 GB	60 - 200
Raspberry Pi Zero 2 W	12 GFLOPS	0.7 W	512 MB	22
Rock Pi 4 Model C+	36 GFLOPS	6.2 W	4 GB	85
Tinker Board R2.0	26 GFLOPS	5.0 W	2 GB	140
Hardkernel Odroid N2L	32 GFLOPS	5.0 W	4 GB	90
Jetson Nano	472 GFLOPS	7.5W	2 GB	85
Jetson AGX Xavier	11 TFLOPS	20W	16 GB	850
Jetson AGX Orin	25 TFLOPS	27.5W	32 GB	2400

The actual Floating-point Operations per Second (FLOPS) required for various computational tasks can vary widely depending on specific requirements, algorithms, sensor resolutions, and more. Also, these tasks are often divided among different types of processors, such as CPUs, GPUs, and specialised AI accelerators, each of which might have different performance characteristics. However, here is a rough idea of the processing power that might be needed for various tasks.

IoT Data Processing: IoT data processing can vary greatly in complexity. For simple tasks, there is the need for only few GFLOPS, e.g., 10 GFLOPS.

LIDAR Data Processing: LIDAR systems generate a high volume of point cloud data that needs to be processed in real-time. While specific TFLOPs figures are hard to pin down without knowing the exact algorithms and LIDAR

specifications, this is certainly a task that would benefit from high-performance processors. Some of the advanced GPUs or AI accelerators available as of 2023 would likely be used for this. LIDAR data processing is a computationally intensive task that involves processing high-density 3D point cloud data and depending on the complexity of the processing task, a rough estimate for the computation requirement might be between 1 to 10 TFLOPS.

Video Analytics: Again, the specifics can vary, but high-resolution video at high frame rates can be very data intensive. For example, advanced driver-assistance systems often use video analytics for tasks like object detection, classification, and tracking, as well as lane detection and traffic sign recognition. These are highly complex tasks that can require substantial computational power. A rough estimate might be in the range of 10 to 40 TFLOPS, but it could be depending on the specifics.

AI Training: Training AI models is an extremely compute-intensive task that is typically done in the far edge or the near edge or the cloud with high-performance GPUs. The required TFLOPS can vary greatly depending on the size and complexity of the models and the datasets. A basic deep learning model might be able to train with a few TFLOPS, but state-of-the-art models might require 100 TFLOPS or more.

4.2.2 Far and Near Edge Processing Enablers

The far edge zone includes devices which can be installed close to network equipment owned by network operators. Latency is a critical factor in the far edge layer, as data must be processed quickly to support real-time applications. On the other hand, near edge zone is the edge computing infrastructure that is deployed in a location between the far edge and the cloud data centres. The typical latency of the far edge layer in an edge computing architecture is generally less than five milliseconds, and the near edge a couple of tens of milliseconds respectively.

Some indicative systems that could serve as solutions for near and far edge deployments can be single board computers for light-weight applications and have already been discussed thoroughly in 4.2.1 sub-section, as well as mini PCs or tower PCs for more demanding loads and applications:

- Mini PCs, also known as minicomputers or mid-range computers, are a variant of computers that possesses most of the features and capabilities of a large computer but is smaller in physical size, as depicted in Figure 25. Mini PCs are a scaled-down cost-effective versions of desktop computers and are optimised for the small space and low power requirements of edge deployments. However due to their small form-factor it is difficult to support CPUs with tens of cores due to limited heat-dissipation mechanisms as well as large components such as external GPUs for the offloading of demanding tasks.

Figure 25: Indicative mini PCs [59]

- A tower server is a computer that is built in an upright cabinet that stands alone in a cabinet called a 'tower' (Figure 26). They can be easily added to existing networks and although bulkier than mini PCs, they are highly customisable and upgradable. Since they take a lot of space, they are easier to cool, reducing chances of damage by overheating when incorporating CPUs with large number of cores. They typically accommodate less than 100 GBs of RAM and only a few drives and GPUs. However, because of their relatively large size and management complexity, tower servers are not suitable when the application demands multiple server instances.

Figure 26: Indicative tower servers [60]

Different types of CPUs and GPUs are being used in far and near edge systems. Comparing GFLOPs (billion floating point operations per second) across different CPUs and GPUs can provide insights into their performance, particularly for compute-intensive tasks like computer vision tasks, LIDAR data filtering or AI model training. However, it's important to note that GFLOPs don't tell the entire story about a chip's overall performance, as different architectures can achieve better or worse real-world performance even with similar GFLOPs figures. Also, note that the GFLOPs measurement is more relevant to GPUs than CPUs.

CPUs can nowadays be realised as vector processors (e.g., Intel's Xeon Phi x200 and Skylake-X CPUs) with an ability of working with multiple data elements simultaneously rather than with a single data element at a time. Vector processors have multiple Arithmetic Logic Units (ALUs) that work synchronously and perform an instruction on a vector of data. Therefore, vector processors use the Single-Instruction-Multiple-Data (SIMD) technique. Among the available hardware platforms, CPUs are often the least used for DNNs (Deep-Neural-Networks) inference or training, as they provide lower FLOPS and FLOPS/W performance (Figure 27). However, manufacturers have recently undertaken measures to accelerate the deployment of Neural Networks CPUs. For example, at the instruction level, Intel has added the AVX-512 Vector Neural Network Instructions (AVX-512 VNNI) to the AVX-512 Instruction Set to accelerate Convolutional Neural Networks. In addition, Intel announced that the next generation of Cooper Lake and Sky Lake processors will support Brain Floating Point (bfloat16) operations. bfloat16 is a floating-radix-point format on 16 bits with a dynamic range comparable with the dynamic range of the 32-bit IEEE 754 floating-point format. bfloat16 is also supported by ARMv8.6-A and AMD's ROCm library. Figure 27 below describes comparison between CPUs and GPUs and Table 15 provides a benchmark between different types of GPUs.

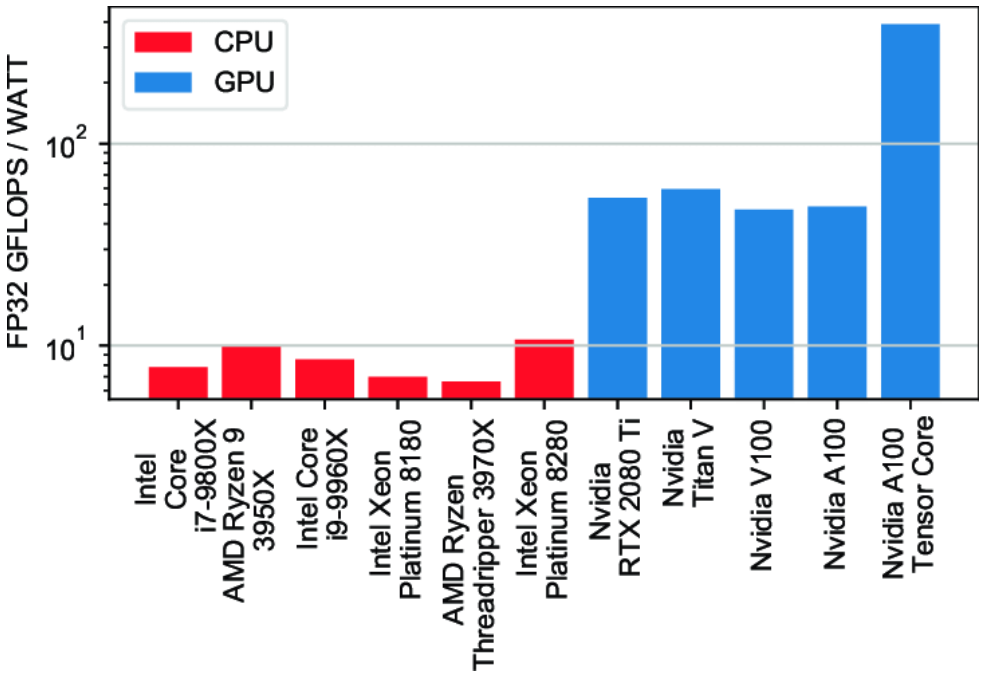


Figure 27: Comparison between CPUs and GPUs based on GFLOPS/W [61]

FP32 and FP64 performance for different GPUs are provided in the following table:

Table 15: Comparison of different types of GPUs [62]

Identification	FP32 (FLOAT)	FP64 (DOUBLE)	TDP	Memory size	Cost (€)
RX 6500 XT	5.765 TFLOPS	0.360 TFLOPS	107W	4 GB	170 - 220
RX 6700 XT	13.21 TFLOPS	0.825 TFLOPS	230W	12 GB	400 - 2700
RX 6900 XT	23.04 TFLOPS	1.440 TFLOPS	300W	16 GB	700 - 2300
RX 7900 XTX	61.42 TFLOPS	1.919 TFLOPS	355W	24 GB	1150 - 1900
RTX 3050	9.098 TFLOPS	0.142 TFLOPS	130W	8 GB	270 - 400
RTX 3070	20.31 TFLOPS	0.317 TFLOPS	220W	8 GB	550 - 1500
RTX 3090	35.58 TFLOPS	0.556 TFLOPS	350W	24 GB	1400 - 3500
RTX 4090	82.58 TFLOPS	1.29 TFLOPS	450W	24 GB	1900 - 2700
ARC 770	19.66 TFLOPS	Not Supported	225W	16 GB	330 - 350

Hardware accelerated processing

Unlike CPUs and GPUs, which are software-programmable fixed architectures, FPGAs are reconfigurable, and their compute engines are defined by the user. With CPUs and GPUs, instruction stages are pipelined, and new instructions start executing every clock cycle. With FPGAs, operations are pipelined so new instruction streams operating on different data start executing every clock cycle. This means that FPGAs will have a very high level of optimisation for the needs of the application (more than CPUs and GPUs). Custom instructions not supported by CPUs and GPUs can be easily implemented and efficiently executed.

ASICs work the same way as FPGAs, except that the hardware cannot be reconfigured by the user and is “burnt” into the silicon chip.

The NPU (Neural Processing Unit) is a chip dedicated to neural networks. Its particularity is that it will provide an answer not only based on a calculation but on a knowledge base that will be able to be enriched over time to continuously improve its precision and speed. It will be used, for example, to do image recognition and improve the quality of the photos taken by the device. NPUs are not limited to mobile devices and can be used on a much larger scale for heavier applications.

A few examples are provided below:

- FPGA manufacturers: AMD (Xilinx), Intel (Altera), Lattice Semiconductor, Microchip Technology, etc.
- ASIC manufacturers: Infineon Technologies, STMicroelectronics, Texas Instruments, Analog Devices, Advanced Linear Devices, NXP, etc.

5 Service Processing Optimisation and Green ICT

Based on the characteristics of the various connectivity and processing technologies, this section focuses on the design of an algorithm for selecting the set of technologies that best meet the requirements emanating from the Use Cases. While doing so, the algorithm optimises towards user-specified objectives, e.g., cost and power consumption. The output of the algorithm includes connectivity/processing technology pairs, the number of resources to deploy, e.g., component quantity, and the way these are connected, i.e., the topology.

For the considered algorithmic process, information from the following Tables is used: (a) 'Comparison of access network technologies' (Section 193.1), referred to as **AT Matrix**, (b) 'Comparison of transport network technologies' (Section 3.2), referred to as **TT Matrix**, (c) 'Terrain constraints and communication technologies' (Section 3.4), referred to as **Terrain Matrix**, (d) 'General characteristics of different compute continuum zones' (Section 4.1.2) specifically the 'Latency' column, referred to as **EL Matrix** and (e) 'Extreme processing enablers' (Section 4.2.1), referred to as **PD Matrix**.

Combining the information of the **Terrain Matrix** with the rest of the information, we result to the **Filter Matrix** in Table 16 which will be used for the network filtering process of the algorithm (see Section 5.1). The matrix below contains dummy values, which serve as an example for demonstration purposes. Please note that the columns considered currently are indicative and will be updated in the future. Also, although not shown, Table 16 includes all access and transport network technologies considered – a value of 1 is set in cases where a particular network technology applies and 0 otherwise. A dummy example is provided for 2 access network technologies (Wi-Fi 6 and 5G) and 3 indicative feature categories, i.e., Latency, Terrain and Area Size.

Table 16: Filter matrix

	Latency		Terrain			Area Size	
	Low	High	Mountain	Flat	Sea	Small	Big
Wi-Fi 6	1	0	0	1	0	1	0
5G	1	0	1	1	0	1	0
…	…	…	…	…	…	…	…

In this example, we see that while Wi-Fi 6 and 5G both can provide low latencies, Wi-Fi is not a suitable technology for mountainous terrains due to the propagation properties and typical operation frequencies, as discussed in Section 3.

5.1 Technology Selection Algorithm

An important task tackled by the XGain KFT is to find the proper mix of technologies and to dimension it properly to satisfy the needs of a service and sector (or a specific UC). In the context of dimensioning problems, where the goal is to determine the appropriate technology or infrastructure requirements for a given application or system, a technology selection algorithm plays a critical role. Dimensioning problems often involve complex considerations such as capacity planning, resource allocation, and cost optimisation. With a wide range of available technologies and options, it becomes increasingly challenging to manually identify the most suitable technology for a specific scenario.

To address this challenge, a technology selection algorithm serves as a valuable tool to automate and streamline the decision-making process. It takes into account various factors, such as performance requirements, resource availability, cost constraints, and scalability considerations, to identify the optimal technology or combination of technologies that meet the desired objectives. By employing a technology selection algorithm, the dimensioning problem can be approached systematically and objectively. The algorithm considers both quantitative and qualitative factors, enabling a comprehensive evaluation of available technologies. It allows for efficient exploration of different options, considering trade-offs between performance, cost, scalability, and other relevant criteria.

Moreover, a technology selection algorithm helps to overcome the limitations of manual decision-making, which can be time-consuming, error-prone, and subjective. It leverages computational power to process large datasets, perform complex calculations, and analyse multiple scenarios, resulting in more accurate and reliable decisions. In the following, we state the specific problem that needs to be solved and the relevant algorithmic steps.

The technology selection problem can be formulated as follows: select (a) appropriate network topology, access and transport networks (the type of network and the number of network devices), and (b) the type of processing devices, their placement in the compute continuum, and their number to satisfy specific requirements coming from a UC. Specifically, considering the service that a Use Case covers, the requirements are the **total network demand** to be satisfied (throughput requirement in bps), **maximum coverage** (Km^2), **minimum computational** requirement (FLOPS) and **maximum end-to-end latency** (s) as described in 4.1.2. In the following algorithmic steps, we make the simplifying assumption that the studied areas are square regions.

4.2.1

Algorithmic Steps

A) Input

Information included in the KFT (about Network and Processing Devices)

1) For each k^{th} Access Technology (AT) we have: $Throughput_{k,AT}$ (bps), $Range_{k,AT}$ (m), Communication $Latency_{k,AT}$ (s), $Power_{k,AT}$ (W) and $Cost_{k,AT}$ (€) (the term 'cost' represents the Capex cost) for each infrastructure device, i.e. the **AT Matrix**. The Power and Cost represent the corresponding values of indicatively a 5G Base Station (BS), a Wi-Fi access point or an IoT GW.

2) For each b^{th} Transport Technology (TT) we have: $Throughput_{b,TT}$ (bps), $Range_{b,TT}$ (m), Communication $Latency_{b,TT}$ (s), $Power_{b,TT}$ (W) and $Cost_{b,TT}$ (€) (the term 'cost' represents the Capex cost) for each infrastructure device, i.e. the **TT Matrix**.

3) Networking technologies are filtered based on the characteristics of each access and transport network, i.e., their applicability in small, large, flat and mountainous terrains presented in **Terrain Matrix,** and their communication latency, water and weather resistance and foliage support all combined in **Filter Matrix**. Particularly, the separation in small and large areas assumes, indicatively, square regions smaller than $10Km^2$ in the first case and larger than $10Km^2$ in the second case.

4) For each i^{th} zone of the compute continuum we have the maximum latency (l_i) incurred, i.e., the **EL Matrix**.

5) For each j^{th} device of Processing Devices (PD) we have: Computing $Capacity_j$ (FLOPS), $Power_j$ (W) and $Cost_j$ (€) (the term 'cost' represents the Capex cost), i.e., the **PD Matrix**.

Information provided by the End User

6) (a) **Area Information**: size (m^2), type (mountainous, sea, flat, foliage etc.), meteorological conditions (wind, rain, fog, snow)

(b) **Type of service**: IoT Data Processing, LIDAR Data Processing, Video Analytics, AI Training, etc.

(c) **Number of end devices** (e.g., drones), referred to as Dev, to implement the service.

(d) **Demand (bps) per end device**: If the demand per d^{th} end device is given by Dem_d the total demand (TD) is simply calculated by $TD = \sum_{d=1}^{Dev} Dem_d$. Moreover, we assume that the total demand is uniformly distributed inside the considered area (i.e., for example each BS offers the same capacity to end devices) and each end device has to be served by only one access device.

(e) **Tolerable communication latency**: the maximum latency (s) that can be tolerated by the requested service.

B) Filter the Network and Processing technologies, thus reducing the Search Space

7) Given **6(a)** and **6(e)** the appropriate type of AT and TT networks is selected by considering the **Filter Matrix**. This results to smaller AT and TT sets, referring to them as AT_n and TT_n, respectively. A possible filtering process is presented below.

Specifically, the Filter Matrix includes different categories of features (e.g., Latency, Terrain, Area size), referred to as CF and there are binary values to present under which conditions (in total F features) each network technology is applicable. Based on **6(a)** and **6(e)** a binary vector (PF of shape 1xF) is derived from the End User preferences (during requirement gathering) including the same features of the Filter Matrix, i.e., the area information and the tolerable latency that can be encoded as high or low. In case that we have N possible networks we can multiply the user's vector with the transposed Filter Matrix (of shape FxN) and result in the vector 1xN that represents which network technologies can meet the requirements.

Particularly, in the latter vector, the network technology whose value is equal to CF means that it applies to every feature's category of the End User's preferences and is an appropriate network technology. Hence, the AT_n and TT_n sets are constructed. A dummy example for outlining the process is shown in Figure 28.

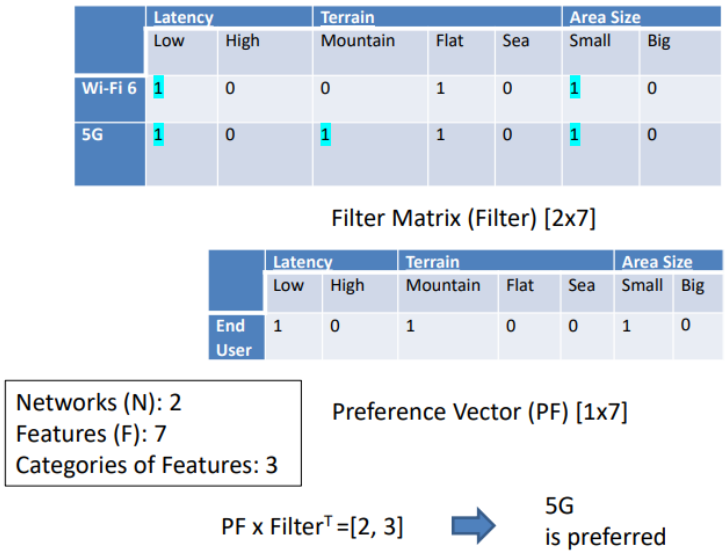


Figure 28: Filter matrix and preference vector

8) Given **6(b)** the appropriate type of processing devices by considering the **PD Matrix**. This results to a smaller PD set, referred to as PD_n. A similar filtering process as (7) can be applied by considering for each processing device the type of services it satisfies based on its computing capacity (in FLOPS).

C) Derive the Network Infrastructure (Access and Transport Networks and Links between them and End Devices)

The Ceil(x) function in the following analysis determines the smallest integer i, such that i >= x.

Number of Access Devices and Links between End and Access Devices

9) (a) The end devices, the number of which is represented by **Dev**, produce the TD given in **6(d)**. The access network devices have to serve this demand and cover the examined area having in mind a load balancing criterion, i.e., the same or a similar load is produced by each access device. In the analysis below we make the logical assumptions that each end device is served by one access device and that the number of end devices has to be at least equal to the number of access devices. Considering these we conclude in:

For each n^{th} network in AT_n set:

- Compute the number of c_{ATn} network's devices to cover the square area given in **6(a)**:

$$c_{ATn} = \text{Ceil}(\text{Area_size}/\text{Range}^2_{n,AT})$$

- Compute the number of t_{ATn} network's devices to satisfy the total demand given in **6(d)**:

$$t_{ATn} = \text{Ceil}(\text{TD}/\ \text{Throughput}_{n,AT})$$

- Keep as current number of U_{ATn} network devices: $U_{ATn} = \max\{c_{ATn}, t_{ATn}\}$
- Connect the end devices from first to last to access devices based on a **round robin** scheduling algorithm, where the end devices are distributed in simple rotation until the maximum throughput of an access device is exhausted.
- Increase U_{ATn} if access devices cannot support the required throughput.

(b) We have the U_{AT} vector (of shape $1x|AT_n|$, where $|AT_n|$ is the number of elements in AT_n set) of the number of devices for each Access Network in the AT_n set. The symbol |s| represents the number of elements in s.

Number of Transport Devices and Links between Access and Transport Devices

10) (a) For each p^{th} network in TT_n set:

For each n^{th} network in AT_n set:

- Compute the number of $U_{TTp,n}$ network devices to satisfy the total access network demand AD_n = TD (check that $U_{TTp,n} \leq U_{ATn}$ considering that transport network devices satisfy larger throughput than access network devices)

$$U_{TTp,n} = \text{Ceil}(\text{TD}/\ \text{Throughput}_{p,TT})$$

- Connect the access devices (from first to last) to transport devices based on a **round robin** scheduling algorithm, where the access devices are distributed in simple rotation among transport devices until the maximum throughput of a transport device is exhausted.

- Increase $U_{TTp,n}$ if transport devices cannot support the required throughput.

(b) We have the U_{TT} matrix (of shape $|TT_n|x\ |AT_n|$) containing the $U_{TTp,n}$ number of devices for each Transport Network in the TT_n set, given each Access Network in the AT_n set.

11) Until this point we have: (a) the number of devices in the access (U_{AT}) and transport (U_{TT}) networks, and (b) the links between end devices and access devices as well as between access devices and transport devices.

D) Determine the number and placement of processing devices in the compute continuum

12) Having derived the appropriate infrastructure of Access and Transport Networks (various different combinations) that satisfy the requirements of the **total demand** and the **total coverage**, the problem hereafter is to determine the number and placement of the processing devices from the PD_n set in the compute continuum (at which zone) subject to an objective and some criteria. To do that we formulate the following optimisation problem.
13) **Problem**: We have M (i=1,…,M) compute (edge) zones (maximum M=4) and N (j=1,…,N) different types of processing devices from the PD_n set. Each i^{th} compute continuum zone has a maximum communication latency, $\mathbf{l_i}$, as well as maximum ($\mathbf{CRE^{max}_i}$) and minimum ($\mathbf{CRE^{min}_i}$) tolerable computational requirements of its allocated devices. Moreover, each j^{th} device has $\mathbf{cost_j}$, $\mathbf{power_j}$ and computational capacity $\mathbf{cc_j}$.

    A UC has a maximum latency requirement **L** and a specific computational requirement that has to be satisfied. The problem (presented in 5.2) is to allocate the processing devices to compute continuum zones in order to minimise some cost function (e.g. CAPEX, power consumption or other performance metrics) while satisfying the communication latency and the computational requirements.

    Finally, for the extreme edge we consider that there is a one-to-one pairing of processing devices with end devices, hence either zero or Dev processing devices are allocated to that zone.

14) For each of n^{th} network in AT_n set:

    For each of p^{th} network in TT_n set:

    i. Substitute the U, R in the objective, (C3), (C4) and (C5) of Figure 29 as U_{ATn} and $U_{TTp,n}$ respectively.
    ii. Solve the Problem and compute the corresponding processing devices' cost objective (or other performance metric), i.e. $process_cost_{n,p}$.
    iii. Overall cost of the $topology_{n,p}$:

$$process_cost_{n,p} + U_{ATn} * Cost_{n,AT} + U_{TTp,n} * Cost_{p,TT}$$

E) Output

15) We rank the n*p feasible solutions, i.e., access and transport network infrastructure, number and placement of processing devices, from the most cost effective to the least based on the overall cost in step (14) (other performance metrics could also be considered).

5.2 Optimisation Problem

As shown in Figure 29, the problem is formulated as an integer minimisation problem with a linear objective, along with five sets of constraints that include linear and quadratic terms.

$$\min_{X,W,G,O} \sum_{d=1}^{Dev}\sum_{j=1}^{N} g_{dj}\text{cost}_j + \sum_{k=1}^{U}\sum_{j=1}^{N} w_{kj}\text{cost}_j + \sum_{r=1}^{R}\sum_{j=1}^{N} o_{rj}\text{cost}_j$$

$$s.t. \quad \text{Latency constraints} \quad (C1)$$
$$\text{Extreme edge constraints} \quad (C2)$$
$$\text{Far edge constraints} \quad (C3)$$
$$\text{Near edge constraints} \quad (C4)$$
$$\text{Far-Near edge constraint} \quad (C5)$$

(a)

Latency Constraints

$x_i l_i \le L, \ \forall i \le M$ (C1a)

$x_1 x_2 = 0$ (C1b)

$x_1 x_3 = 0$ (C1c)

$x_i \in \{0,1\}, \ \forall i \le M$ (C1d)

Extreme Edge Constraints

$\sum_{j=1}^{N} g_{dj} = x_1, \ \forall d \le Dev$ (C2a)

$\sum_{j=1}^{N} g_{dj} cc_j \ge x_1 FL_d^1, \ \forall d \le Dev$ (C2b)

$g_{dj}(cc_j - CRE_1^{\max}) \le 0 \ \forall d \le Dev, j \le N$ (C2c)

$g_{dj}(cc_j - CRE_1^{\min}) \ge 0 \ \forall d \le Dev, j \le N$ (C2d)

$g_{dj} \in \{0,1\}, \ \forall d \le Dev, j \le N$ (C2e)

Far-Near Edge Constraint

$\sum_{u=1}^{Num_U[r]} \min\left[\sum_{j=1}^{N} w_{UR[u,r]j} cc_j, FL_u^2\right] + \sum_{j=1}^{N} o_{rj} cc_j \ge (1-x_1) FL_r^3, \ \forall r \le R$ (C5)

Far Edge Constraints

$(1-x_2)\sum_{k=1}^{U}\sum_{j=1}^{N} w_{kj} = 0$ (C3a)

$\sum_{j=1}^{N} w_{kj} cc_j \le (FL_k^2 + t_{far}), \ \forall k \le U$ (C3b)

$w_{kj}(f_{kj} - 1) = 0 \ \forall k \le U, j \le N$ (C3c)

$f_{kj}(cc_j - CRE_2^{\max}) \le 0 \ \forall k \le U, j \le N$ (C3d)

$f_{kj}(cc_j - CRE_2^{\min}) \ge 0 \ \forall k \le U, j \le N$ (C3e)

$0 \le w_{kj}$, integer $\forall k \le U, j \le N$ (C3f)

Near Edge Constraints

$(1-x_3)\sum_{r=1}^{R}\sum_{j=1}^{N} o_{rj} = 0$ (C4a)

$\sum_{j=1}^{N} o_{rj} cc_j \le (FL_r^3 + t_{near}), \ \forall r \le R$ (C4b)

$o_{rj}(n_{rj} - 1) = 0 \ \forall r \le R, j \le N$ (C4c)

$n_{rj}(cc_j - CRE_3^{\max}) \le 0 \ \forall r \le R, j \le N$ (C4d)

$n_{rj}(cc_j - CRE_3^{\min}) \ge 0 \ \forall r \le R, j \le N$ (C4e)

$0 \le o_{rj}$, integer $\forall r \le R, j \le N$ (C4f)

(b)

Figure 29: Optimisation problem: (a) Overall Problem and (b) Constraints of the Problem

Variables of the Problem:

- $\mathbf{x_i}$ (binary) means that i[th] compute continuum zone can be used if x_i=1.
- $\mathbf{g_{dj}}$ (binary) means that the j[th] type device will be used in the d[th] end device if g_{dj}=1.
- $\mathbf{w_{kj}}$ (integer) represents the number of the j[th] type devices that will be used in the k[th] access device.
- $\mathbf{o_{rj}}$ (integer) represents the number of the j[th] type devices that will be used in the r[th] transport device.

Parameters of the Problem:

- **Dev, U** and **R** represent the number of end, access and transport devices, respectively.
- FL_d^1 represents the computational requirement (FLOPS) of the d[th] end device.

- FL_k^2 represents the computational requirement (FLOPS) of the k[th] access device considering its connection with the end devices.
- FL_r^3 represents the computational requirement (FLOPS) of the r[th] transport device considering its connection with the access devices.
- In C5: **UR[u,r]** is the binary allocation matrix among the access and transport devices and **Num_U[r]** is the number of paired access devices to the r[th] transport device.

Constraints of the Problem:

- **C1 (Latency Constraints):** The i[th] compute continuum zone is active only if the latency requirement is achieved. Moreover, we assume that the processing devices allocated to the d[th] end device have to satisfy the computational requirement of this device. For example, if there are drones and they have video analytics service, then the SBC allocated in each drone has to satisfy this service, otherwise the SBCs are allocated to the remaining compute continuum zones. Hence, if the extreme edge is selected for the allocation, the other zones will not participate in the solution because the total number of FLOPS are already satisfied.
- **C2 (Extreme Edge Constraints)**: One processing device can be allocated to each end device, in total **Dev** processing devices can be used and the processing device has to satisfy the computational requirement of the end device's service. Moreover, there are maximum (**CRE^{max}**) and minimum (**CRE^{min}**) computational requirements in all zones for the processing devices that can be allocated to each zone These are used to avoid the allocation of lightweight processing devices to higher edge zones and heavyweight processing devices to lower compute continuum zones.
- **C3 (Far Edge Constraints)**: The computational capacity of allocated processing devices to an active device should not be much higher than the computational requirement that this active device has to serve (thus avoiding large spare computational capacity). Hence, we add the non-negative scalar parameter **t_{far}** that is predefined and can control heuristically the aforementioned situation. Finally, there are maximum and minimum computational requirements for the processing devices of this zone.
- **C4 (Near Edge Constraints)**: Under similar assumptions with (C3) we construct the (C4) constraints and we add the non-negative scalar parameter **t_{near}** that is predefined. Finally, there are maximum and minimum computational requirements for the processing devices of this zone.
- **C5 (Far-Near Edge Constraint)**: In case that the extreme edge is not active, the total computational requirement has to be served by the devices in the far or near edge or in both zones. This is captured by this constraint. Hence, constraints (C3b) and (C4b) can be excluded from the optimisation algorithm, because the total computational requirement will be satisfied in (C5). However, the former helps us to achieve a total offered computation 'closer' to the requested computation.

5.3 Indicative Examples of the Algorithm

In the Figures below, we present examples of the network topology and the associated allocation of processing devices to satisfy the requirements. Specifically, we show the type and number of access and transport devices, the type and number of processing devices and their location. The algorithm results in feasible allocations, i.e. the proposed allocations satisfy all the requirements that have been provided by the user. Finally, we depict the total computational capacity allocated, the service's requested capacity, any excess capacity, and the total power and cost of processing devices.

5.3.1 Toy Example (Video Analytics)

This theoretical example below is used to demonstrate the output of the algorithm when dealing with a scenario with high data processing requirements. Specifically, a possible scenario is the existence of drones equipped with cameras for video streaming, such as monitoring large wildlife areas, e.g. dense forests, and processing the stream for keeping track of animals and their living conditions.

Given parameters:

- **Area Information**: $4*10^4$ m^2
- **Number of end devices**: Dev=10 drones
- **Type of service:** Video Analytics (computational capacity requirement of each drone's video is around 60 TFLOPS (according to Section 4.2.1), so in total CR=10*60TFLOPS=600000 GFLOPS)
- **Demand (bps) per end device**: Each drone creates a stream of 50 Mbps (e.g. an ultra-high definition, 4Kp60 video [64]), which is used for video analytics. Hence total demand is TD = 500 Mbps.
- **Access devices**: Wi-Fi access points offering $Throughput_{AT}$=200 Mbps and Range=100 m.
- **Transport devices**: Microwave backhaul transport devices of $Throughput_{TT}$=500 Mbps.
- We assume **3 compute continuum zones** [Extreme, Far, Near] and 4 types of **processing devices** [Ras_PI4, Rock_PI, Jet_Nano, Jet_Xavier]
- **Latency:** l=[10,25,50] ms includes the maximum tolerable latencies in each i^{th} zone
- **Maximum computational requirement**: CRE^{max}=[0.5, 20, 20]*10^3 GFLOPS for each i^{th} zone
- **Minimum computational requirement**: CRE^{min}=[0, 40, 40] GFLOPS for each i^{th} zone
- $\mathbf{t_{far}}$ =$\min\left(cc\left[cc > CRE_2^{min}\right]\right)$ =48 GFLOPS and $\mathbf{t_{near}}$ =$\min\left(cc\left[cc > CRE_3^{min}\right]\right)$=48 GFLOPS
- **Computational capacity: cc**=[48, 36, 472, 11000] GFLOPS for each j^{th} device
- **Power**: p= [5.1,6.2,7.5,20] Watts for each j^{th} device
- **Cost**: cost=[130, 85, 85, 850] € for each j^{th} device

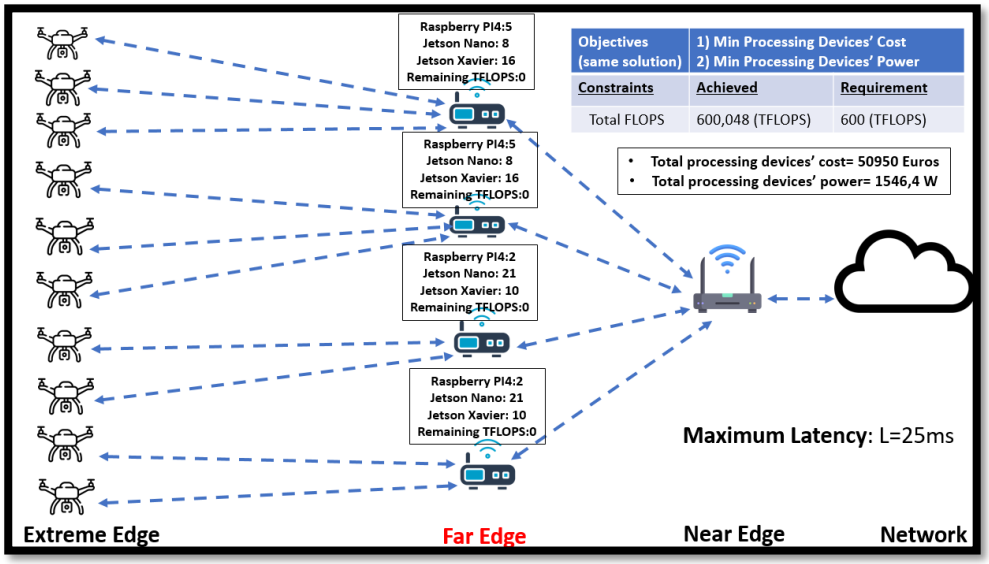
Number of access and transport devices and the corresponding connections:

- Number of Wi-Fi access points:

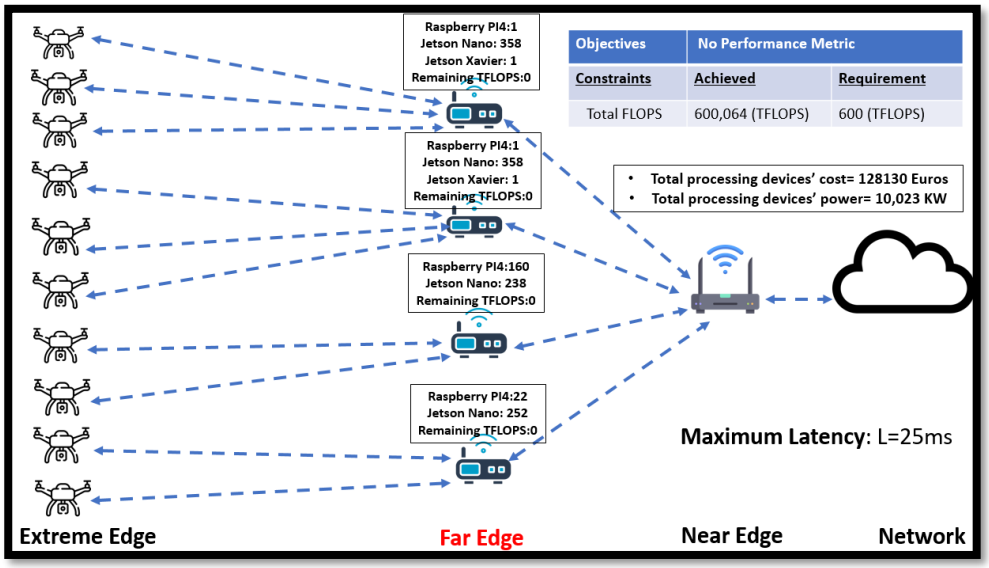
  $\mathbf{U_{AT}}$=max{Ceil(Area_size/$Range^2_{AT}$),Ceil(TD/$Throughput_{AT}$)}=

  max{Ceil($4*10^4/10^4$),Ceil(500/200)}=**4**
- Number of transport devices: $\mathbf{U_{TT}}$= Ceil(TD/$Throughput_{TT}$)= Ceil(500/500)=**1**
- Considering a load balancing objective, two Wi-Fi access devices serve 3 drones each and the other two serve 2 drones each.

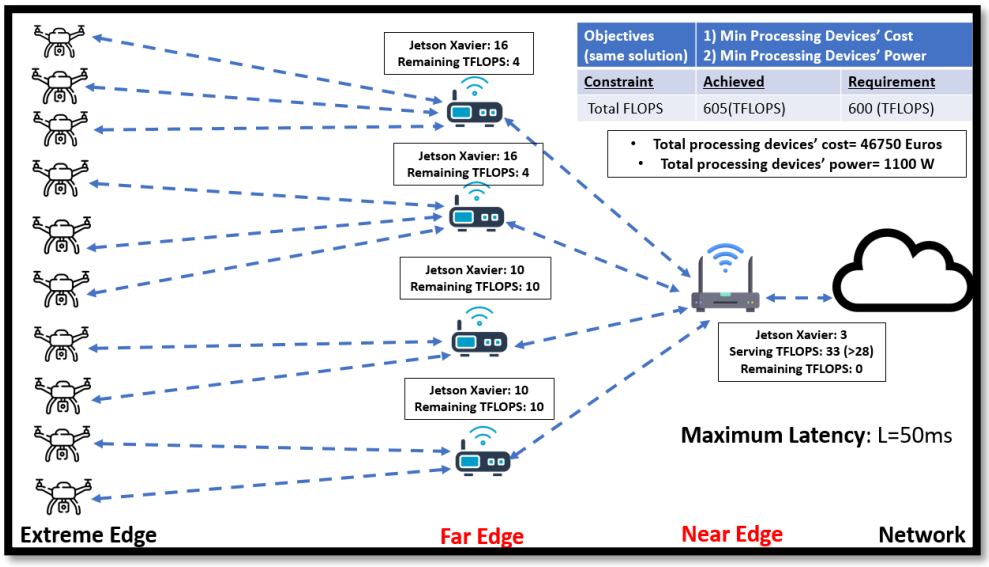
Overall network configuration:



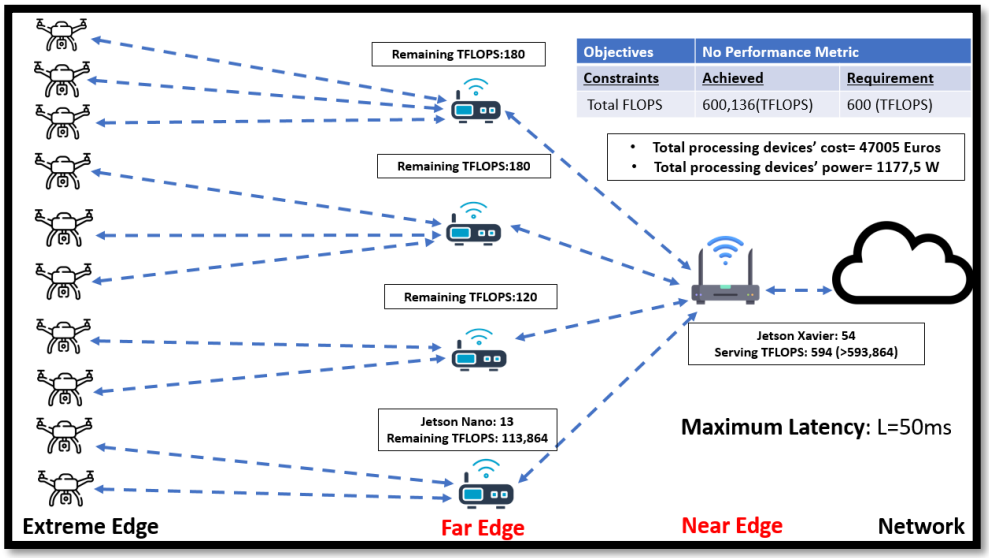
(a)



(b)



(c)



(d)

Figure 30: Solution of optimisation problem no1: Position and number of processing devices in compute continuum for: (a) Maximum Latency =25ms and Cost and Power Objectives, (b) Maximum Latency =25ms and No Objective, (c) Maximum Latency =50ms and Cost and Power Objectives and d) Maximum Latency =50ms and No Objective.

In Figure 30(a) and (b), due to the fact that we have a 25ms latency requirement, the near edge zone is excluded from the allocation. In Figure 30(a) the specific type and number of processing devices that have to be attached to each Wi-Fi access point, to satisfy the total computational requirement, is shown. The minimisation of cost and power objectives result in the same allocation that includes Jetson Xavier, Jetson Nano and Raspberry PI 4 devices. In Figure 30(b) if we do not consider any performance metric we result in a different allocation with worse performance, but with the same type of devices as in the case of Figure 30(a). Finally, in Figure 30(c) and (d) we relax the latency requirement and in addition to the far edge zone, the near edge is used. Figure 30(c) Jetson Xavier is only proposed, while in Figure 30(d) Jetson Nano is also used.

Comparing Figure 30(a) and (b) with Figure 30(c) and (d) it is obvious that when the latency constraint is relaxed from 25 to 50 ms, the near edge zone can be activated. Moreover, the allocation considering either power or cost minimisation is the same (it is possible to have different allocations for different data values) and much better (lower power consumption and cost) compared to the results of the allocation without minimising an objective, i.e. comparing Figure 30(a) with Figure 30(b) and Figure 30(c) with Figure 30(d). Finally, considering of total cost and power with smaller latency, i.e. 25 ms (Figure 30(a) and Figure 30(b)) and higher latency, i.e. 50 ms (Figure 30(c) and Figure 30(d)), we conclude that the relaxation of the latency constraint can result in lower cost and power consumption.

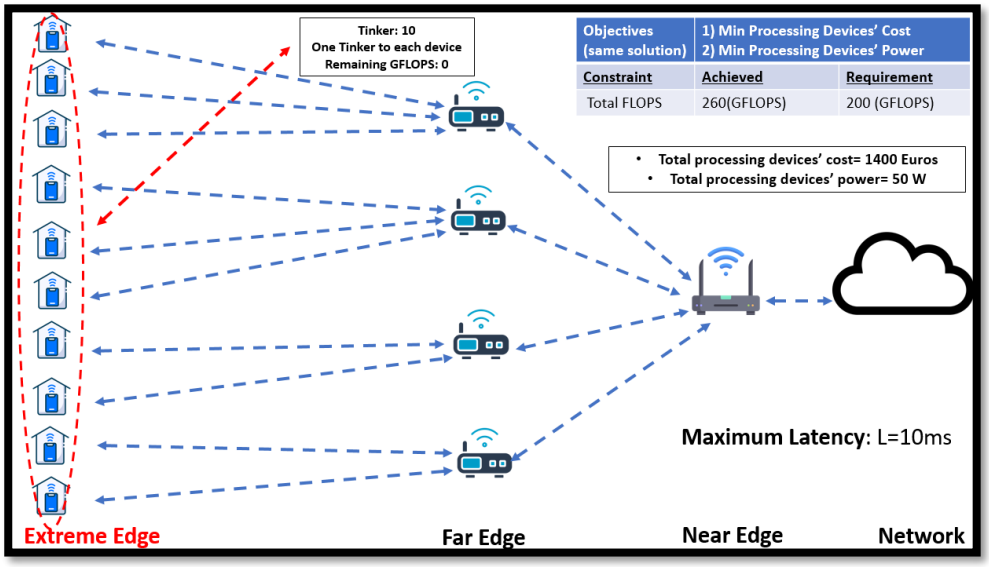
5.3.2 Toy Example (Low Data Processing)

This simple example below is used to demonstrate the output of the algorithm when dealing with a theoretical scenario with low data processing requirements. Processing devices with lower processing capabilities could be attached in the end devices of the extreme edge.

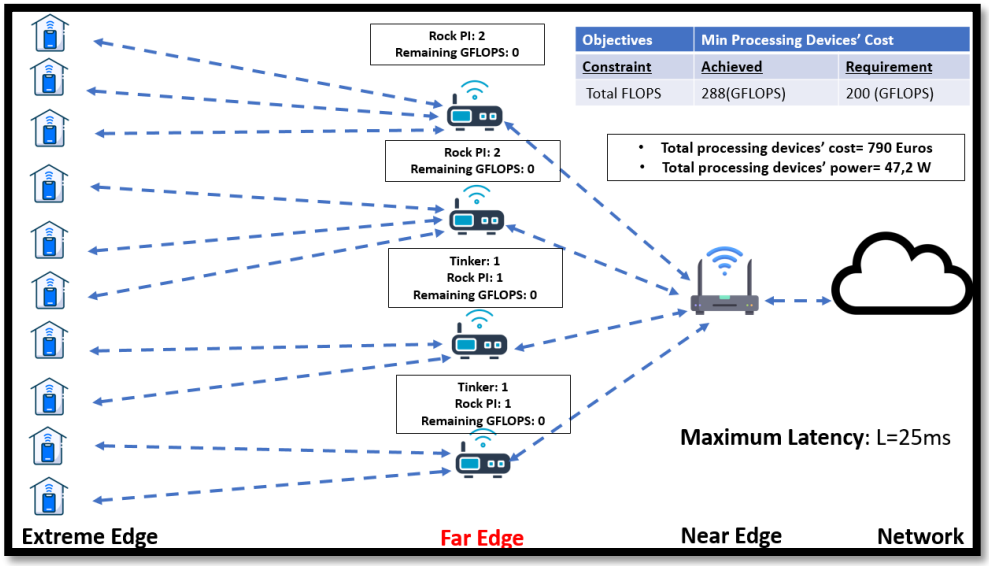
Given parameters:

- For demonstration purposes we assume that we have **2** access devices and **1** backhaul device.
- **Number of end devices**: Dev=10 IoT devices
- **Type of service**: Data Processing (e.g. computational capacity requirement for an IoT device is around 20 GFLOPS (according to Section 4.2.1) so in total CR=10*20 GFLOPS=200 GFLOPS)
- We assume **3 compute continuum zones** [Extreme, Far, Near] and 4 types of **processing devices** [Ras_Zero, Tinker, Rock_PI, Jet_Nano]
- **Latency:** l=[10,25,50] ms which includes the maximum latencies in each i^{th} zone
- **Maximum computational requirement**: CRE^{max}=[0.03, 0.038, 20]*10^3 GFLOPS for each i^{th} zone
- **Minimum computational requirement**: CRE^{min}=[0, 20, 300] GFLOPS for each i^{th} zone
- $\mathbf{t_{far}} = \min\left(cc\left[cc > CRE_2^{min}\right]\right)$=26 GFLOPS and $\mathbf{t_{near}} = \min\left(cc\left[cc > CRE_3^{min}\right]\right)$=472 GFLOPS
- **Computational capacity**: **cc**=[12, 26, 36, 472] GFLOPS for each j^{th} device
- **Power**: p= [0.7, 5,6.2,7.5] Watts for each j^{th} device
- **Cost**: cost=[22, 140, 85, 85] € for each j^{th} device

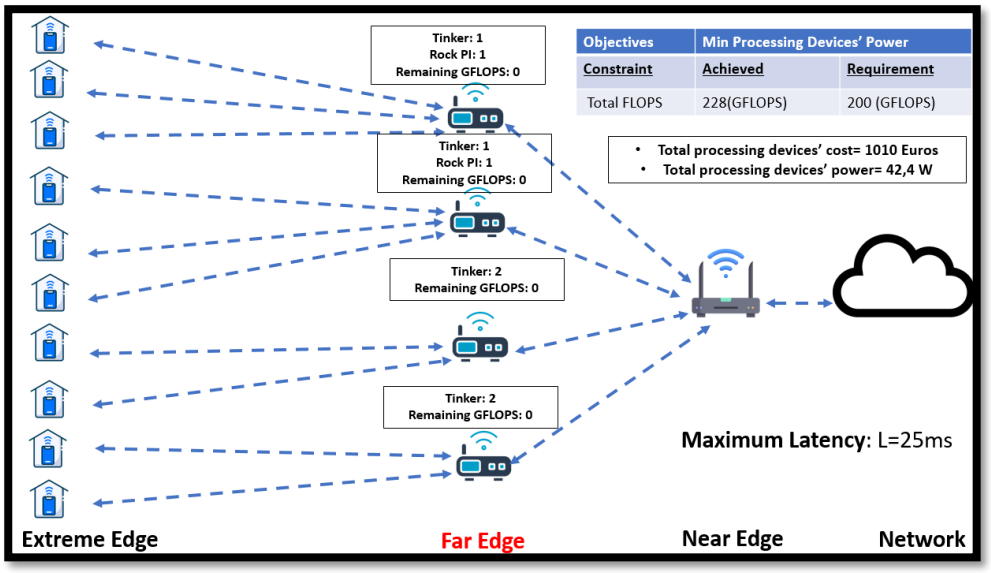
Overall network configuration:



(a)



(b)



(c)

Figure 31: Solution of optimisation problem no2 for: (a) Maximum Latency =10ms and Cost and Power Objectives, (b) Maximum Latency =25ms and Cost Objective and (c) Maximum Latency =50ms and Power Objective.

In Figure 31(a) we observe that in order to achieve the latency and processing capacity requirements of the service, 10 Tinker Board R2.0 devices are used and attached in the end devices. Moreover, in Figure 31(b) 6 Rock PI 4 and 2 Tinker Board R2.0 devices are used to capture the computational requirements considering the minimisation of cost (Tinker's cost is higher than Rock Pi 4's cost), all attached to the far edge zone. However, in case of power consumption minimisation (Tinker's power consumption is lower than Rock PI 4's) in Figure 31(c) 2 Rock PI 4 and 6 Tinker Board R2.0 devices are used, all attached to the far edge zone.

Comparing Figure 31(a) with Figure 31(b) and (c) it is obvious that when the latency constraint is relaxed from 10 to 25 ms, the far edge zone can be activated, while only the extreme edge is used for a latency of 10 ms. Moreover, we permit only the Tinker and Rock PI 4 devices by setting appropriately the CRE^{max} and CRE^{min} values in the given parameters, hence we observe only these types of devices in the far edge. Furthermore, the minimisation of cost results in an allocation that has a lower cost and a higher power consumption. The opposite for the power minimisation objective as we can observe by comparing Figure 31(b) with Figure 31(c). Finally, considering cost and power consumption with a smaller latency, i.e. 10ms (Figure 31(a)), and higher latency, i.e. 25ms (Figure 31(b) and (c)), we conclude that the relaxation of the latency constraint can result in a lower cost and power consumption.

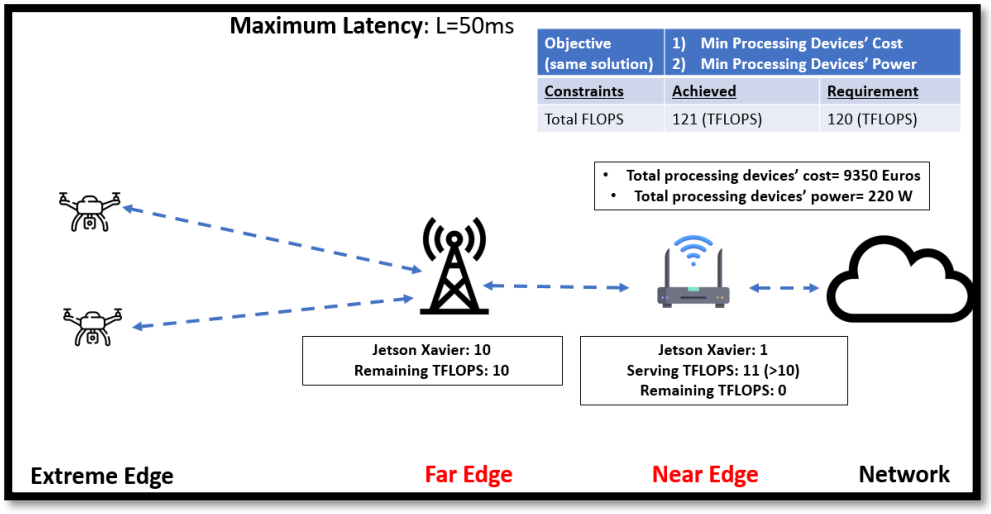
5.3.3 Drone-based Video Analysis

This example is inspired by one of the potential services that could be implemented in the Spanish UC (UC1): drones equipped with cameras record a video that is immediately sent and processed with a video analytics application. Such an application could be for example the recording and analysis of crops to identify potential sickness or pests that affect plants, or it could be a live and remote assessment of the structural stability of a building. We assume, for demonstration purposes, two application scenarios with 50 ms and 25 ms tolerable latency.

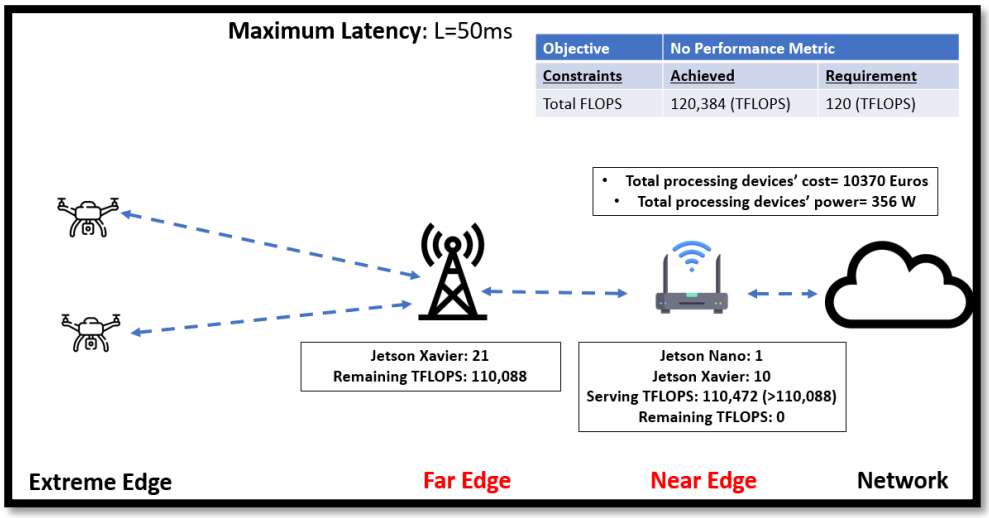
- Given parameters:
    - **Area Information:** 10^4 m^2
    - **Number of end devices**: Dev=2 drones
    - **Type of service**: Video Analytics (computational capacity requirement of each drone's video is around 60 TFLOPS (according to Section 4.2.1) so in total CR=2*60TFLOPS=120000 GFLOPS)
    - **Demand (bps) per end device**: Each drone creates 25 Mbps of UHD quality encoded video for video analytics. Hence total demand is TD = 50 Mbps.
    - **Access devices**: We assume for demonstration purposes that we have Sub – 1 GHz 5G access devices of $Throughput_{AT}$=400 Mbps and Range=10 Km.
    - **Transport devices**: Microwave backhaul transport devices of $Throughput_{TT}$=500Mbps.
    - We assume **3 compute continuum zones** [Extreme, Far, Near] and 4 types of **processing devices** [Ras_PI4, Rock_PI, Jet_Nano, Jet_Xavier]
    - **Latency:** l=[10,25,50] ms which includes the maximum latencies in each i^{th} zone
    - **Maximum computational requirement**: CRE^{max}=[0.5, 20, 20]*10^3 GFLOPS for each i^{th} zone
    - **Minimum computational requirement**: CRE^{min}=[0, 40, 40] GFLOPS for each i^{th} zone
    - $\mathbf{t_{far}} = \min\left(cc\left[cc > CRE_2^{min}\right]\right)$ =48 GFLOPS and $\mathbf{t_{near}} = \min\left(cc\left[cc > CRE_3^{min}\right]\right)$=48 GFLOPS
    - **Computational capacity**: cc=[48, 36, 472, 11000] GFLOPS for each j^{th} device
    - **Power**: p= [5.1,6.2,7.5,20] Watts for each j^{th} device
    - **Cost**: cost=[130, 85, 85, 850] € for each j^{th} device
- Number of access and transport devices and the corresponding connections:
    - Number of 5G base stations:

      $\mathbf{U_{AT}}$=max{Ceil(Area_size/$Range^2_{AT}$),Ceil(TD/$Throughput_{AT}$)}=

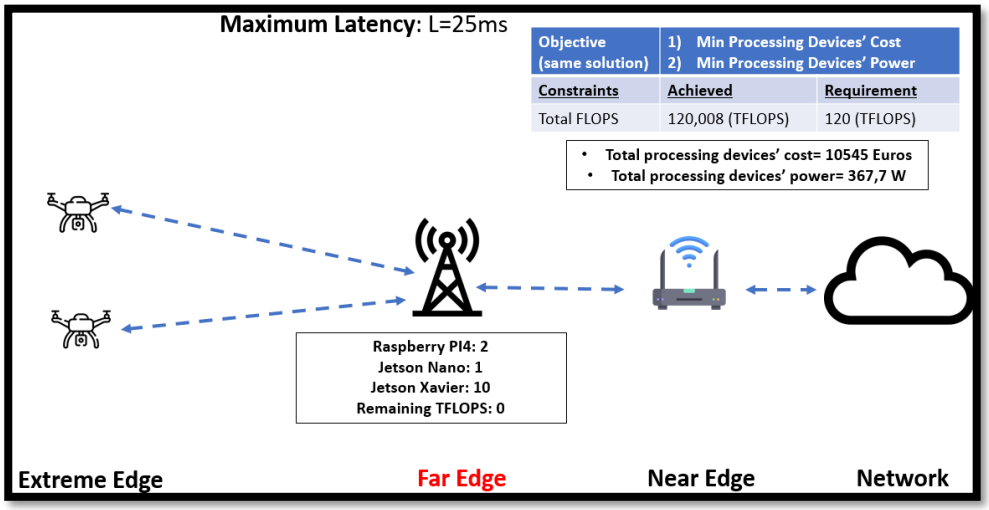
      max{Ceil(10^4/10^8),Ceil(50/400)}=**1**
    - Number of transport devices: $\mathbf{U_{TT}}$= Ceil(TD/$Throughput_{TT}$)= Ceil(50/500)=**1**
- Overall network configuration:



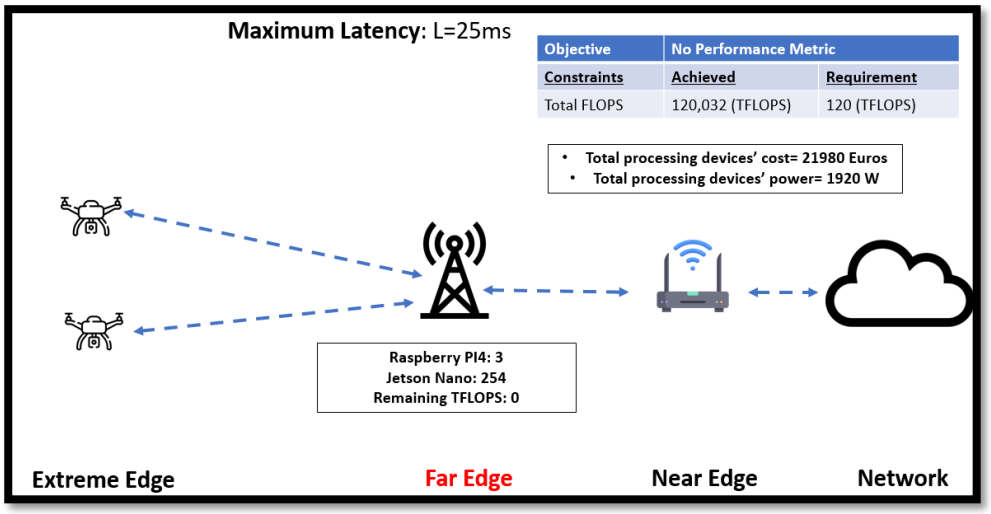
(a)


Maximum Latency: L=50ms
Objective
No Performance Metric
Constraints
Achieved
Requirement
Total FLOPS
120,384 (TFLOPS)
120 (TFLOPS)
Total processing devices' cost= 10370 Euros
Total processing devices' power= 356 W
Jetson Xavier: 21
Remaining TFLOPS: 110,088
Jetson Nano: 1
Jetson Xavier: 10
Serving TFLOPS: 110,472 (>110,088)
Remaining TFLOPS: 0
Extreme Edge
Far Edge
Near Edge
Network

(b)


Maximum Latency: L=25ms
Objective (same solution)
1) Min Processing Devices' Cost
2) Min Processing Devices' Power
Constraints
Achieved
Requirement
Total FLOPS
120,008 (TFLOPS)
120 (TFLOPS)
Total processing devices' cost= 10545 Euros
Total processing devices' power= 367,7 W
Raspberry PI4: 2
Jetson Nano: 1
Jetson Xavier: 10
Remaining TFLOPS: 0
Extreme Edge
Far Edge
Near Edge
Network

(c)



(d)

Figure 32: Solution of optimisation problem no3 for: (a) Maximum Latency =50ms and Cost and Power Objectives, (b) Maximum Latency =50ms and No Objective, (c) Maximum Latency =25ms and Cost and Power Objectives and (d) Maximum Latency =25ms and No Objective.

In Figure 32(a) we observe that in order to satisfy the latency and processing capacity requirements of the service, 10 Jetson Xavier are attached in the base station and 1 Jetson Xavier in the backhaul router. The same solution is achieved for both cost and power consumption minimisation objectives. In case that we do not minimise any objective (Figure 32(b)) 21 Jetson Nano are attached in the base station, as well as 10 Jetson Xavier and 1 Jetson Nano in the backhaul router. Considering a stricter latency requirement (i.e. 25 ms) for the service, in Figure 32(c), all the devices, i.e. 2 Raspberry PI 4, 21 Jetson Nano and 10 Jetson Xavier are attached in the base station for both the cost and power minimisation. Finally, in Figure 32(d) 4 Raspberry PI 4 and 254 Jetson Nano are attached to the base station for the same latency requirement.

Comparing Figure 32 (a) and (b) with Figure 32(c) and (d) it is obvious that when the latency constraint is relaxed, less power consumption and cost is achieved and also the near edge zone is used. Furthermore, the minimisation of a performance metric results in allocations with better performance compared to allocations without performance metric optimisation. These results are consistent with the conclusions in parts (A) and (B).

6 Conclusions

This deliverable constitutes the first version of the “Technology and edge processing enablers, and service optimisation solutions” (M10). In this deliverable, a thorough investigation and evaluation of the available connectivity solutions for rural connectivity environments was conducted. Furthermore, a rich portfolio of edge computing solutions was developed, to address the different set of requirements and challenges of applications in rural areas. Finally, algorithmic techniques were implemented for the end-to-end system design and service provisioning, taking into consideration the technological capabilities and limitations of both connectivity and edge computing enablers, along with (non-)functional requirements of WP2 Use Cases.

The second deliverable “Technology and edge processing enablers, and service optimisation solutions”, due M26, will be an updated version of the current deliverable. It will provide improvements and enhancements on the content of connectivity and edge computing solutions as well as more complex/sophisticated algorithmic techniques based on the requirements of the XGain’s living labs in WP5.

7 References

[1] Kantor Management Consultants, “The European Commission allocates €95,5 billion on Rural Development” – Online Article: https://www.kantor-group.eu/Kantor-rural-development

[2] World Health Organization – One health. Online: https://www.who.int/news-room/questions-and-answers/item/one-health

[3] International Telecommunication Union (ITU), “The last-mile Internet Connectivity Solutions Guide – Sustainable connectivity options of unconnected sites”, 2020, Online: https://www.itu.int/pub/D-TND-01

[4] International Telecommunication Union (ITU), “Telecommunications/ICTs for rural and remote areas, Output Report on ITU-D Question 5/1, Study period 2018-2021”, Online: https://www.itu.int/pub/D-STG-SG01.05.1-2021

[5] Michael Wang, “5G, C-RAN, and the Required Technology Breakthrough”, Online: https://medium.com/@miccowang/5g-c-ran-and-the-required-technology-breakthrough-a1b2babf774

[6] Stefan Pongratz, The Value of Low-Band 5G, August 6, 2021, DELL’ORO GROUP, Online: https://www.delloro.com/the-value-of-low-band-5g/

[7] Ronan McLaughlin, “5G Low Latency Requirements”, Broadband Library, Online: https://broadbandlibrary.com/5g-low-latency-requirements/

[8] Chen Dongxu, “5G Power: Creating a green grid that slashes costs, emissions & energy use”, Huawei, Online: https://www.huawei.com/en/huaweitech/publication/89/5g-power-green-grid-slashes-costs-emissions-energy-use

[9] Dappworks, “Front Line Data Study about 5G Power Consumption | You need to know about 5G”, https://dappworks.com/front-line-data-study-about-5g-power-consumption-you-need-to-know-about-5g/

[10] Qorvo, “5 Things to Consider When Designing Fixed Wireless Access (FWA) Systems”, Online: https://www.qorvo.com/design-hub/blog/5-things-to-consider-when-designing-fixed-wireless-access-fwa-systems

[11] Rupert Wood, “A joined-up approach to operators’ technology choices could deliver green benefits”, November 2021, Online: https://www.analysysmason.com/contentassets/82ca172602eb49d88bf6eca853b4c3e7/analysys_mason_green_access_technology_nov2021_rdfi0_rdnt0.pdf

[12] Deliberant, “2.4/5 GHz legacy point to multi point”, Online: http://www.deliberant.com/cn/point-to-multi-point

[13] Versa Technology, “Wi-Fi 7: The Next Generation”, 05 Jan, 2021, Online: https://versatek.com/wi-fi-7-the-next-generation/

[14] Sigfox 0G technology, Online: https://www.sigfox.com/

[15] LoRa Alliance, Online: https://lora-alliance.org/

[16] R. Sanchez-Iborra and M.D. Cano, "State of the art in LP-WAN solutions for industrial IoT services," Sensors 16(5), 2016.

[17] K. Mekki, et al., “A comparative study of LPWAN technologies for large-scale IoT deployment,” ICT express, 5(1), pp.1-7, 2019.

[18] C. Gomez, et al., “A sigfox energy consumption model,” Sensors. 19(3), 2019.

[19] M.I. Hossain and Jan I. Markendahl, "Comparison of LPWAN technologies: Cost structure and scalability." Wireless Personal Communications 121(1), pp. 887-903, 2021.

[20] Vladislav Yordanov, “Things to Consider When Picking A LoRaWAN® Gateway”, Online: https://news.rakwireless.com/things-to-consider-when-picking-a-lorawan-gateway/

[21] Starlink – Reddit account, Online: https://www.reddit.com/r/Starlink/comments/pkcgn4/starlink_power_consumption/

[22] Steve Baker, “Low-cost satellite IoT: around the corner, but no free lunch for operators”, Online: https://www.ttp.com/insights/low-cost-satellite-iot-around-the-corner-but-no-free-lunch-for-operators/
[23] Eugenio Pasqua, “Satellite IoT connectivity: Three key developments to drive the market size beyond $1 billion”, August 25, 2022, Online: https://iot-analytics.com/satellite-iot-connectivity/
[24] Astrocast, Online: https://docs.astrocast.com/dl/0532-DOC-M2M-ASTRO-Astronode_S-Datasheet.pdf
[25] Swarm, Online: https://swarm.space/product/swarm-m138-modem/
[26] Kineis, Online: https://www.kineis.com/wp-content/uploads/2021/03/EN-25-mars-French-Tech-EN.pdf
[27] EchoStarMobile, Online: https://www.echostarmobile.com/wp-content/uploads/2022/07/EchoStar-Mobile-LoRa-enabled-Network-EM2050-Datasheet.pdf
[28] Iridium, Online: https://www.iridium.com/products/iridium-9602/
[29] Myriota, Online: https://developer.myriota.com/assets/myriota-module-data-sheet-v1_5.pdf
[30] C. Lin, “Broadband Optical Access Networks and Fibre-to-the-Home: Systems Technologies and Deployment Strategies”, John Wiley & Sons, England, 2006
[31] Fibre Broadband Association, Online: http://www.ftthcouncil.org/
[32] C.F. Lam, “Passive Optical Networks: Principles and Practice”, Academic Press 2007
[33] Federal Communications Commission (FCC), “Millimeter Wave Propagation: Spectrum Management Implications”, July 1997
[34] ITU, “Specific attenuation model for rain for use in prediction methods”, Recommendation ITU-R P.838-3, (Question ITU-R 201/3)
[35] Faragó, Ákos, Kántor, P., Bitó, J. Z. “Rain effects on 5G millimeter wave ad-hoc mesh networks investigated with different rain models”, Periodica Polytechnica Electrical Engineering and Computer Science, 60(1), pp. 44–50, 2016. https://doi.org/10.3311/PPee.8644
[36] H. M. Rahim, C. Y. Leow and T. A. Rahman, "Millimeter wave propagation through foliage: Comparison of models," 2015 IEEE 12th Malaysia International Conference on Communications (MICC), Kuching, Malaysia, 2015, pp. 236-240, doi: 10.1109/MICC.2015.7725440.
[37] Weissberger, M. A., “An initial critical summary of models for predicting the attenuation of radio waves by foliage”, Electromagnetic Compatibility Analysis Center, Annapolis, MD, ECAC-TR-81-101, 1981.
[38] COST235, “Radio propagation effects on next-generation fixed-service terrestrial telecommunication systems,” Final Report, Luxembourg, 1996.
[39] Recommendation ITU-R P.833-9 (09/2016), Attenuation in vegetation, P Series Radiowave Propagation
[40] J. S. Lu, X. Han and H. L. Bertoni, "The Influence of Terrain Scattering on Radio Links in Hilly/Mountainous Regions," in IEEE Transactions on Antennas and Propagation, vol. 61, no. 3, pp. 1385-1395, March 2013, doi: 10.1109/TAP.2012.2231919.
[41] Wei Liao and Naveen Chilamkurti, "Edge Computing for Internet of Things: A Review," Journal of Ambient Intelligence and Humanized Computing, Volume 9, 2018.
[42] Intel, "The Compute Continuum: Seamless, Scalable Computing from Edge to Cloud," Intel, 2020, https://www.intel.com/content/www/us/en/internet-of-things/what-is-edge-computing.html
[43] Arm, "The compute continuum: A vision for computing," 2021, Online: https://www.arm.com/why-arm/compute-continuum
[44] Sunku Ranganath, “4 Types of Edge Computing — Broadly Categorized”, Online: https://itnext.io/4-types-of-edge-computing-broadly-categorized-be557f3556ea
[45] Docker, "Docker: The Leading Container Platform," Docker, 2021, Online: https://www.docker.com/
[46] Kubernetes, "Kubernetes: An open-source platform for automating deployment, scaling, and operations," Kubernetes, 2021, Online: https://kubernetes.io/
[47] VMware, "VMware vSphere: The World's Leading Virtualization Platform," VMware, 2021, Online: https://www.vmware.com/products/vsphere.html

[48] OpenStack, “OpenStack: The most Widely Deployed Open Source Cloud Software in the World”, OpenStack 2023, Online: https://www.openstack.org/
[49] Unikernels, "Unikernels: A new approach to virtualization," Unikernels, 2021, Online: https://unikernels.org/
[50] D. Chopra, "Why Containers are Better than VMs for Edge Computing," Container Journal, 2020, https://containerjournal.com/topics/container-journal/why-containers-are-better-than-vms-for-edge-computing/
[51] P. Chacko, "Containers vs. Unikernels: Which is the Best Fit for Edge Computing?" Packt, 2020, https://hub.packtpub.com/containers-vs-unikernels-which-is-the-best-fit-for-edge-computing/
[52] Dogo, E.M., Salami, A.F., Aigbavboa, C.O., Nkonyana, T. (2019). Taking Cloud Computing to the Extreme Edge: A Review of Mist Computing for Smart Cities and Industry 4.0 in Africa. In: Al-Turjman, F. (eds) Edge Computing. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-319-99061-3_7
[53] J. Portilla, G. Mujica, J. -S. Lee and T. Riesgo, "The Extreme Edge at the Bottom of the Internet of Things: A Review," in IEEE Sensors Journal, vol. 19, no. 9, pp. 3179-3190, 1 May1, 2019, doi: 10.1109/JSEN.2019.2891911.
[54] Social Compare, “RaspberryPI models comparison”, Online: https://socialcompare.com/en/comparison/raspberrypi-models-comparison
[55] Radxa, “Introduce ROCK 4”, Online: https://rockpi.org/rockpi4
[56] Tinkerboard, “ ASUS Tinker Board Series”, Online: https://tinker-board.asus.com/series.html
[57] Hardkernel, “ODROID-N2L with 4GByte RAM”, Online: https://www.hardkernel.com/shop/odroid-n2l-with-4gbyte-ram/
[58] Nvidia, “NVIDIA Jetson AGX Orin Series”, Technical Brief v1.2 | July 2022, Online: https://www.nvidia.com/content/dam/en-zz/Solutions/gtcf21/jetson-orin/nvidia-jetson-agx-orin-technical-brief.pdf
[59] Coolblue, “The differences between a mini PC and a stick PC”, Online: https://www.coolblue.be/en/advice/difference-between-mini-pc-and-stick-pc.html
[60] NHR ENGINEERING & SERVICES SDN BHD, “Tower Servers”, Online: http://nhr-engineeringservices.com/index.php/tower-servers-servers/
[61] M. Capra, B. Bussolino, A. Marchisio, G. Masera, M. Martina and M. Shafique, "Hardware and Software Optimizations for Accelerating Deep Neural Networks: Survey of Current Trends, Challenges, and the Road Ahead," in IEEE Access, vol. 8, pp. 225134-225180, 2020, doi: 10.1109/ACCESS.2020.3039858.
[62] Techpowerup, Online: https://www.techpowerup.com/
[63] Pcpartpicker, Online: https://de.pcpartpicker.com/
[64]“Upload Speeds for 4K Live Streaming”, Online: https://www.boxcast.com/blog/internet-speeds-for-4k-live-streaming