Alliance for IoT
and Edge Computing
Innovation

Computing Continuum Scenarios, Proof of Concepts, Requirements and Optical Communication enablers

Release 2.0

AIOTI WG Standardisation

2 May 2024

Executive Summary

This report introduces Computing Continuum use cases, requirements and KPIs on communication infrastructures, IoT and edge computing platforms. Compared to many current activities, the computing continuum enables a more flexible allocation of compute and communication resources and workload placement. Many novel applications require rather stringent KPIs since the IoT becomes more and more mission-critical. The new system requirements include strong security, very high bandwidth, very low delays, and very high reliability. Depending on the use case and deployment scenario, various technology enablers are currently under standardization, including the F5G optical network architecture, as well as novel approaches for computing, networking and establishing security.

Regarding computing continuum platforms, high-performance secure computing together with optical communication can be considered as an ideal combination fulfilling the high-end IoT requirements. In fact, many basic requirements for high-end IoT can be satisfied by optical communication enablers. Such enablers can also help meeting stringent communication requirements stemming from the distributed nature of the continuum including multiple, potentially different, computing locations.

Furthermore, high-end IoT devices, that may be connected via optical communication technologies, can enable a whole new application area to be explored and supported. For example, some use cases might need very high resolution and high frame rate of camera sensors that send uncompressed video to AI-enabled analytics platforms with minimum delay. These analytics platforms typically need to react fast on any given situation and steer actors appropriately.

The proof of concepts validates the great potential of passive optical networks for new use cases such as fibre to the room, but also for industrial applications where low latency In combination with Wi-Fi is required.

Recommendations for the evolution of the current technologies employed for high-end IoT systems running on computing continuum platforms are discussed.

Finally, it is shown that passive optical networks play an important role for decarbonizing vertical sectors due to their improved footprint over conventional switch-based networking solutions, especially in scenarios with large numbers of access points.

Table of Content

Table of Figures

List of Tables

Acronyms

AggN	Aggregation Node
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AP	Access Point
AR	Augmented Reality
BNG	Broadband Network Gateway
BSS	Basic Service Set
BW	Bandwidth
CAD	Computer-Aided Design
CCTV	Closed Circuit Television
CPE	Customer Premises Equipment
CPN	Customer Premises Network
CPN-A	Customer Premise Network-Aggregator
CPU	Central Processing Unit
CS	Computation Service
CT	Computer Tomography
DBA	Dynamic Bandwidth Allocation
DC	Data Centre
DDS	Data Distribution Service
DICOM	Digital Imaging and Communications in Medicine
DPI	Deep Package Inspection
DSA	Digital Subtraction Angiography
DSP	Digital Signal Processing
E2E	End-to-End
EC	Edge Cloud
ECO$_2$	Emitted CO_2
E-CPE	Edge CPE
EMS	Element Management System
ETH	Ethernet
ETSI	European Telecommunications Standards Institute
EV	Electric Vehicle
F5G	Fifth Generation Fixed Network
F5G-A	F5G Advanced
FEC	Forward Error Correction
fgOTN	fine-grain OTN
FTTH	Fibre to the Home
FTTO	Fibre to the Office
FTTM	Fibre to the Machine
FTTR	Fibre to the Room
GE, GigE	Gigabit Ethernet
GPON	Gigabit Passive Optical Network
GPU	Graphics Processing Units
GPRS	General Packet Radio Service
GW	Gateway
HMI	Human Machine Interface
ICT	Information and Communications Technology
IFDN	Indoor fibre Distributed Network
IPC	Industrial PC
ISG	International Study Group
IoT	Internet of Things
IT	Information Technology
ITU	International Telecommunication Union

KPI	Key Performance Indicator
LAN	Local Area Network
LCA	Life Cycle Assessment
LEA	Law Enforcement Agent
LiFi	Light Fidelity
MFU	Main FTTR Unit
ML	Machine Learning
MPLS	Multiprotocol Label Switching
MRI	Magnetic Resource Imaging
MTBF	Mean Time Between Failure
MQTT	Message Queuing Telemetry Transport
M&C	Management and Control
NCIe	Network Carbon Intensity energy
NFV	Network Function Virtualisation
OAA	Observe-Analyse-Act
O-E-O	Optical-Electrical-Optical
OLT	Optical Line Terminal
ONT	Optical Network Termination
ONU	Optical Network Unit
OSM	Open Source MANO
OTN	Optical Transport Network
OWC	Optical Wireless Communication
P2MP	Point-to-Multipoint
P2P	Point-to-Point
PACS	Picture Archiving and Communication System
PC	Personal Computer
PDU	Power Distribution Unit
PE	Provider Edge
PON	Passive Optical Network
PPPoE	Point-to-Point Protocol over Ethernet
QoE	Quality of Experience
QoS	Quality of Service
RaaS	Robotics as a Service
RIS	Radiology Information System
RF	Radio Frequency
RMS	Remote Management System
ROS	Robot Operating System
SBI	South Bound Interface
SFU	Subordinate FTTR Unit
SME	Small and Medium-Sized Enterprise
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TPU	Tensor Processing Unit
UBP	Urban Platform
UDP	User Datagram Protocol
USB	Universal Serial Bus
VIS	Visual Inspection Station
VNF	Virtualized Network Function
vPLC	virtual Programmable Logic Controller
VR	Virtual Reality
WDM	Wavelength Division Multiplexing
WiFi	Wireless Fidelity
XGS-PON	10 Gigabit Symmetrical PON

1 Introduction

This report introduces the Computing Continuum use cases, requirements and KPIs on communication infrastructures, IoT and edge computing platforms and describes optical communication enablers that can meet these KPIs and requirements.

Due to the huge increase of connected devices and systems, several computing deployments are embracing the notion of computing continuum, where the right compute resources are placed at optimal processing points, i.e., cloud data centre, edge computing systems and end devices.

Currently, due to the near real-time decisions that are directly affecting the operation of, e.g., buildings and homes, transportation, factories, cities, it is required that computing resources are fast, efficient, secure and are located both near the data sources for time-sensitive tasks, and farther away enough for intensive computations and data aggregation.

However, challenges arise for the situations that the near real-time decisions need to collect simultaneously the data processed in the different processing/computing points, i.e., cloud data centre, edge computing systems and end devices.

In this situation, the underlying communication technologies, connecting these processing points that are typically distributed over large distances (e.g. cross-boarders) need to support low latency and high bandwidth requirements.

Also, it is assumed that some IoT sensors and actors require high bandwidth connectivity to the nearest possible place to compute.

2. Use Cases

This section focuses on identifying the computing continuum requirements and KPIs that are imposed to the underlying infrastructure. The derivation of these requirements is based on computing continuum use cases and related literature.

These requirements, which are an output of this document, will be used as input to define the KPIs for the network connecting edge computing platforms and cloud. This section focuses on applications in medical, industrial and smart city environments.

2.1 Cloud-based medical imaging

Description

The cloudification of medical imaging uses systems such as Picture Archiving and Communication Systems (PACS) or Radiology Information Systems (RIS). To ensure optimal experience, the image system requires high bandwidth, low latency, low packet loss rate, high security, high reliability, and flexible scheduling capabilities. This use case describes key components and service data flows in the cloud-based medical imaging system.

PACS is a medical imaging technology, which provides storage and convenient access to images from multiple medical imaging equipment. Electronic images and reports are transmitted digitally via PACS. The universal format for PACS image storage and transfer is the Digital Imaging and Communications in Medicine (DICOM®) which is the standard for the communication and management of medical imaging information and related data.

PACS consists of four major components:

- The imaging equipment such as X-ray plain film, Computed Tomography (CT) and Magnetic Resource Imaging (MRI)
- Secured network for the transmission of patient information
- Workstations for interpreting and reviewing images
- Archives for the storage and retrieval of images and reports

The migrating of medical images to the cloud allows for remote access and all-round PACS services for medical institutions. It also allows for resource sharing required for Artificial Intelligence (AI) based image processing. Figure 1 gives a high-level view.

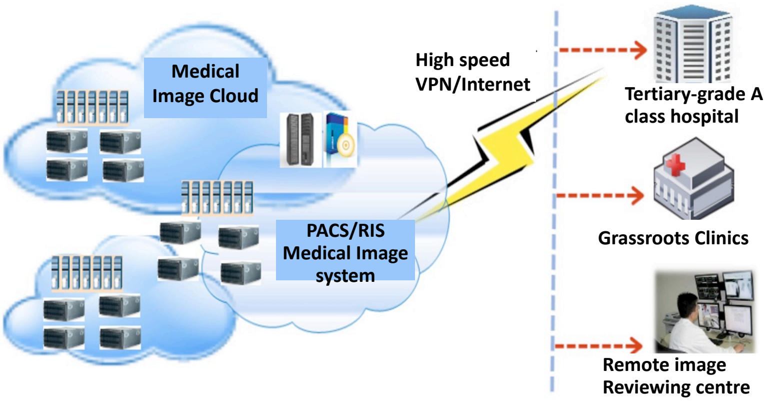


Figure 1: Medical image migration to the cloud.

This migration provides a wide range of applications such as medical image data storage, image retrieval via a doctor's desktop and mobile terminals, diagnosis and treatment assistance and training material for teaching at medical institutions.

The medical image cloud provides medical image data back up and archiving, which provides complete, fast and efficient services of image data collection, conversion, integration, storage, verification and access control.

Medical image cloud provides services for remote consultation, imaging specialist diagnosis, image teaching, mobile image reading/consultation and image big data analysis services. These services enable medical personnel to quickly query and search medical records, improving their work and scientific research efficiency.

The medical image cloud provides necessary resources for AI-based image analytics.

Source

- [ETSI GR F5G 008 V1.1.1 (2022-06), "F5G Use Cases Release #2"]()
- [Digital Imaging and Communication in Medicine]()

Roles and Actors

- Imaging Cloud Provider: Is providing the imaging system as a service.
- Hospitals and other medical institutions: User of the imaging system.

Pre-conditions

The assumption of this use case is to move the imaging system to the cloud such that the images are better accessible, sharable under security constraints and viewable independent of the location.

Triggers

Any of the actions in the flow (see below) is triggered through an image creation, image processing, or image retrieval action.

Normal flow

Figure 2 illustrates the key components (A1 to A4, N1 to N2) and the main data flows (F1 to F3) involved in the cloud-based medical image migration. Different imaging types generate different size image data. Patient's image data can be as high as 2 GB.

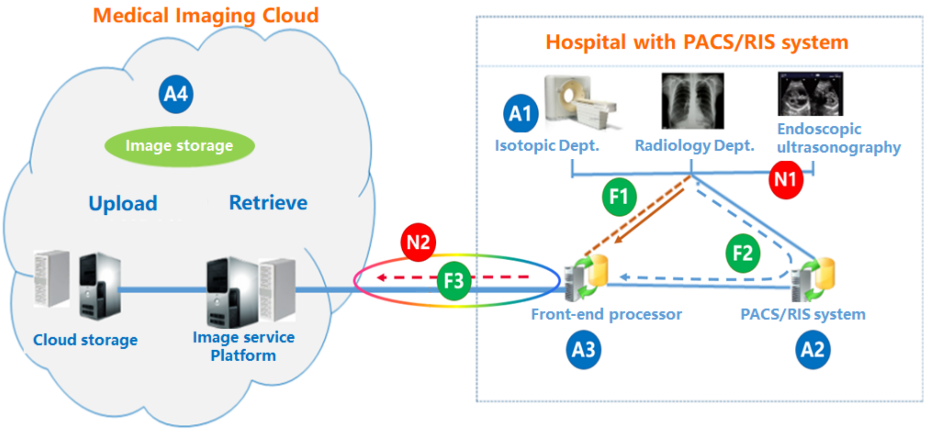


Figure 2: Key components and data flows in the cloud-based medical image migration.

The followings are the key components in the cloud-based medical image migration:

- A1 - image terminal: It may generate image data in either DICOM® or non-DICOM® format. Non-DICOM® format is converted to DICOM® format by the front-end processor before upload.
- A2 - medical imaging system: It is an IT system that stores medical image data locally in the hospital.
- A3 - image cloud front-end processor: It processes the local image data before upload to the image cloud. The front-end processor mainly performs the following functions:
  - Image transmission and backup
  - Providing a temporary local PACS for the hospital when the cloud PACS network has a fault
  - Non-standard DICOM® image conversion
- A4 - medical imaging cloud: It is a Data Centre (DC) where compute and storage servers are deployed to provide cloud storage and cloud retrieval services.
- N1 network: It is a Local Area Network (LAN) in the hospital campus with a coverage area of several km^2 and provides communication service within the hospital.
- N2 network: It is the hospital cloudification network, which connect the image cloud front-end processor and image cloud. N2 Network is either owned by the hospital itself or provided by a network operator.

The following are the main data flows in the cloud-based medical image migration:

- F1 data flow: For hospitals without local medical image storage systems, the data flow generated by image terminals is sent directly to the front-end processors deployed in hospitals.
- F2 data flow: For a hospital that has a local medical image storage system, the data flow generated by the image terminal is first stored in the local medical image storage system. Then the data is sent to the front-end processor deployed in the hospital.
- F3 data flow: After local image data is processed by the front-end processor in the hospital, the image data is uploaded to the medical imaging cloud deployed outside the hospital.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

Figure 2.

Potential Requirements

Data security and privacy: Over and above the security provided by PACS, consideration should be given to data security. The hospital campus communication system is LAN based and the internal link data security may be considered. The connection from the front-end processor to the DC may also need data level and or link level security.

Flexible bandwidth allocation: Hospital have regular visit from non-resident specialist consultants, these consultants move from hospital to hospital on a regular basis. These onsite visits can increase the demand on cloud service for the duration of their visit. This will require temporarily increase in bandwidth to meet the additional cloud data access for the patient's image data. This additional bandwidth is only required for the duration of the visit and the available bandwidth should return to normal once the visits are over to keep hospital IT cost down. This means the network service provider or operator needs to support flexible bandwidth allocation to match the needs of the hospital.

High reliability: Since medical imaging depends on the situation, there are situations, where the reliability and time to receive an image does not matter, for example, in the case when a visit is well planned and images can be pre-loaded or the doctor already checks the images before the patient visit. However, there are situation, where receiving the images on-time is mission critical and might be lifesaving. For the latter cases, it is mission-critical to have the medical imaging system up and running all the time. This includes very low latency of the networking components from the cloud to the terminal, such that the images are available reliably on time. For example, in emergency situations or surgery, the medical images created in other medical institutions requires to be available and visible immediately. Also in cases, where remote surgeries are done, or patients are handled at different locations, the images require to be at the place of surgery quickly.

High performance requirements: Digitalized medical images require high accuracy and need to meet the diagnostic-level image quality requirements. To ensure the quality of medical images, it is recommended that the image data should not be compressed for transmission and storage (or only loss-less compression is enabled). Therefore, the network bandwidth and storage requirements are high.

The image sizes are very dependent on the type of image, the image creation equipment, and the number of images needed for 3D pictures (one picture per slice in case of a CT). For details refer to the PACS storage and network calculator.

As a rough estimate, the medical image sizes currently used are typically in the order of:

- CT, MRI images: 100 MByte – 1 GByte
- DSA images: 10 GByte

However, the imaging technologies improve over time and higher-data volumes can be expected in the future.

Given a certain medical image data size, the network bandwidth can directly affect the transmission efficiency of the image data to and from the cloud. Network degradation, such as network packet loss and increased network latency, can slow down image transmission, and prolong transmission time This also affects the image data transmission experience to and from the cloud.

Depending on the size of the imaging department and the number of medical personnel in the hospital, the number of patients that the hospital can process may vary with the size of the hospital. Table 1 lists some suggested network bandwidths for hospitals of different sizes to access the imaging cloud.

Optical Network specific Requirements

Table 1: Network bandwidths of hospitals with different scales

Hospital Size	Daily patients/visits	Image and image reading terminal in the hospital	Network bandwidth for image storage to the cloud in Mbit/s
Large hospital	20,000	2,000	15,840
Medium-sized hospital	7,000	800	6,336
Small hospital	1,000	100	792

Note that these numbers assume only the imaging part. In case of (remote) surgery scenarios also video is required and needs to be calculated on top. Also, for regulatory reasons some videos need to be stored for later use as proof or teaching material. The surgery- and video-oriented use cases are different compared to this one and can be handled in a different use case.

Notably, high Quality of Experience (QoE) of users, e. g., doctors and hospital staff, is very important and time-sensitive when browsing through large sets of images to avoid delays and display medical images instantly. The use of computing continuum technologies enables and improves such requirements.

2.2 Cloud-based visual inspection in production

Description

Background: This use case focuses on a particular aspect of industrial production processes, namely the quality assurance supported by visual inspection based on video analytics. A scenario is considered, where a closed control loop is desired between video cameras at factory shop floors, edge computing resources hosting the video analytics and control functions to control robotic actors at the factory shop floors (see Figure 3).

Business drivers and motivation: The current trend in the industry goes towards virtualization of control functions in the form of virtual Programmable Logic Controllers (vPLCs), which are hosted in edge cloud environments. This has the benefit of using standard off-the-shelve IT hardware in a dedicated environment instead of ruggedized and specially hardware IT components, which can operate directly in the production environment. Employing edge cloud solutions connected via a real-time communication network to the factory shop floor offers new, economically highly attractive possibilities, especially for smaller manufacturing companies due to less infrastructure and acquisition costs. However, outsourcing of control functions to edge clouds may add particular requirements on the networking infrastructure between production lines and edge compute resources.

Operation of the use case: An overview on the operation of the use case is provided in Figure 3. Video streams of industrial-grade video cameras are transported in real-time to an edge DC. Video analytics solutions extract metrics to estimate the quality of the produced parts. These metrics are fed to the virtual control logic to provide automatic quality control measures on the factory shop floor by directly controlling robotic actors over a time-sensitive network connection supporting the required industrial Ethernet protocols.

Source

- ETSI GR F5G 008 V1.1.1 (2022-06), "F5G Use Cases Release #2"

Roles and Actors

- Actors/Parties
    - Large corporations, small and medium sized enterprises
- Roles
    - Factory owner/vertical industry:

Runs productions lines with video sources and robotic actors, benefits from quality assessment.

- Edge Cloud Service Provider:

Provides edge cloud resources to the use case, this may comprise both hardware and software and different service levels such as e. g. infrastructure-as-a-service, platform-as-a-service or software-as-a-service.

- Communication Service Provider:

Provides the communication between factory and edge DC.

Pre-conditions

- Edge DC (e.g. on-premise edge or colocation edge)
- Edge cloud environment to host video analytics and the virtual control logic
- Real-time capable/time-sensitive communication between factory shop floor and edge DC

Triggers

- The use case is triggered when a new vision inspection station is introduced into the production process.

Normal Flow

- Produced parts at the production lines are monitored by cameras acting as video sources.
- Video streams from the video sources are transported in real-time over a time-sensitive network to an edge DC where the edge cloud environment hosts the video analytics and virtual control logic functions.
- The video analytics service performs assessment of the quality of the produced parts and reports the resulting metrics to the virtual control logic.
- In case that regulatory action is required, a vPLC communicates the appropriate control signals via a time-sensitive network to the robotic actor at the production line.
- The robotic actor performs the required regulatory action on the produced parts. This completes the control loop.

Alternative Flow

N/A.

Post-conditions

There are no post-conditions as long as production is running and quality is assessed.

High Level Illustration

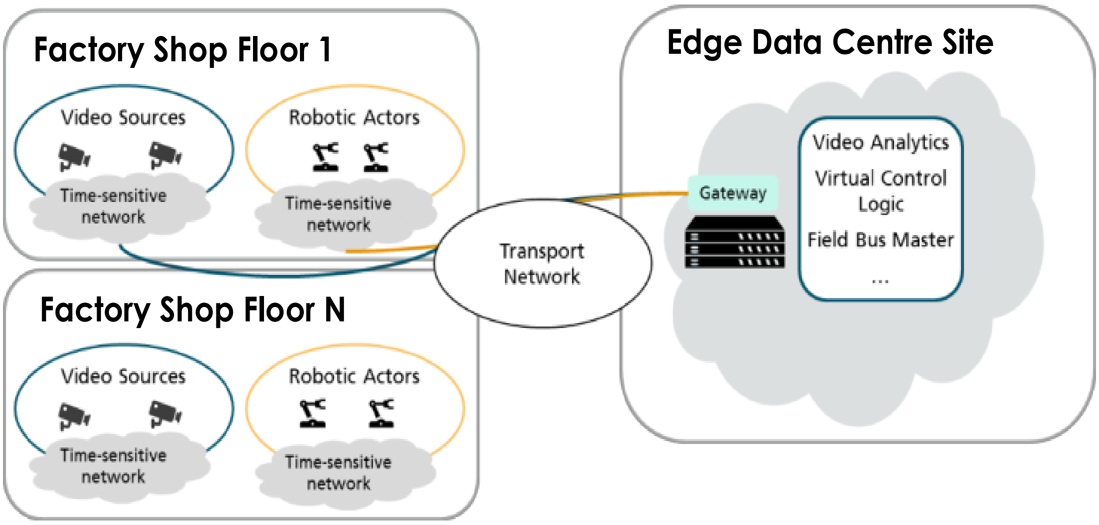


Figure 3: Schematic depiction of the visual inspection in production use case.

Potential Requirements

<u>Functional Requirements</u>

- Reliable communication between video sources, edge DC and robotic actors.
- High-bandwidth connectivity between production line and edge DC to provide significant data rates in the upstream (typically 1 Gbit/s to 20 Gbit/s per vision inspection station).
- Isochronous, low latency and deterministic communication between video sources, edge DC and robotic actors supporting low cycle times (s. Table 2).

<u>Non-Functional Requirements</u>

- Interoperability with industrial Ethernet standards (e. g. Ethernet/IP, PROFINET, Sercos III).
- Secure connection between production facilities and edge DC.
- Guaranteed data privacy.

Optical Network specific Requirements

Table 2: Target KPIs for cloud-based visual inspection for automatic quality assessment in production.

Target KPI	Value
Upstream data rate per vision inspection station	1 Gbit/s (single GigE Vision camera) – 20 Gbit/s (4× USB3 Vision cameras)
Downstream data rate per vision inspection station	> 400 kbit/s (control signals only)
End-to-End (E2E) cycle time*	5 - 10 ms typical < 2 ms time-critical scenarios
Reach (max. distance to edge DC)	< 80 km

**cycle time is determined by the time required for the vPLC to send all control signals to its assigned targets and to receive all of their feedback in return*

2.3 Cloud-based control of automated guided vehicles

Description

Background, business drivers and motivation: Modern production facilities have to support on-demand product customization to satisfy customer needs. This can be enabled by making the manufacturing of small lot sizes very cost-efficient. One key technology to make this happen are Automated Guided Vehicles (AGVs). These are mobile transport robots, which distribute raw materials and parts on the factory shop floor and potentially among different manufacturing halls and warehouses. The navigation of the AGVs on the factory shop floor or in outdoor areas is a computationally complex task requiring significant computing resources. In order to save battery life on the AGVs and minimize down-times for loading, navigation and control algorithms are often offloaded to an edge cloud, which can provide sufficient computing resources (e. g. Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs) for acceleration of AI-tasks). Additionally, cloud-based AGV navigation enables cooperation and centralized information exchange between multiple robots and AGVs.

Operation of the use case: A high-level overview on the operation of the use case is provided in Figure 4. On the hardware layer, AGVs transport goods, materials and other objects to and between robotic production cells. The hardware layer is governed by the service layer, where production processes are flexibly described by sets of microservices, managed by process controllers. The service layer, containing services of the production cells, AGVs, navigation, guidance control systems and so on is hosted on an edge cloud. The connectivity between AGVs, production cells and edge DC is provided by a combination of wireless and wireline networks. The connection must be highly reliable and provide an E2E roundtrip latency of less than 30 ms.

Source

- ETSI GR F5G 008 V1.1.1 (2022-06), "F5G Use Cases Release #2"

Roles and Actors

- Actors/Parties
    - Large corporations, small and medium-sized enterprises
- Roles

  - Factory owner/vertical industry:

Owns productions facilities and controls the hardware layer (i. e. AGVs, production cells and so on). Controls and configures service layer.

  - Edge Cloud Service Provider:

Provides edge cloud resources to the use case, this may comprise both hardware and software and different service levels such as, e. g., infrastructure-as-a-service, platform-as-a-service or software-as-a-service.

  - Communication Service Provider: Provides the communication between factory and edge DC.

Pre-conditions

- Reliable wireless network for AGVs (e. g. 5G, WiFi, LiFi)
  - Current and upcoming generations of WiFi: WiFi 6 and WiFi 7
  - LiFi depends on the availability of LoS channel between the AGV and at least one LiFi access point
- Edge DC (e.g. on-premises edge or colocation edge)
- Edge cloud environment to host the service layer
- Reliable, low-latency communication between AGVs and edge DC

Triggers

- The use case is triggered when new production processes are introduced or running processes are changed (e. g. onboarding of new AGVs).

Normal Flow

- AGV communicates its sensor data to the service layer.
- Process information, navigation and guidance control systems in the service layer are updated and control information for the AGV is generated.
- Control information is communicated back to the AGV.
- AGV performs the required actions.

Alternative Flow

N/A.

Post-conditions

There are no post-conditions.

High Level Illustration

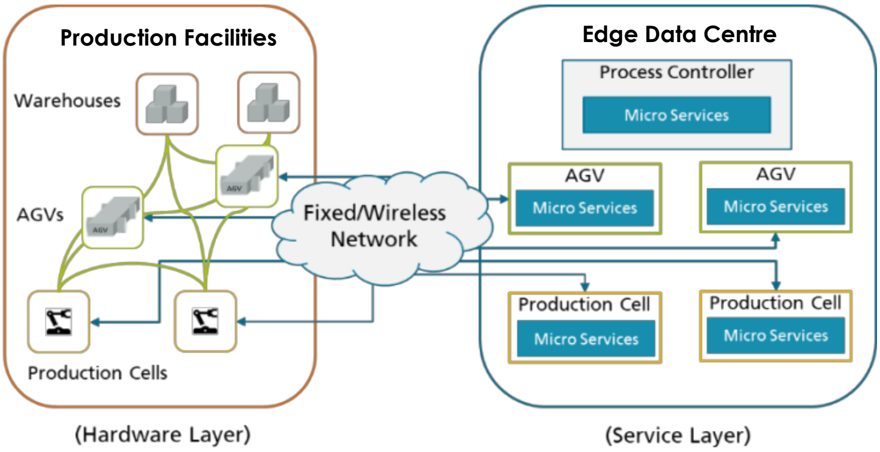


Figure 4: Schematic depiction of the automated guided vehicles use case.

Potential Requirements

Functional Requirements

- Low-latency wireless/wireline communication between AGV and edge DC (s. Table 3).
- Reliable communication between AGV, production cells and edge DC.
- Cyclic data communication with 10 – 50 ms cycle time.

Non-Functional Requirements

- Interoperability with industrial Ethernet standards (e.g. Ethernet/IP, PROFINET, Sercos III).
- Secure connection between production facilities and edge DC.
- Guaranteed data privacy.

Optical Network specific Requirements

Table 3: Target KPIs for cloud-based control of automated guided vehicles.

Target KPI	Value
Upstream data rate from AGV to edge	> 400 kbit/s per AGV > 10 Mbit/s per AGV in case of video upstream
Downstream data rate from edge to AGV	> 400 kbit/s per AGV
E2E roundtrip latency	< 30 ms*
Reach (max. distance to edge DC)	< 80 km

**including processing time at edge DC*

2.4 Cloud-based control of production via optical wireless communication

Description

Background, business drivers and motivation: The increasing digitalization of the production and future smart factory approaches (Industry 4.0) demand for a reliable wireless communication infrastructure. However, this wireless infrastructure must fulfil the quality standards of currently used cable connections. Due to electromagnetic interference in loaded environments, e. g., the use of radio based mobile communication systems in production halls can be very challenging. As commonly used radio waves can be detected far beyond the area of the actual operating site, opening a physical gate for hacking or jamming.

Optical Wireless Communication (OWC) systems use light as the communication medium. OWC is very well suited for dense deployments, with data rate per area factors 10-times higher compared to WiFi due to the possibility of sharply limited communication cells. OWC is inherently robust against EMI, as it operates in the optical spectrum. OWC can provide a complement to existing radio-based infrastructures without any interference. Additional features like sub-centimetre positioning have already been demonstrated.

Operation of the use case:

- Use case autonomous vehicle or mobile robot: Movement is monitored and/or tracked via OWC, operations are performed according to continuously updated schemes.
- Use case Augmented Reality (AR) / Virtual Reality (VR) based maintenance: Bidirectional data exchange between the end user device (e. g. Microsoft HoloLens) and the company server and/or remote company sites for remote maintenance or production support.
- Use case flexible production floor: OWC provides bidirectional data exchange between production systems and company data server.

A high-level overview on the operation of the use cases is provided in Figure 5 and Figure 6.

Source

- H2020 Project ELIOT
- SESAM - Sichere, softwarebasierte Zugangsnetze für die intelligente Fabrik von morgen (ip45g.de)
- The Light Communications Alliance is an association of companies, as well as academia, involved in OWC systems.

Roles and Actors

- End User: Industry production
- Responsibility on the production site: IT team, production quality team
- OWC system must be embedded seamlessly in IT-infrastructure and must fulfil Industry 4.0 requirements

Pre-conditions

- OWC systems can be installed in parallel to existing infrastructure and exchange data with the factory network. System architecture is similar to WiFi deployment.
- OWC Access Points (AP) must be deployed in the production area, in order to provide sufficient area coverage
- As the corresponding standardization is still under development, a seamless handover between WiFi and OWC systems needs to be provided for as a separate solution.

Triggers

- Evolution of industry production (Industry 4.0)
- Accelerated use of sensors
- Increase in system mobility (autonomous vehicles, mobile robots)
- Wish for better flexibility of system positioning on the production floor

Normal Flow

- On a general level, there is a bidirectional data exchange between the company cloud and the end user device.
- Use case autonomous vehicle or mobile robot: Movement is monitored and/or tracked via OWC. At the final position, operations are performed according to continuously updated schemes, allowing for a high production flexibility ("lot size = 1"). Operation data (visual material, other) are sent back to company data server for quality control.
- Use case AR/VR-based maintenance: Bidirectional data exchange between the end user device (e. g. Microsoft HoloLens) and the company server and/or remote company sites for remote maintenance or production support. Documents necessary for maintenance or operation (e. g. operation manual, CAD schemes, circuit schemes etc.) are sent to the end user device. Visual material is sent back to company data server and/or remote company site.
- Use case flexible production floor: OWC provides bidirectional data exchange between production systems and company data server. Thus, production system position can be varied without extra data cabling installation.

Alternative Flow

- Mobility on the production floor requires a wireless communication solution.
- If WiFi is accepted/available, a parallel OWC/WiFi operation can be considered. OWC would then be implemented in areas with high data density.
- For positioning/tracking of movements, camera based solutions can be considered, as well.

Post-conditions

Data exchange is continuous on a production floor.

High Level Illustration

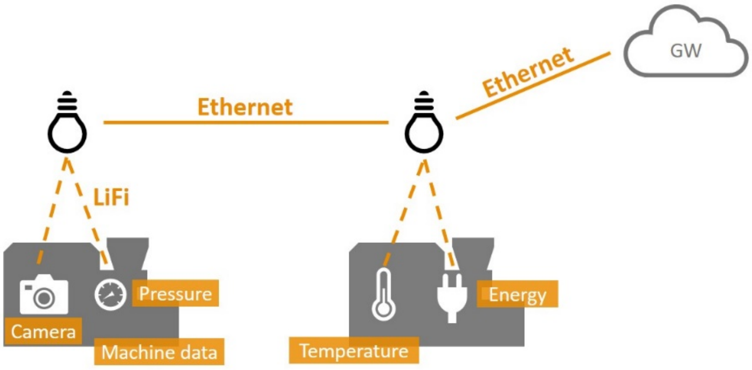


Figure 5: Schematic view of robust and reliable communication via OWC cells in IoT network: Machines are wirelessly connected via the OWC access points with the cable-based Ethernet network.

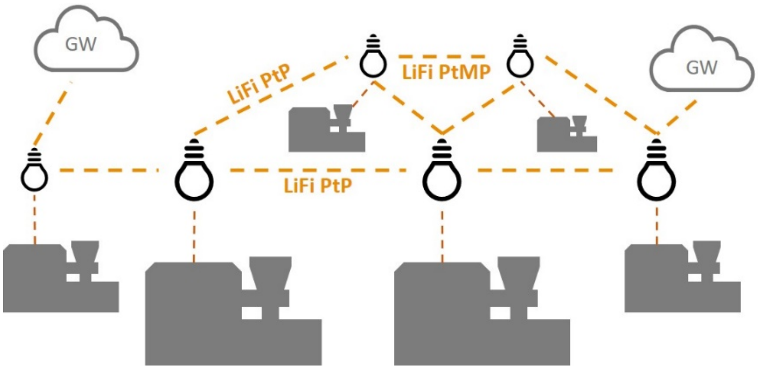


Figure 6: Schematic view of OWC system implementation for a flexible production floor: Production devices are wirelessly connected via point-to-point (P2P) and/or point-to-multipoint (P2MP) connections.

Potential Requirements

Functional requirements

- OWC needs to cover at least the area of one machine. Optionally, full area coverage through OWC
- Achievable data rates need to fulfil usual machine operations
- OWC system must offer “no link loss”
- Latency values of about 10 ms appear sufficient for most applications. However, latency jitter must be minimized.
- TCP/IP + UDP/IP are considered for real-time Ethernet protocol

Non-Functional requirements

- OWC cells need to scale with the operation area. Seamless handover between neighbour OWC cells needs to be provided for.
- Failure of OWC needs to be covered by an alternative (if necessary low bit rate) data connection (e. g. WiFi).
- In case of parallel use of OWC with RF-based mobile communication, seamless handover between the two protocols is necessary.

Optical Network specific Requirements

Table 4: Target KPIs for cloud-based control of industrial production via OWC.

Target KPI	Value
OWC cell (coverage area)	4 m x 5.5 m x 5 m (height x width x length)
Minimum achievable speed inside an OWC cell	100 Mbit/s
Minimum achievable speed in backhaul	1 Gbit/s
E2E roundtrip latency	< 10 ms*

2.5 Protecting sensitive data within smart cities

Description

Background: FogProtect's "Smart Cities" use case describes a network of Closed Circuit Television (CCTV) cameras that monitor selected places of a city, typically installed in connected street furniture, such as the smart lampposts (a modular lamppost capable of supporting different modules such as cameras, fog nodes, small cell antennas, electric vehicle (EV) chargers, etc.). For this use case, smart lampposts are equipping with fog nodes (computing units at the edge of the network) that run processes to anonymise sensitive data from the videos recorded by CCTV cameras (for example, by blurring peoples' faces and vehicle license plates). The ultimate goal is to process the sensitive data before it goes through the Internet, maintain citizens' trust in the system. Given that street furniture is vulnerable to physical attacks or other severe conditions, it is very important to implement the right tools in order to protect the data within the system.

Business drivers and motivation: A CCTV system is crucial for municipal entities, such as the city's decision-makers and operational staff as well as first-responders, to quickly understand what is happening within the urban environment and react accordingly. By embedding this system in smart lampposts, one can not only install the cameras and obtain footage but can use local fog nodes to process the videos and anonymise sensitive data. On the one hand, it allows the distribution of the processing power throughout the fog, while on the other, it allows data processing before uploading it to the cloud, helping to preserve everyone's privacy.

Operation of the use case: The use case is shown in

Figure 7 and Figure 8. Citizens can use mobile apps to report occurrences or incidents within the urban environment. These are pushed to an Urban Platform (UBP), hosted in the cloud. City operators can then request video footage of the location of the incident. The UBP will know which fog node to query and fetch the requested video to the user. One important question is that the fog nodes can return three different types of data according to the defined policies: original video, anonymised video and inferred data (e. g. number of people and vehicles captured on video at that given time). Within the use case, role-based access control is done, where the Law Enforcement Agents (LAEs) can access all types of data, city managers cannot access raw (unblurred/original) video, and city analysts can only access the inferred data for their urban planning activities.

Source

- FogProtect D2.2 - Validation Results of the 1st Iteration (Available, 2021-09-22).

Roles and Actors

- Actors/Parties
    - City entities, such as municipalities, police, firefighters, energy utilities
- Roles/Policies
    - Law enforcement agent: Entities such as first respondents (police and firefighters) where access to a clear video might save lives.
    - City managers: Managers of the city where they would only need access to anonymized data to understand what is going on and react accordingly. No need to access sensitive data.
    - City analysts: City employees that just want to obtain data that has been inferred from the video footage in order to run their analysis for urban planning.

Pre-conditions

- Street furniture with video camera and fog node capable of processing and storing video streams
- Urban platform running on a cloud centre capable of receiving requests from users and understanding which fog nodes to contact
- Mobile application that users can use to report occurrences
- Urban platform dashboard capable of communicating with the cloud centre to showcase the information based on roles and policies

Triggers

- Citizens reporting occurrences they witnessed around the city
- Fog node computer vision algorithm detecting incident
- Street furniture IoT device detecting vulnerable conditions (door opened without access, severe weather conditions etc.)

Normal Flow

- Areas of the city are monitored through video cameras, whose video streams are processed and stored locally for a given period of time;
- Citizens report occurrences that happen around the city;
- End-users of the urban platform receive notifications of the occurrences in the platform and, if necessary and given their level of access, request data from the relevant fog nodes;
- End-users analyse the video/data their policies and roles allow and act accordingly.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

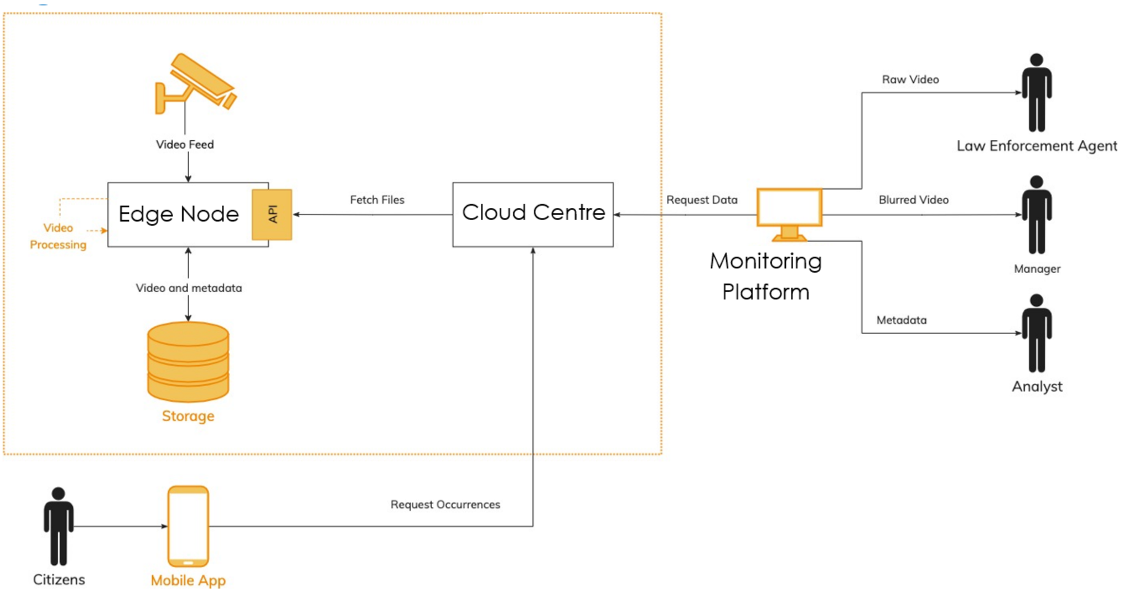


Figure 7: Schematic depiction of the protecting sensitive data within smart cities use case.

Figure 8: Schematic depiction of the smart cities use case deployment (left), video frame sample from the fog node after blurring sensitive data (right).

Potential Requirements

The use case requires a reliable and efficient high-bandwidth connectivity between the CCTV camera and the fog node (Ethernet, GPRS, WiFi). Such fog nodes must comprise CPU, GPU, memory, power management and high-speed interfaces, to process video streams and run computer vision algorithms that identify objects and people to blur and anonymise sensitive data.

Optical Network specific Requirements

The data exchange between video sources, fog nodes and cloud DCs need to support isochronous, low latency and deterministic communication. High-speed Passive Optical Network (PON) architectures allow an efficient support of this use case (Section 0).

2.6 Computing collaboration in optical access networks

Description

Access networks, usually PONs have unique characteristics: providing connection capability for a large number of network terminals, conveying comprehensive service profiles, and being close to the end users. Therefore, to make good use of computing power in the access network it is important to improve network performance, user experience, operation efficiency, etc. Traditionally, the computing power of an access network is distributed in the Optical Network Unit (ONU), Optical Line Terminal (OLT) and cloud, respectively, working independently, i.e., without much collaboration.

This use case explores the collaboration of computing power in an access network, showing the benefits on enabling high-quality service, enhancing network performance, etc.

The collaboration of computing power in PON access networks facilitates better quality for the network service. For example, correct identification of service requirement (e.g. service type, priority, latency, jitter, etc.) and transmission status (e.g. packet waiting time, channel quality, etc.) in the end device (e.g. residential gateway or access point) can be shared with the central office (e.g. OLT or cloud platform). This also helps to improve the dynamic configuration of the OLT (e.g. enabling slicing configuration, reserving buffer or bandwidth, etc.).

The collaboration of computing power needs to be adapted to different service scenarios to achieve the systematic integration and collaboration of the entire access network computing power, which brings multiple benefits:

- The service Quality of Experience (QoE) can be dynamically improved
- Network status can be reported in real-time based on the demands
- More application can be supported in the future

With the evolution of network services, end users pay more and more attention to QoE. Service providers need to improve service quality to maintain registered users and attract new users. To dynamically monitor and improve service quality, collaboration of computing power is a useful way. The followings show the basic functionalities of different network elements in computing power collaboration of PON access network, shown in Figure 9:

For the ONU, the network terminal for broadband access, is the closest location to sense in-premised network status. Therefore, the ONU is an appropriate network node to initially evaluate and analyse the status of the in-premises network based on its computing capabilities. Additionally, some simple operations such as WiFi automatic tuning can also be done. The ONU could also give feedback of the analysed results to the OLT, especially when the ONU cannot solve the problem by itself and needs help from the more powerful upper network due to its limited computing capability.

For the OLT, the computing specialized line card may have opportunities to analyse service data through mirroring and collaborate with ONU through the feedback to derive a deeper and more accurate view of the network status. Moreover, the OLT is able to perform certain operations such as service identification, DBA, QoS, etc. Obviously, the OLT may not solve all the problems. For example, it is difficult for the OLT to analyse user service quality in details and solve upper-layer network problems such as P2P Protocol over Ethernet (PPPoE) or to collect and analyse multiple user services data simultaneously for a long period. In many of these cases, interim analysis results and processed data will be uploaded to the cloud platform for further processing.

For the cloud platform, it is considered that computing power is here the most powerful, but it is far away from the user. The platform can proceed the deepest analysis of the service quality, analyse the service for a long time to get more accurate conclusions, and request other system/platform for collaboration to perform necessary operations. Moreover, the platform can monitor the computing power status of the OLT and ONU and adjust the computing tasks dynamically when the ONU's computing capability is not enough during processing task.

To further demonstrate the functionality of computing power collaboration, two cases are described as follows:

1. Example: Online interactive service quality assurance

The online interactive services such as online education, online conference, live streaming, have strict requirements on dynamic network quality in terms of throughput, latency, jitter, etc. In order to improve the user's QoE, the platform, OLT and ONU should collaborate to sense, analyse, process so as to satisfy QoE requirements.

In this case, the ONU should sense the in-premise network status including bandwidth usage, number of connected user end devices, WiFi quality (e. g. signal strength, interference), etc. in order to avoid any defects affecting the service quality of online services. For example, the P2P download application seriously affects the online interactive services. A traffic flow working in the same frequency band competes with data streams of the online interactive services. Weak WiFi signals in the user location creates communication paths of bad quality. Therefore, the ONU should sense, analyse and make the conclusion by leveraging its computing power. Then the ONU can proceed the operation like adjusting the WiFi configuration or speed limit if necessary and feedback a warning to the OLT for further processing.

After receiving the feedback from the ONU, the OLT can identify the service type and analyse the traffic status based on the undergoing traffic data (IP address, throughput, bandwidth usage, configuration, QoS, etc.) and identify the network problem. Then the OLT can proceed QoS strategy such as resource priority configuration and slicing to improve the quality of online interactive service. Furthermore, the OLT could also feedback the processed data and analysed result to cloud platform for further processing.

The cloud platform can analyse data (flow characteristics, time, latency, jitter, etc.) in details based on powerful computing power to get accurate results (user portrait, overall network status, whether the bandwidth of users matches the service, AP adopted by users, etc.) through southbound interface (MQTT/Telemetry), while coordinating with other platforms to ensure service quality, such as coordinating with BNG to improve uplink and downlink bandwidth, or synchronizing the priority applied in access networks to metropolitan area networks. Moreover, the platform can analyse user behaviour habits and main applications to promote more suitable services to users.

2. Example: Coordination function for FTTR across the network in different scenarios

NOTE: MFU/SFU are the terms defined in ITU-T SG15 Q3 recommendation G.9940. These terms have the equivalent meanings of P-ONU/E-ONU, defined in F5G documents.

FTTR provides a foundation for the good-quality application of WiFi. Moreover, a coordination mechanism over fibre and WiFi can provide a better collaboration among APs, which is defined in G.9940. This mechanism can avoid interference in air interface over multiple APs without any change of the WiFi protocol. In addition, as shown in Figure 10, the interference between different neighbours should also be avoided in the systematic point of view. The computing power in different network location should be capable to determine the coordination strategy, i. e. MFU as the network terminal gives strategy for a single FTTR network while OLT provides guides for coordination between different neighbours. Collaboration between the OLT and ONU controllers (MFU) is conducted to improved resource utilization in the frequency, time, and spatial domain.

The typical network coordination in a FTTR system (shown in Figure 10) can be described as follows:

- WiFi is provided by one FTTR system, so the MFU can process the coordination function over fibre and WiFi based on MFU's computing power in real-time.
- Parts of the Wi-Fi network in an area are provided by two FTTR systems located in the same OLT. This requires a cross-network coordination function within a single OLT based on the computing power of the OLT. The OLT can handle functions such as intra-BSS coordination.
- Parts of the WiFi network in an area are provided by two FTTR systems located in different OLTs. This requires a cross-network coordination function within a platform based on its computing power. The platform can process the network configuration including channel selection, resource allocation to dynamically manage the network. Therefore, in this complex WiFi network, the platform, OLT and MFU should cooperate with each other to perform a high-quality wireless network.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 3 (work in progress).

Roles and Actors

The actors in that use case are operators of networking and compute infrastructure and end-users in the residential or SME market segment.

Pre-conditions

Availability of compute resources at (or near) the network elements near the edge of the network.

Triggers

N/A.

Normal Flow

When a new service is provisioned, decision needs to be made on what compute infrastructure the software components of the service need to be installed. If the software components of the service need interaction with the network the location and speed of access to the right information is essential for an efficient operation.

For example 1: The service functions for optimizing the QoE are installed in an ONU or OLT, gathers network information about the service, and might decide to optimize the configuration for the optical or WiFi networking.

For example 2: The compute for the service is allocated at various places, which requires a level of coordination between the WiFi APs of different customers handing off the same OLT (or different OLTs). The software components read a lot of network related Information, analyse it and might change the configurations to improve the overall WiFi performance.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

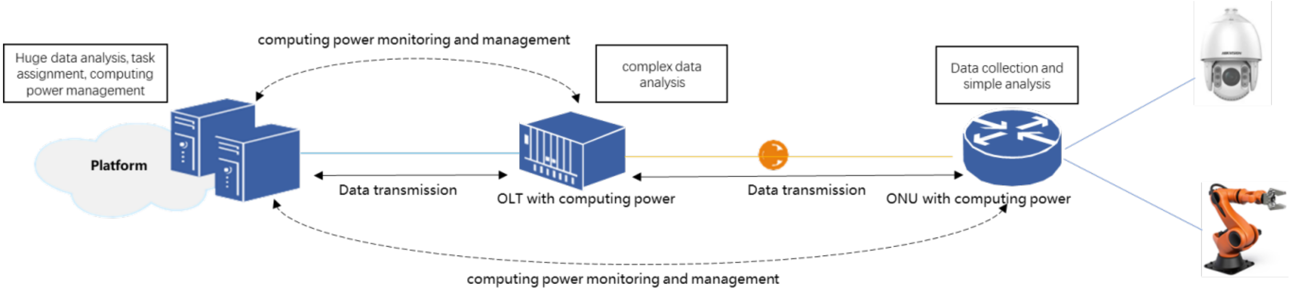


Figure 9: The basic network elements in computing power access networks.

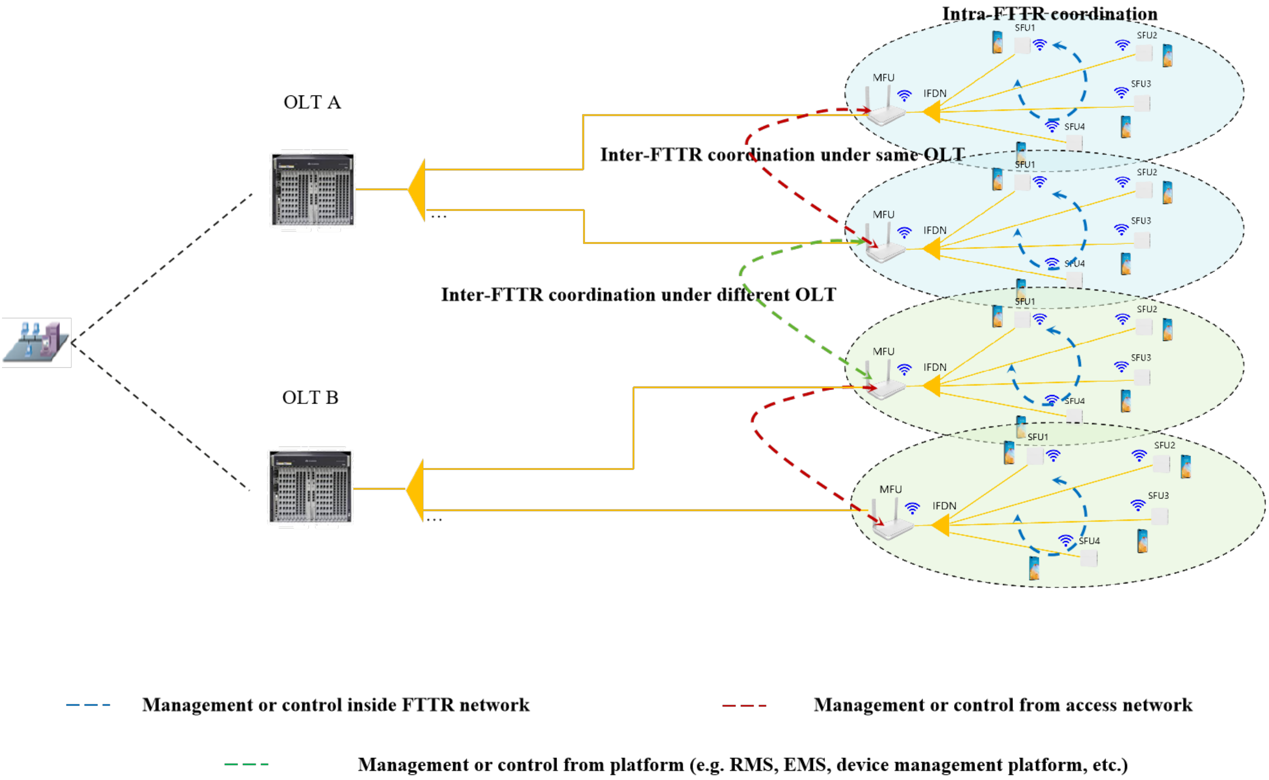


Figure 10: Coordination functions for FTTR across the network (main FTTR unit: MFU, subordinate FTTR unit: SFU, indoor fibre distributed network: IFDN).

Potential Requirements

N/A.

Optical Network specific Requirements

R1: The residential or SME network shall use optical backhaul of the WiFi APs such that control and management traffic for an intelligent optimization is not getting to large compared to the user traffic.

R2: The software components running on the optical network equipment shall have access to some local and eventually remote network status configuration and shall have the privilege to make changes to the configurations.

R3: Coordination mechanisms between the different compute locations are required for the large scale optimizations.

2.6 Robotics as a service

Description

Move to cognitive robots

In the past, the operation of robots in factories was characterized by a repetitive and static nature. These robots would tirelessly carry out the same task over and over again without any deviation or adaptability. However, with the emergence of Industry 4.0 and the subsequent demand for increased flexibility in manufacturing systems, robots had to undergo significant advancements to meet these new requirements. This transformative shift in robotics paved the way for the development of cognitive robots, marking a significant milestone in the field. Unlike their predecessors, cognitive robots are equipped with advanced sensors that enable them to perceive and comprehend their surrounding environment. This newfound perception allows them to adapt their behaviour dynamically based on real-time inputs, making them more versatile and capable of handling complex tasks (Figure 11).

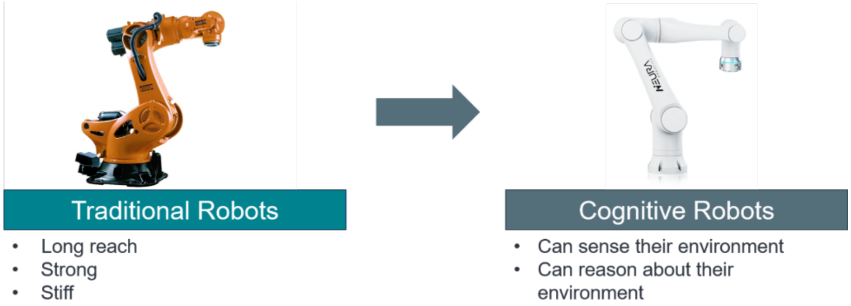


Figure 11: Move to cognitive robots.

Example applications

One of the most prevalent example applications of cognitive robots is bin picking (Figure 12), a process in which a robot is tasked with retrieving specific parts from a bin. Accomplishing this task involves a series of intricate steps. First and foremost, the robot must accurately detect the target part amidst a cluttered bin, a task often achieved through the utilization of sophisticated 3D cameras. Once the target part is identified, the robot must then devise a path or trajectory to reach the desired object and successfully grasp it. The execution of these tasks demands significantly more computing resources and the utilization of more complex algorithms compared to the conventional robots of the past.

Figure 12: Bin picking example application.

However, with this increased complexity and reliance on advanced technology, maintaining and managing such cognitive robotic systems becomes a more intricate endeavour. The reconfiguration of these systems for new tasks, as well as the troubleshooting and optimization of their performance, necessitates a higher level of expertise and technical knowledge.

Currently, most of the computational processes required by these systems are performed at the edge on an industrial PC in the control cabinet of the robot, necessitating the presence of a dedicated robotics and IT technician within the factory premises. Note that the location of the edge is deployment and application specific. The key characteristic is that the computational processes are located very near to the robot.

Move computing to the cloud

Nevertheless, the advent of cloud technology presents an opportunity to overcome these challenges and enable remote feedback control of cognitive robots. By leveraging the power of the cloud, the computational burden can be shifted away from the edge devices into centralized on-premise servers. This shift allows for more efficient resource utilization, as the cloud provides virtually limitless computing capabilities, accommodating the complex algorithms and demanding processing requirements of cognitive robots. Specifically, dedicated AI-optimized compute hardware can be leveraged.

The implementation of cloud-based compute for cognitive robots offers numerous advantages, including the ability to perform maintenance tasks remotely. Rather than relying on an on-site technician, the cloud infrastructure enables service technicians to diagnose, troubleshoot, and optimize the performance of cognitive robots from any location. This remote maintenance capability streamlines operations, minimizes downtime, and facilitates prompt resolution of issues, leading to increased productivity and cost savings.

While many of these potentials can be realized with standard Ethernet connections the real-time control of these systems and achieve a large multiplexing gain in computing is still a challenge. Ensuring the control signals from the cloud reach the robots promptly and in sync is pivotal for maintaining peak performance. Navigating these challenges demands innovative solutions, such as tapping into the capabilities of optical communication in F5G-A.

For illustration, consider the precision and ultra-low latency required in milling machine control systems, wherein control loops must respond within microsecond to ensure accuracy and quality in operations. Transferring such crucial real-time feedback control to a cloud system, via standard Ethernet technologies, could introduce larger and more importantly non-deterministic latencies, making accurate and immediate control unattainable.

In case of sensors or actuators mounted to robots, the fibre connected to those sensors like cameras needs to be robust enough for a large number of movements and large temperatures or temperature differences.

By harnessing the advanced capabilities of F5G-A networks, real-time control of cognitive robots from the cloud becomes a viable proposition. The high bandwidth and low- and deterministic-latency characteristics of F5G-A networks allow for the seamless and instantaneous transmission of control signals, ensuring that the cognitive robots respond swiftly and accurately to commands issued from the cloud. This opens up new possibilities for enhanced collaboration, increased efficiency, and improved overall performance of cognitive robotic systems in various industrial settings. In the end this could lead to new business models in which robot cognition is available as a service model straight from the cloud, or as we call it Robotics as a Service (RaaS).

Current robotics stacks, such as the Robotic Operating System (ROS), have already facilitated easy multiprocessing, enabling control nodes to be distributed across different machines. However, the introduction of a network into the communication structure inevitably leads to latency. This latency becomes particularly problematic for lower-level functions like motion control, hindering real-time responsiveness in the magnitude of 1 ms from sensor message to control command execution.

Emerging from this is a 'cloud barrier' in robotics, whereby lower-level manufacturing operations are executed at the edge, and high-level functionalities are relegated to the cloud.

A pivotal goal of our F5G-A use case is to strategically lower this cloud barrier by leveraging the outstanding capabilities of F5G-A enabling more efficient real-time control straight from the cloud, even with geographical flexibility around large manufacturing zones.

By utilizing F5G-A network technologies, companies can unlock the potential to offer software solutions for real-time process control as a service, effectively establishing a RaaS market.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 3 (work in progress).

Roles and Actors

There three distinct roles:

1) The robot vendor

2) The robotics control software vendor

3) The manufacturer

Pre-conditions

The robot's software is networked and can interact with control software in the cloud or on the optical network equipment. The robot and the robotics control software might be procured from different vendors. The robot's control software might be supplied as a service running in the either the robot's vendor cloud, the robot control software cloud, the manufacturer's cloud or the optical network's compute equipment.

Triggers

Normal Flow

Since robots are flexible, the use case starts with the definition of a task for one or several robots to do something. Then the control software figures out the particular robot configuration eventually with different tools attached for this task. The robot's control software for this task is created, loaded, initialized and connected over the fibre network with the robot.

Slight changes in the task can be made by changing the software component controlling the robot. For bigger changes the robot and its control software might need to be reconfigured.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

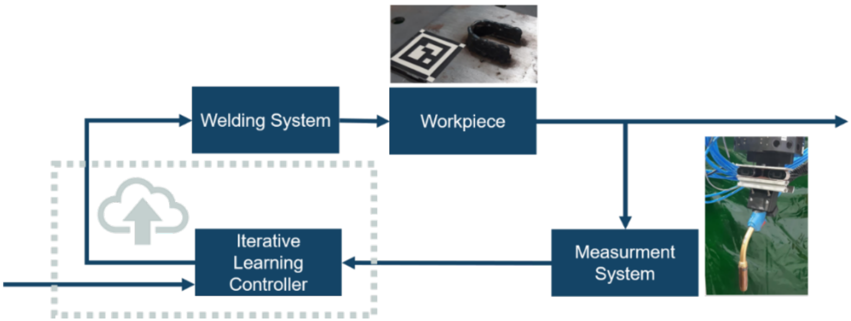


Figure 13: RaaS welding application.

While welding provides an illustrative example (Figure 13) – where AI-driven camera systems monitor and adjust the heat distribution – the envisioned RaaS model extends far beyond this. It empowers companies to separate the AI control component from the hardware, enabling seamless integration in different hardware and manufacturing scenarios.

Potential Requirements

To create a thriving market for RaaS and facilitate seamless remote control of robots from the cloud, several key factors need to be considered. These factors not only contribute to the success of RaaS but also foster a collaborative ecosystem where different components can easily integrate and work together. The following points outline some essential considerations:

The optical network shall enable high bandwidth and low-latency sensor data transfer to the cloud: The foundation of RaaS relies on the ability to transmit sensor data, including visual feeds, environmental information, and other relevant inputs, to the cloud with high bandwidth and minimal latency. This ensures that the data can be processed, analysed, and utilized in real-time, empowering cloud-based systems to make informed decisions and provide accurate control instructions to the robots.

The space needed for ICT in manufacturing can be removed from the shop floor and the space needed for ICT gets minimal.

Facilitate low-latency control of the robot from the cloud: To achieve real-time control, it is imperative to minimize the delay between cloud-based instructions and the robot's response. By leveraging low-latency communication channels offered by technologies like F5G-A, companies can significantly reduce the time lag in transmitting control signals. This enhances the synchronization and coordination between the cloud-based control systems and the robots, enabling rapid and precise execution of tasks.

Establish common interfaces for RaaS: A critical aspect of driving the adoption and integration of RaaS solutions is the development of standardized interfaces. These interfaces serve as a bridge between different components of the robotic ecosystem, enabling seamless interoperability and easy integration of diverse hardware, software, and AI-driven modules. Common interfaces simplify the process of connecting various components, reducing complexity, and fostering collaboration among different stakeholders in the RaaS market.

All the processing needed in such a use case can run in the cloud, either provided by the manufacturing stakeholder or by specialized software for the robotics stakeholder.

Enable high-bandwidth, low-latency transfer of sensor data to the cloud: By leveraging the capabilities of F5G-A networks, companies can ensure that sensor data, such as uncompressed camera feeds or other relevant information, can be swiftly transmitted to the cloud with minimal delay. This enables real-time analysis, processing, and decision-making based on the acquired data.

Enable low-latency control of the robot from the cloud: F5G-A networks offer the potential for rapid and synchronized communication between the cloud and the robotic systems. This low-latency control allows for real-time instructions to be sent from the cloud to the robots, enabling immediate and precise responses. This advancement paves the way for enhanced coordination, adaptability, and overall performance of robotic systems in various applications.

By addressing these key areas, the foundation for a vibrant RaaS eco-system can be laid. High bandwidth, low-latency sensor data transfer ensures real-time data analysis and decision-making capabilities in the cloud. Low-latency control facilitates instant responsiveness of robots to instructions issued from the cloud. Lastly, common interfaces promote the compatibility and smooth integration of different components, streamlining the adoption and expansion of RaaS offerings.

Optical Network specific Requirements

To seamlessly integrate the ROS with the F5G-A, it is crucial to consider the communication protocols employed by ROS. ROS uses the Data-Distribution Service (DDS), a publish-subscribe protocol that can operate over different transports such as TCP and UDP.

To enable ROS messages to be transmitted over the F5G-A network, a seamless integration between F5G-A and the DDS protocol is essential. This integration must prioritize low latency, high bandwidth, and deterministic latency. Deterministic latency ensures that the latency remains bounded and consistent, minimizing variations in communication delay.

Achieving this integration requires designing an F5G-A network infrastructure that optimizes real-time data exchange. By minimizing processing delays, optimizing data transmission protocols, and efficiently utilizing network resources, low-latency and high-bandwidth communication channels can be established.

However, it is still beneficial to keep computing resources separate from the networking equipment as many advanced robotic applications will require special accelerators such as TPUs.

The seamless integration of DDS with the F5G-A network enables efficient transmission of ROS messages, ensuring responsive and reliable communication between the cloud and the robots. This integration lays the foundation for RaaS and facilitates advanced robotics applications in various domains.

In the case of actuators and sensors mounted on the moveable part of the robot, the fibre cable and connectors shall be durable for a lot of movements, torsions, and stretches.

3. Computing continuum requirements and KPIs for optical communications

This section gives an overview about the general and optical networking specific requirements for the Computing Continuum.

3.1 Computing continuum requirements

Requirements derived from use cases

High performance requirements: Network bandwidth and storage requirements are high, especially for use cases where high-resolution image and video processing is dominant.

Flexible bandwidth allocation: Network service providers and operators may need to support flexible bandwidth allocation to match the needs of the different use cases. Temporary increases in bandwidth have to be supported, e.g. when downloading necessary data from the cloud, to provide a good quality of user experience.

High reliability: This includes very low latency of the networking components from the cloud to the terminal, such that, e.g., video, images or processed data are available reliably on time. And since some of the use cases are mission critical, e.g. downloading a medical image in emergency situations, the overall system including the edge compute and the network needs to be highly reliable.

Data security and privacy: Communications between terminal, edge and cloud have to be protectable at different levels of security. Personal data has to be protectable from any unauthorized third-party access or malicious attacks and exploitation of data.

Direct optical network support in the computing continuum platform: Several use cases require low delay between the sensor location and the compute location. In order to achieve that optical cut-through technology and direct optical access to the compute resources are required.

Scalability of resources: The computing continuum must support scalable computing and storage resources to efficiently handle fluctuations in demand across various use cases. This includes the ability to dynamically provision and de-provision resources in real-time to accommodate the needs of applications ranging from IoT devices to high-powered computing tasks, ensuring optimal performance and resource utilization without manual intervention.

The following optical networking specific requirements were derived via the previously described use cases:

Use case: Cloud-based medical imaging (Section 0)

Table 5: Network bandwidths of hospitals with different scales.

Hospital Size	Daily patients/visits	Image and image reading terminal in the hospital	Network bandwidth for image storage to the cloud in Mbit/s
Large hospital	20,000	2,000	15,840
Medium-sized hospital	7,000	800	6,336
Small hospital	1,000	100	792

Note this number assumes only the imaging part. In case of (remote) surgery scenarios also video is required and needs to be calculated on top. Also for regulatory reasons some videos need to be stored for later use as prove or teaching material. The surgery and video-oriented use cases are different compared to this one, and are for further study or can be handled in a different use case.

High-QoE for users

Good Quality of Experience (QoE) for doctors and staff (instantly visible medical images), browsing through large sets of images (delay sensitive).

The use of computing continuum technologies enables and improves such requirements.

Use case: Cloud-based visual inspection in production (Section 0)

Table 6: Target KPIs for cloud-based visual inspection for automatic quality assessment in production.

Target KPI	Value
Upstream data rate per vision inspection station	1 Gbit/s (single GigE Vision camera) – 20 Gbit/s (4× USB3 Vision cameras)
Downstream data rate per vision inspection station	> 400 kbit/s (control signals only)
E2E cycle time*	5 - 10 ms typical < 2 ms time-critical scenarios
Reach (max. distance to edge DC)	< 80 km

*cycle time is determined by the time required for the vPLC to send all control signals to its assigned targets and to receive all of their feedback in return

Use case: Cloud-based control of automated guided vehicles (Section 0)

Table 7: Target KPIs for cloud-based control of automated guided vehicles.

Target KPI	Value
Upstream data rate from AGV to edge	> 400 kbit/s per AGV > 10 Mbit/s per AGV in case of video upstream
Downstream data rate from edge to AGV	> 400 kbit/s per AGV
E2E roundtrip latency	< 30 ms*
Reach (max. distance to edge DC)	< 80 km

*including processing time at edge DC

Use case: Cloud-based control of production via optical wireless communication (Section 0)

Table 8: Target KPIs for cloud-based control of industrial production via OWC.

Target KPI	Value
OWC cell (coverage area)	4 m x 5.5 m x 5 m (height x width x length)
Minimum achievable speed inside an OWC cell	100 Mbit/s
Minimum achievable speed in backhaul	1 Gbit/s
E2E roundtrip latency	< 10 ms*

*including processing time at edge DC

Use case: Protecting sensitive data within smart cities (Section 0)

- The data exchange between video sources, fog nodes and cloud DCs need to support isochronous, low latency and deterministic communication. High-speed Passive Optical Network (PON) architectures allow an efficient support of this use case.

Use case: Robotics as a Service (Section 0)

- The residential or SME network shall use optical backhaul of the WiFi APs such that control and management traffic for an intelligent optimization is not getting to large compared to the user traffic.
- The software components running on the optical network equipment shall have access to some local and eventually remote network status configuration and shall have the privilege to make changes to the configurations.
- Coordination mechanisms between the different compute locations are required for the large scale optimizations.

Use case: Robotics as a Service (Section 0)

- To seamlessly integrate the ROS with the F5G-A, it is crucial to consider the communication protocols employed by ROS. ROS uses the Data-Distribution Service (DDS), a publish-subscribe protocol that can operate over different transports such as TCP and UDP.
- To enable ROS messages to be transmitted over the F5G-A network, a seamless integration between F5G-A and the DDS protocol is essential. This integration must prioritize low latency, high bandwidth, and deterministic latency. Deterministic latency ensures that the latency remains bounded and consistent, minimizing variations in communication delay.
- Achieving this integration requires designing an F5G-A network infrastructure that optimizes real-time data exchange. By minimizing processing delays, optimizing data transmission protocols, and efficiently utilizing network resources, low-latency and high-bandwidth communication channels can be established.
- However, it is still beneficial to keep computing resources separate from the networking equipment as many advanced robotic applications will require special accelerators such as TPUs.

- The seamless integration of DDS with the F5G-A network enables efficient transmission of ROS messages, ensuring responsive and reliable communication between the cloud and the robots. This integration lays the foundation for RaaS and facilitates advanced robotics applications in various domains.

- In the case of actuators and sensors mounted on the moveable part of the robot, the fibre cable and connectors shall be durable for a lot of movements, torsions, and stretches.

Other requirements

In Figure 14, see [ZaAh19], a list of requirements for enabling edge computing and computing continuum is provided, which are more detailed:

- **Real-time applications support**: Edge computing provides numerous services and particularly supports real-time based applications.

- **Joint business model for management and deployment**: Is needed due to the fact that the edge computing systems are typically owned by different service providers and work under different business models.

- **Resource management**: A dynamic resource management approach is needed in order to adapt the various service demands to resources, which need to be allocated and distributed in different processing/computing points, i. e., cloud DC, edge computing systems and end devices.

- **Scalable architecture**: The number of IoT devices in an edge network has significantly increased and with it the demand of edge-based services and resources. Therefore, a scalable edge computing architecture is considered as vital as it can lower the cost.

- **Redundancy and fail-over capabilities**: These requirements are needed for the reliable functioning of edge computing systems in order to support many critical business applications with strict performance requirements, such as low latency and uninterrupted content delivery services. Moreover, in order to develop reliable and resilient edge computing systems, redundancy and fail-over capabilities should be as well considered.

- **Security**: Due to the heterogeneous nature of edge computing systems, security is very important.

- **Optical networks**: Network service provider or operator need to support flexible resource allocation to match the bandwidth, latency and resilience needs.

The following requirements can be considered as open challenges:

- **Users trust on edge computing and on computing continuum systems**: The success of edge computing and computing continuum is related and linked to trust that is regarded as one of the most important factors for the acceptance and adoption of these edge computing and computing continuum systems by consumers and users.

- **Dynamic and agile pricing models**: The rapid increase of the edge computing applications and services creates the need for dynamic pricing and market places.

- **Service discovery, service delivery and mobility**: Distributed and federated edge computing systems require service discovery and delivery support, in particular, for scenarios where multiple mobile devices are used that require services simultaneously and uninterruptedly.

- **Collaborations between heterogeneous edge computing systems**: The ecosystem of edge computing systems consists of a collection of different processing/computing points, e. g., cloud DC, edge computing systems and end devices, and different underlying communication infrastructures, which makes the collaboration between such systems a challenging task.

- **Low-cost fault tolerant deployment models**: Deployment models that are fault tolerant are important because they ensure the continuous operation of any system in the event of failure with little or no human involvement.

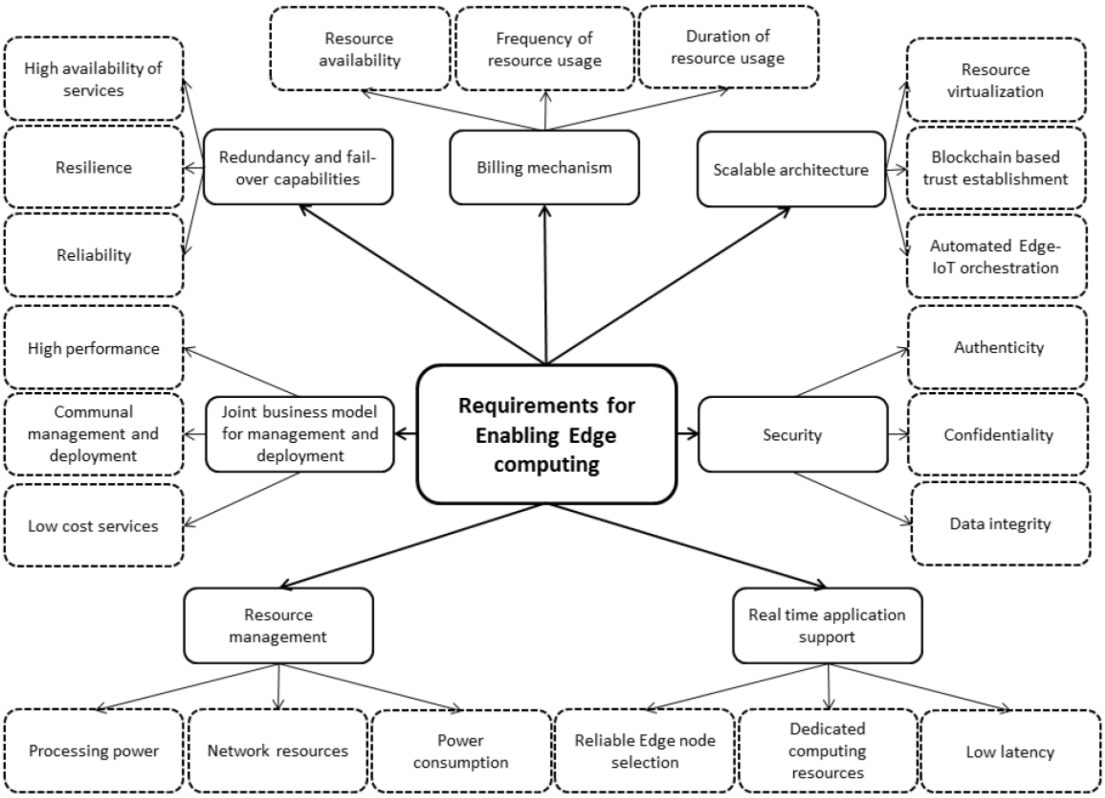


Figure 14: Requirements for enabling edge computing and computing continuum [ZaAh19].

References:

[ZaAh19] Wazir Zada Khan, Ejaz Ahmed, Saqib Hakak, Ibrar Yaqoob, Arif Ahmed, "Edge Computing: A Survey", Elsevier, Future Generation Computer Systems, Volume 97, August 2019, Pages 219-235.

3.2 KPIs for optical communications

This section applies the requirements derived in Section 2. on deriving the KPIs for the network connecting edge computing platforms and cloud, considering that an optical communication infrastructure is used as underlying network.

For the access network, 10G PON has become the dominant broadband access technology and has been continuously optimized. Currently, the 50G-PON generation of access networks are getting available on the market and around 200 Gbit/s are in the research phase. It achieves full coverage of gigabit access to the customer premises. Coexistence with Gigabit PON (GPON) enables smooth network migration.

High-bandwidth technologies, such as 100GE and Optical Transport Network (OTN), are deployed at access sites to implement large-bandwidth backhaul for access networks and ensure E2E gigabit-per-second bandwidth capabilities.

Wavelength Division Multiplexing (WDM) nodes are moved from the backbone network down to the access network central offices and are directly interconnected with OLTs to implement E2E all-optical connections.

The capacity of OTN is continuously improved. 200 Gbit/s and 400 Gbit/s single-wavelength OTN are fully deployed, 800 Gbit/s is ready and the C-band and L-band are widely used, achieving high-performance transmission of more than 40 Tbit/s per fibre. OTN is responsible for DC interconnection and even provides high-speed connections between servers inside the DC.

PON shall support distinct types of services based on different latency, jitter and bandwidth requirements.

Table 9: KPI targets for various dimension in the fixed broadband, see the F5G Advanced generation definition [F5GA23]

Fixed Network Generation	F4G	F5G (revised)	F5G Advanced
Generation reference	UltraFast BB (UFBB)	Gigabit BB (GBB)	MultiGigabit BB (mGBB)
	Mbits	Gigabit	10 Gigabit
Reference Downstream Bandwidth per User	100-1000 Mbps	1-5 Gbps	5-25 Gbps
Reference Upstream Bandwidth per User	50-500 Mbps	1-5 Gbps	5-25 Gbps
Reference services	UHD 4K Video	VR Video Cloud Gaming Smart City	Extended reality Metaverse Digital twins Industrial optical network
Reference Architecture	FTTH/FTTdp	FTTH/FTTR	FTTR/FTTM/FTTT
Access Network Technology Reference	GPON/G.Fast	10GPON	50GPON
Technical Specification reference	G.984.x G.9701	G.987.x (XG-PON) G.9807.x (XGS-PON)	G.9804.x
On-Premise Network Technical Specification reference	FE/GE+Wi-Fi4/Wi-Fi5	GE/10G 2.5 Gbps FTTR (G.FIN) WiFi6 (802.11.ax)	10 Gbps FTTR (G.FIN) WiFi7 (802.11be)
Radio Frequency (RF) Video over Fibre (LAN Coaxial) reference	Yes	Yes	Yes
Aggregation and core network	IP/MPLS WDM	IP/Eth OTN/ROADM	IP/Eth OTN/fgOTN/fgMTN/OXC
Reference Bandwidth per wavelenght	100 Gbps	200/400 Gbps	400/800 Gbps
Autonomous network level	-	3	4

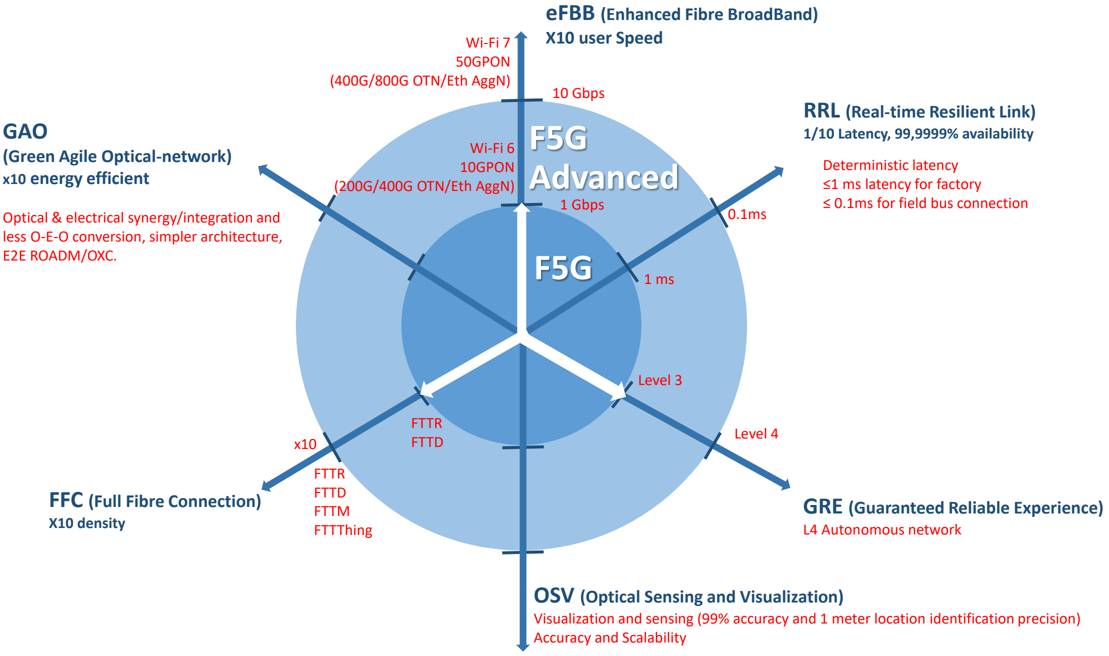


Figure 15: F5G Advanced generation dimensions and KPIs

References:

[F5GA23] F5G Advanced Generation Definition ETSI GR F5G 021 V1.1.1 (2023 11) Fifth Generation Fixed Network (F5G), F5G Advanced Generation Definition

4. Enabling technologies

This section provides the specification of the optical communication infrastructure as the enabling technology to support the KPIs described in Section 0. In particular, features as the following ones have to be considered:

✓ Slicing for IoT with soft or hard isolation

The concept of slicing has been described in 5G for mobile networks and F5G for fixed networks. Slicing is basically a virtual network service and the level of guarantees and service quality can be configured to the various need. The basic concept is that traffic from different tenants, like IoT applications, can choose the service quality expected. Both hard as well as soft isolation between different slices can be configured. The slicing concept is an E2E concept and therefore well suited for IoT applications with various needs.

Also, in the context of computing continuum, slicing is a concept which can be applied and enable more freedom to place the IoT workload at the best place in order to guarantee and achieve the application's required quality.

✓ Hard pipe between IoT devices and cloud

Enterprise access and PON-based OLTs backhaul need to communicate to multiple clouds, therefore the transport network requires to provide P2P, P2MP and multipoint-to-multipoint interconnection capabilities.

On traditional transport networks, enterprise IT leases multiple P2P private lines (L2 E-Line/MPLS PWs) from the carriers to implement single-point to multiple clouds and multi-point to multiple clouds access. However, those technologies provide a certain degree of resource reservation and separation of traffic, however, for demanding services and some of the use case described above that is not enough. Hard pipes are dedicated resources and guarantee latency and reliability. For example, fine-grain OTN (fgOTN) provides constant latency, guaranteed reliability, and reserved resources for the total path. Also, such technologies maintain timing transparency. The fgOTN technology is enhanced with the capability to have rather small E2E connection fully guaranteed and hard isolated (starting at 10 Mbit/s). But also, the technology has been extended for high end connections beyond 800 Gbit/s.

These hard pipes can be the foundation of hard isolated slices.

✓ Soft pipes between OLT and cloud based on IP/Ethernet

As described for the case of hard pipe, the soft pipes are more using packet-based technologies like IP/Ethernet. It is basically private link with certain network characteristics. Since the resources are shared, scheduling mechanisms are needed to guarantee a particular bandwidth to a client. However, there is more packet delay due to store and forward of packet networks and packet jitter introduced. Depending on the use case such technologies are suitable due to the capability of resource sharing and multiplexing gain, which has its commercial benefits.

Depending on the IoT application and its requirements this is acceptable or other means are needed.

These soft pipes can be the foundation of soft isolated slices. Network slicing and service identification and mapping are effective means to ensure Internet access quality. Network slicing is not a new technology. However, most network slices are soft slices, which are mainly reflected on the management plane.

Actual resources can still be shared among different slices, and hardware resource reservation for high-priority services is not supported. Hardware slicing reserves dedicated hardware resources (such as buffers, CPU computing capabilities, air interface resources, and PON timeslots) for high-priority services that are not shared with low-priority services, to implement hard isolation between different priorities.

Hardware slicing of the Customer Premises Equipment (CPE), and PON shall be supported. E2E slicing shall be supported to isolate private line service from other prioritized users such as home broadband users and other SMEs for quality assurance. E2E slicing shall be supported to isolate different applications of a private line service for application quality assurance.

✓ TSN over PON

Time-Sensitive Networking (TSN) is the IEEE 802.1Q defined standard technology to provide deterministic messaging on standard Ethernet. TSN technology is centrally managed and delivers guarantees of delivery and minimized jitter using time scheduling for those real-time applications that require determinism.

Incorporating TSN features in the access and transport network is expected to unleash the potential of E2E deterministic communications, especially in industrial environments and time-critical applications like factory automation.

✓ 50G-PON (seen by many as the next step after 10G-PON)

Higher speed PONs, such as 50G-PON, allow the support of broadband services with higher data-rates as well as lower latency. Sharing requirements with existing systems by supporting the same loss budgets and distances will allow for cost efficient deployments. Usage of Digital Signal Processing (DSP) and enhanced Forward Error Correction (FEC) will provide the required improvement.

✓ AI based application perception and mapping to proper pipes

Different broadband applications are required to be recognized by the network in order to guarantee the application experience.

Application identification could be implemented based on an artificial intelligence mechanism. The legacy method for application identification is based on packet analysis, such as Deep Packet Inspection (DPI). To protect the privacy of broadband users, it is recommended to use AI to analyse traffic at application level (instead of using packet analysis such as DPI).

Depending on the required application performance the application traffic is then mapped to the right tunnels with the appropriate quality assurance.

✓ Management and control

The Management and Control (M&C) of the optical infrastructure plays a critical role in the realization of the computing continuum, primarily through its role in setting up and tearing down the E2E services interconnecting computing resources that are geographically placed apart (e. g. enterprise edge belonging to a manufacturer with multiple sites). In addition to the communication services, the M&C stack can play a more direct role in rolling out edge cloud services. This can be realized by having the M&C stack controlling the underling optical networks, but been orchestrated by a centralized orchestrator, which assigns not only virtualized network function (VNF) (e. g. OpenStack-managed VMs or Kubernetes-managed containers), but also the E2E communication links that connect the VNFs in a chain across the whole infrastructure. In such scenario, M&C stack will control the optical network elements through open or proprietary South Bound Interfaces (SBI), while they communicate with an orchestrator (e. g. ETSI OSM) in the North Bound Interface (NBI).

Moreover, to support the computing continuum requirements, the current M&C functions should consider both centralized solution as well as distributed M&C, where there are multiple of M&C agents running next to the edge to significantly reduce latency that could be imposed by the M&C itself. Furthermore, it should benefit from AI/ML functionalities to make smarter and more proactive decisions.

Optical cloud networks are optical networks adapted to the cloud paradigm with on-demand service provisioning, quick and dynamic bandwidth adjustment, and high reliability for also mission critical workloads.

- ✓ Cascaded PON for FTTR, FTTO, and FTTM

Cascaded PON directly extends optical fibres to each room (Fibre To The Room, FTTR), office (Fibre to the Office, FTTO), machine (Fibre to the Machine, FTTM) achieving gigabit coverage everywhere at home, offices, or enterprise/vertical industries network. As shown in **Figure 16**, cascaded PON is comprised of two stages of PON interfaced by a Customer Premise Network-Aggregator (CPN-A), which acts as a light Optical Line Terminal (OLT) unit. In comparison with previous PON generations, the introduction of cascaded PON (FTTR, FTTO, or FTTM as coined in ETSI ISG F5G specifications) will be a major improvement in fibre connection numbers. This will fundamentally change the network topology, flow model as well as the management. In addition to ETSI ISG F5G, cascaded PON is under further developments in the G.fin-SA project in ITU-T Q18/15.

Cascaded PON delivers higher data rates to each individual rooms or office spaces where WiFi-capable ONUs can offer a remarkably better performance to the end user compared to the previous generations (e. g. FTTH) where a single ONU ends at the end points (e. g. homes).

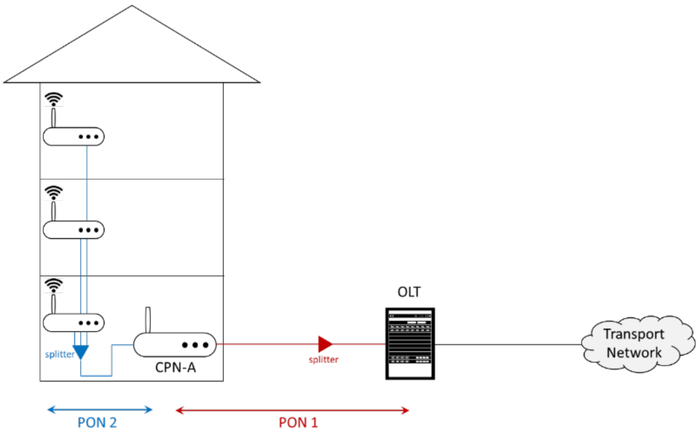


Figure 16: Cascading PON architecture.

- ✓ Energy efficiency

That means to migrate from today's architectures to more and more all-optical architectures and to remove any optical to electrical conversion along a path. Today's technology can run all optical to central offices enabling high-end IoT networking (s. a. Section 0).

- ✓ Computing collaboration

With the availability of computing capabilities on various network nodes and at different places in the network, collaboration between the different compute locations is need to achieve an optimal computing continuum. See below for different edge computing scenarios.

5. Edge computing support on-premise and on-device

This section describes approaches to support edge computing on-premise and/or on-device.

An example of E2E optical network topology for connecting end users to the cloud is shown in Figure 17. In this network, either an ONU or Customer Premise Network-Aggregator (CPN-A) with ONU functions embedded is connected to an OLT via a P2MP fibre based PON. The CPN-A connects to multiple Edge Customer Premises Equipment (E-CPE) with each E-CPE that may connect to one or more end user devices. OLT uplinks to the Aggregation Node (AggN) Edge are possible by either an IP/Ethernet network and/or an Optical Transport Network (OTN). The AggN Edge connects to the core network via core Provider Edge (PE) or connects to the DC via a DC gateway (GW).

Edge computing functionality may be located in: ❶CPN gateway, ❷access node (OLT), ❸aggregation edge, or in the cloud ❹. In general, edge computing functionality can be part of the CPN gateway, OLT or Aggregation Edge (the combination of ❶, ❷, and ❸). This brings various requirements to the link between the device and the edge, and the edge computing to the cloud according to the location of edge computing functionality.

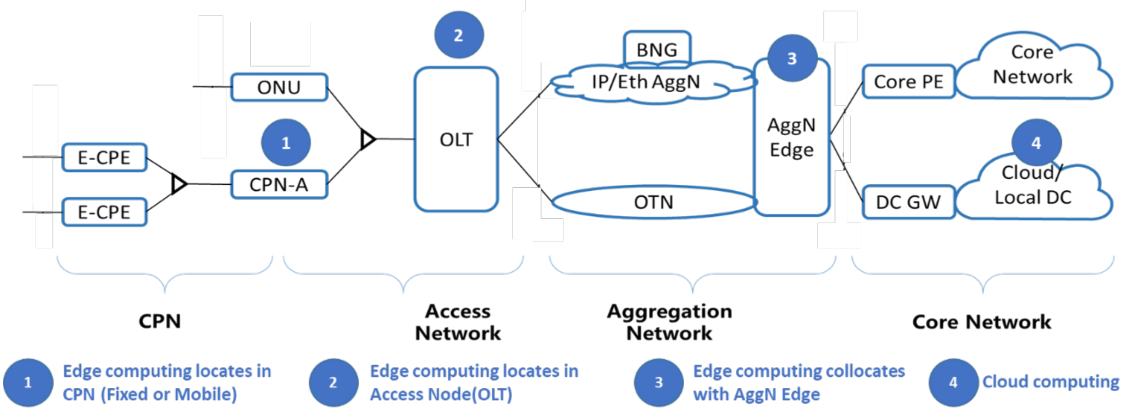


Figure 17: Network topology for edge and cloud computing.

❶ Edge computing functions locates in the CPN-A. In this case edge computing functions are close to the end user, this enables real-time services (e.g. massive IoT link aggregation, industrial IoT data protocol conversion, industrial machine vision data processing, industrial protocol conversion, AR assistant) be pre-processed at the edge cloud, and loosens requirements to capacity, latency and jitter, and gives higher immunity to link outage of the link between edge computing and cloud. The cost brought by this is the demand of dedicated resources at the edge that leads to higher cost (CAPEX and OPEX) and space requirements for installation.

❷ Edge computing functions locates in or aside of the access node (i.e. OLT). In this case edge computing functions are not as close to the end user (s. case ❶). This increases the requirements for capacity, latency, jitter and reliability to the access network. For example, some applications require a jitter free link that is challenging to Time Division Multiple Access (TDMA) based PON systems. As OLT connects to a large number of CPN users, edge computing resources can be shared among more users and this helps for reducing the cost, and loosening requirements to the environment. Moreover, installing stronger computing power and more storage capacity can handle more tasks simultaneously in comparison to case ❶. In this case, the requirements for the link between edge and central cloud will be even looser than for case ❶, as more power full edge computing can handle more pre-processing works. Case ❷ raises stricter requirement to the access network between CPN and OLT, which sometimes exceeds the threshold of PON technology.

❸ This case is similar to case 2 but the edge computing functions locate in an even higher position, which enables resources such as computing power and storage capacity be shared among more users and further lowering down the cost of edge computing functions. As it is closer to the cloud, the requirements to the cloud are lower in comparison to the above discussed cases. As a cost it brings even stricter requirements to the network, in terms of capacity, latency and jitter. Technologies are demanded to guarantee capacity, latency, jitter and reliability of the link: for example, E2E hard slicing, hard pipe link such as OTN between OLT and aggregation network, jitter free PON technologies, etc.

One more possible case is that edge computing functions are divided to several parts and located in combination of CPN, access node and AggNEdge. The split of real-time functions - such as industry protocol conversion - located in the CPN and non-real-time functions - such as IoT link aggregation - located in the access node allows to compromise between link capacity, latency and jitter, etc.

It's hard to simply judge which case illustrated above is better than the others. It may be subject to services and applications for each real deployment.

6. Edge computing platforms with optical cut-through support

This section describes approaches to support optical layer shortcut for extremely low latency.

The factors adding latency to an E2E communication are the distance, the traffic handling in the end-systems, and the traffic handling in network nodes like an edge compute node.

In the network nodes the delay can be attributed to the many O-E-O conversions and the store and forward behaviour of packet-based networks such as IP and Ethernet. In the case that per-node delay can be avoided, distance remains the primary factor. That also means that easy optical cut-through of traffic through edge nodes supports minimal latency. This implies that the computation location can be chosen more flexibly. Moreover, multiplexing gains can be achieved more easily through larger compute nodes and sharing of compute infrastructure; operational cost can be better shared. Still there are applications that need even lower delays and therefore the edge node computation is still relevant.

Figure 18 shows two cases: green a traffic receiving some sort of edge computing, where the red traffic is cut-through directly to the cloud with smallest possible delay and lowest energy usage.

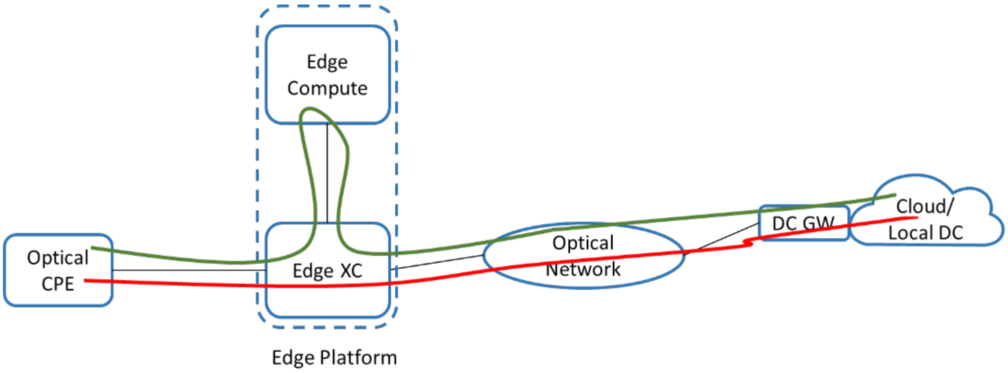


Figure 18: Illustration of the optical cut-through approach (read line).

Fibre to the Machine (FTTM)

The use of fibre to everywhere in the scenarios for machines and robots need to decide where the compute of the data is best suited, depending on the data containment and the level of AI/compute capability needed. With all-optical networking in FTTM, the choice can be made very free, and high bandwidth IoT devices (e. g. 5 Gbit/s industrial video streams) are possible to handle the traffic to the place where compute is available for a particular application.

Fibre to the Office (FTTO)

In the case of FTTO, two aspects with regards to optical communication and cut through are important:

1) Direct connections to the cloud:

With new optical communication technologies available, application of any sort can directly communicate to the cloud computing resources, no matter where they are located, in the fastest possible way. From a packet network perspective, it allows for only one hop to the cloud.

2) Reduced space, energy requirements for campus deployments:

Using optical communication technology on the campus, reduces the number of equipment rooms, the space needed in ducts, and the overall energy need for the communication of high-quality communication. For the computing continuum, it proves free choice of placing compute and the overall system can be optimized along different dimensions in cloud data governance, compute availability, ease of operation, etc.

7. Orchestration of the computing continuum

This section describes approaches to support the orchestration features applied in computing continuum.

The basic technologies needed are the computing management with algorithms for workload placing according to the required QoS parameters. The network - from the end-system to the computing instances - needs to be configured to forward application traffic to the compute nodes and needs to conform to the QoS requirements. Due to all-optical communication, the orchestration is very flexible in placing work-loads, also dimensions outside of QoS dimensions, such as data governance, green energy usage, security, and ease of operation.

For any changes in traffic load and compute load, appropriate scaling actions need to be taken to keep the agreed service level consistent. The optical communication is agile enough to follow the required placement.

For resiliency purposes, the right backup resources need to be available and fast switch-over technologies are required. The carrier grade technology of optical communication has those mechanisms already built-in and can be easily reused for computing continuum applications as well.

To further enhance the efficiency and responsiveness of orchestration within the computing continuum, advanced predictive analytics and machine learning techniques can be integrated. These technologies can forecast potential changes in computing and network load based on historical data and real-time input, enabling proactive adjustments before performance degradation occurs. By utilizing predictive models, the system can pre-emptively scale resources, reroute traffic, and initiate failovers, ensuring continuous service delivery and maintaining system integrity under varying conditions. This approach not only optimizes resource use but also significantly reduces manual oversight, paving the way for more autonomous and resilient computing environments.

8. Security for the computing continuum

This section describes approaches to support security in the computing continuum scenarios.

Security for third-party code on edge computing platforms

Any edge computing platform running third party programme code has security implications. Virtualization is a tool to separate different compute instances from each other. ETSI ISG NFV has specified the base line Virtualization platform and management and orchestration for Network Function Virtualization (NFV). The security aspects are specified in several specifications dealing with virtual network function (VNF) package security (ETSI GS NFV-SEC 021 V2.6.1), security on the management interfaces (ETSI GS NFV-SEC 022 V2.8.1) and security aspects of the different visualization technologies including virtual machines and containers (ETSI GS NFV-EVE 004).

Data protection for edge computing platforms

Through the virtualization technologies and the capabilities of virtualizing memory and storage, as described above, a certain level of data protection can be achieved. The higher challenge is the trade-off between data protection and legal interception capabilities. This depends on the deployment scenario and the regulatory environment. ETSI ISG NFV has described a Legal Interception architecture for NFV (ETSI GR NFV-SEC 011 V1.1.1).

9. PoC report: Edge/Cloud-based visual inspection in production

Overview

The objective of this Proof of Concept (PoC) demonstration is to showcase the use case edge/cloud-based visual inspection in production, in which an AI-based visual inspection model runs on an edge/cloud sorts out 3D printed objects in different classes. The broadband connectivity between the edge/cloud and the Visual Inspection Station (VIS) is provided by a PON. Specifically, the VIS is connected to the edge/cloud through three ONUs. Each ONU supports one camera or a robot arm in the VIS. The demo forms an E2E control loop (camera (observe) → edge/cloud (analyse) → robot arm (act)). The E2E observe-analyse-act (OAA) offers an E2E video processing pipeline with remote compute capability.

Additionally, all the devices are powered by a smart Power Distribution Unit (PDU), which provides real-time energy consumption monitoring that can be used for carbon footprint analysis. The power consumption data together with several networking parameters (e. g. data rate, throughput) are streamed live to a data lake for further pro-cessing or visualization. The telemetry pipeline is based on the architecture presented in [BESH23].

Topics of investigation

Figure 19 shows an overview of the entire setup. The setup involves a VIS comprising two 5GigE cameras (Basler a2A2840-67g5BAS), one robot arm (COBOTTA IP30), and a conveyer belt. A 3D printer (Ultimaker S3) was also used to print the 3D objects. Figure 20 and Figure 21 show the Basler camera and the COBOTTA, respectively. The VIS is provided broadband connectivity through an XGS-PON testbed with three ONUs. The Basler cameras are connected to two OptiXstar P812E ONUs, as they offer 2.5 Gbit/s interfaces. The robot arm is connected to an S892E ONU. We have set up dedicated network slices for each camera and the robot arm to connect them in an isolated slice to the cloud. The network slice for the cameras is set with assured bandwidth (BW) of 2.5 Gbit/s and maximum BW of 5.0 Gbit/s, while the network slice of the robot arm is set with max BW of 100 Mbit/s. As the network slicing feature does not span out of the PON network, three distinct virtual LANs (VLAN) were set up from the uplink of the OLT to the edge/cloud. The routes of the network slices and their extension VLANs to the edge/cloud are illustrated in Figure 19. This specific architecture follows the specifications described in [ETSIGR] and [POSA22]. Finally, in order to monitor the power consumption, a smart PDU is installed in the VIS. The PDU powers the PON elements as well the cameras and the robot arm. When it comes to the edge/cloud, there are three Virtual Machines (VMs) set up, two of them with GPU capability for running the vision inspection models and one for the control of the COBOTTA. The COBOTTA is controlled via an external middleware running in the edge/cloud which sends different commands depending on the output of the AI model. The physical setup Is shown in Figure 25.

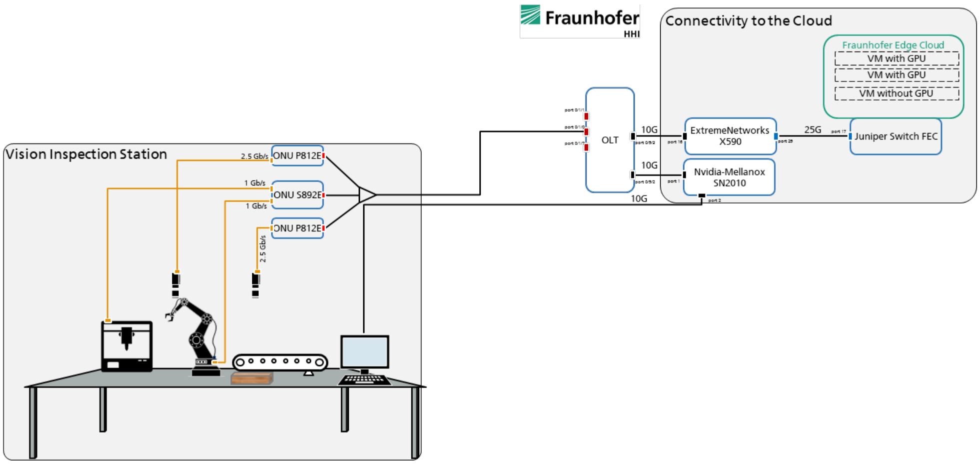


Figure 19: Testbed architecture and network slicing configuration.

Figure 20: Basler Camera.

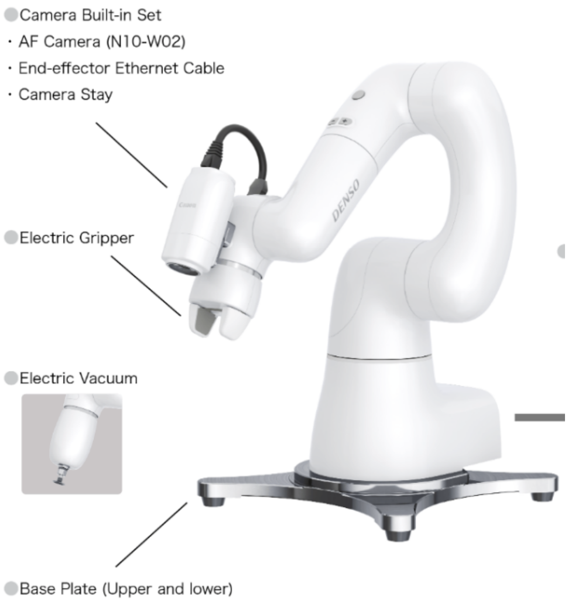


Figure 21: COBOTTA IP30.

The sequence diagram in Figure 22 and Figure 23 explains the entire vision inspection process related to the camera used to sort the objects (camera 1). The main methods involved in the communication between devices are sendVideo, sendDecision, send Command and sendLabelledVideo. The camera calls the sendVideo method to share the recorded images of the objects to be inspected by the AI model. ONU1 forwards the traffic associated with the images to the OLT and the OLT forwards them to the edge/cloud for processing by the AI. The AI model processes the data sent by the cameras and classifies the objects as faulty or non-faulty. The AI model calls the sendDecision method to share the result of the classification with the middleware, which invokes the sendCommand method to instruct the robot on the proper action. Given that the goal of our VIS is to be able to distinguish between faulty (with residue) and non-faulty objects (Figure 24), the two actions will be “discard” and “process”. The robot arm places the faulty objects in a tray (discard), and the non-faulty ones on the conveyor belt for further analysis by the second camera (process).

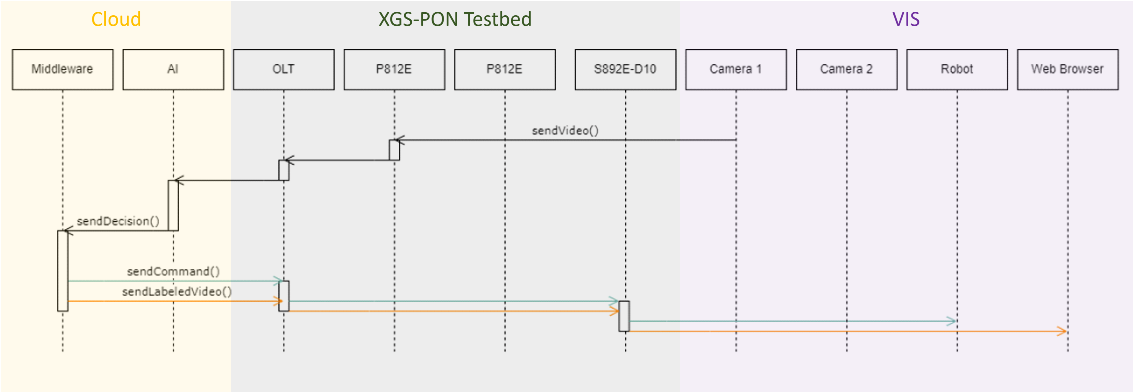


Figure 22: Testbed architecture and network slicing configuration.

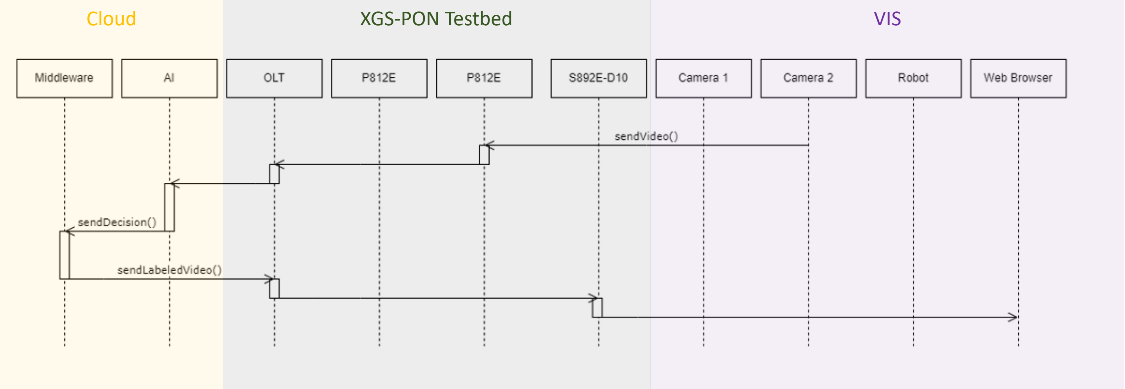


Figure 23: Sequence diagram for camera 2 operation.

Figure 24: Faulty(left), non-faulty (right) objects.

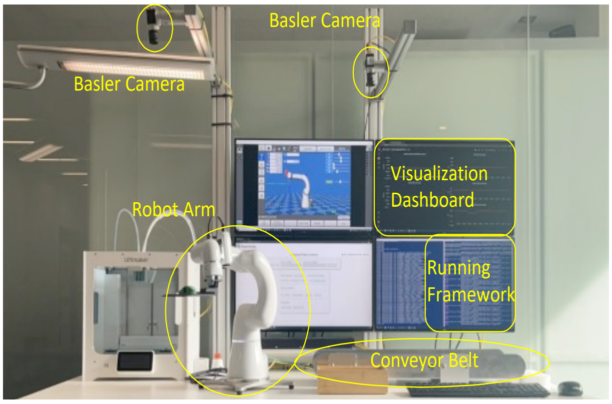


Figure 25: Physical setup.

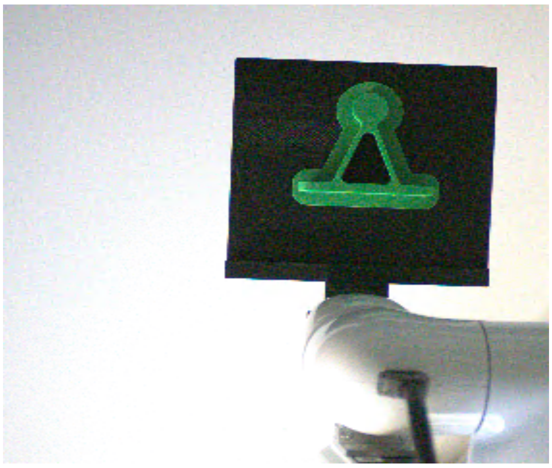
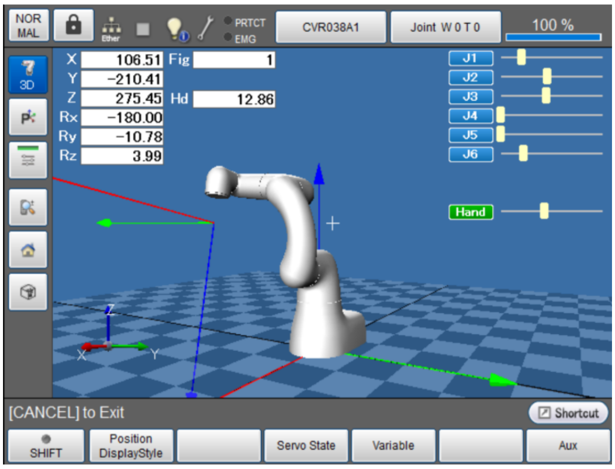


Figure 26: Camera 1 view from PylonViewer.



X 106.51
Y −210.41
Z 275.45
Rx −180.00
Ry −10.78
Rz 3.99
Fig 1
Hd 12.86
[CANCEL] to Exit

Figure 27: VirtualTP’s main screen.

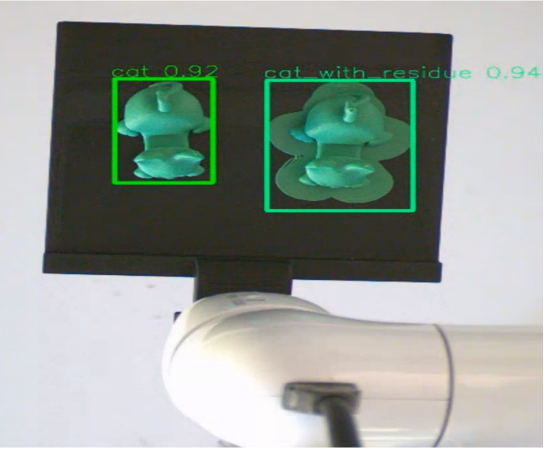


Figure 28: Classification output first camera.

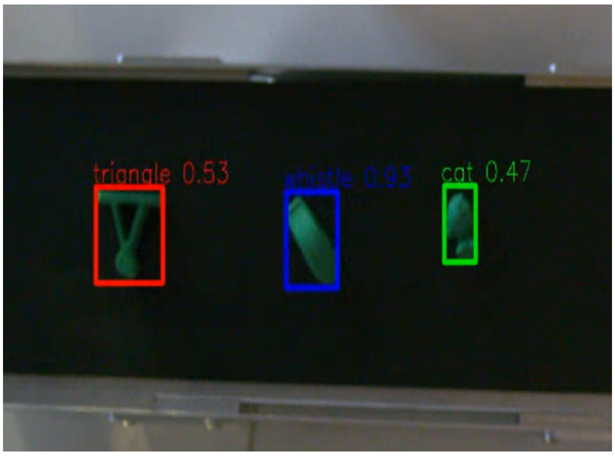


Figure 29: Classification output second camera.

The middleware also calls the sendLabelledVideo command, which is used by the Web Browser to show the results of the classification in real-time within the lab premises (Figure 25).The sequence diagram in Figure 23 shows the routine for the camera located on top of the conveyor belt (camera 2). The process is identical to the one for camera 1 with the exception that no command is sent to the robot since the objective of the camera 2 is to perform an additional screening of the objects on the conveyor belt.

The VIS hardware (cameras and robot) can be controlled by means of two proprietary software's: VirtualTP and PylonViewer. VirtualTP can receive commands from the middleware to control the robot remotely, however the capabilities of VirtualTP extend further than remote control. PylonViewer offers a GUI to configure the cameras and fine tune the recording quality (Figure 26). It is also capable of managing the robot autonomously through a Graphical User Interface (GUI) which can be used for testing and programming (Figure 27). The outcome of the vision inspection models based on the captures from camera1 and camera2 are provided in Figure 28 and Figure 29,respectively.

Monitoring of the PoC

While the selection process takes place, a real-time telemetry framework runs in the background to collect crucial analytics about the PoC operation (data rates, energy, etc.). The framework can provide real-time visibility with second granularity into the network's energy consumption and traffic data. The high-level architecture diagram in Figure 30 shows how the different components of the framework interact together. At the bottom we have the energy source, renewable or not, which feeds the ICT infrastructure. From the infrastructure, the network and energy data streams are processed by the data pipeline described in [BESH23] with an updated Kafka broker. We redesigned the Kafka broker by increasing the number of devices from which data is collected (Figure 31). Each network device has its own topic, which is then divided into as many partitions as, the number of data outlets (ports, sockets, etc.), available. Data consumers can selectively query only the information they are interested in, reducing the network overhead associated with data transfer and by limiting the number of topics. Figure 32 and Figure 33 show the telemetry retrieval process for traffic and energy data, respectively, as a sequence diagram.

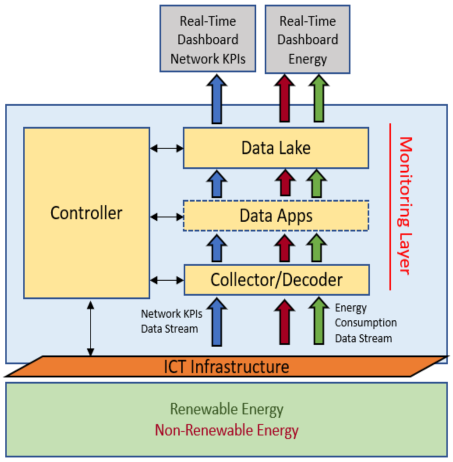


Figure 30: Framework architecture.

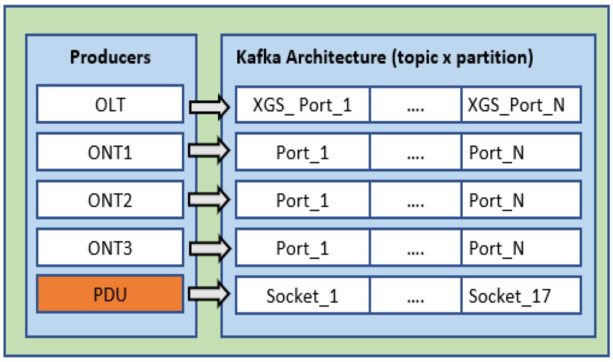


Figure 31: Kafka architecture.

The code running in the edge/cloud starts the traffic monitoring process (Figure 32) by instructing the OLT on how to configure the ONUs via a method called configureNetconf which carries the XML commands needed to configure the telemetry subscriptions on the network devices. The OLT sends a Netconf sendSubscr command to all the ONUs specifying the needed data and the retrieval granularity. The sendSubscr command configures the ONUs to send traffic data (throughput, packet loss etc.) every 10 seconds to the OLT via the sendTrafficData command. The latter sends the data from the OLT to the edge/cloud where it is displayed in a Grafana dashboard accessible on-premise via the Web Browser thanks to port forwarding. The code running in the cloud starts the energy collection process as well (Figure 33). It sends a pollSNMP command to the PDU with information regarding the data to collect and the associated granularity. Once the data is ready, the PDU forwards it to the cloud for display in the Grafana dashboard, where also the traffic data is shown.

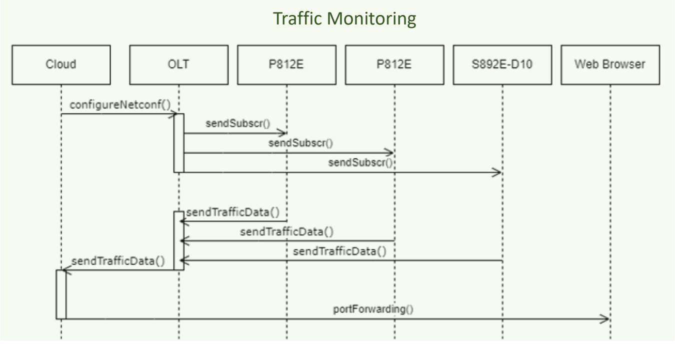


Figure 32: Sequence diagram for traffic monitoring.

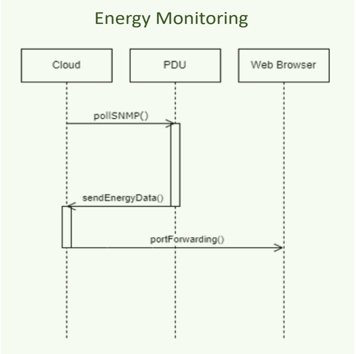


Figure 33: Sequence diagram for energy monitoring.

There has been a growing interest in the scientific community towards the CO_2 emissions of the telecommunications infrastructure due to the raising concerns related to global warming. A piece of evidence is the 22% increase of energy consumption of telecom networks in Germany from 2015 to 2020. In this regards we decided to make use of the telemetry provided by the PDU to study the carbon emissions and provide some projections over multiple scenarios. To prove the capabilities of our framework, we decided to model three different VIS setups involving a different number of cameras transmitting at different data rates. All the setups model a variation of the standard VIS shown in Figure 19. The first setup consists of 2 cameras transmitting at 1 Gbit/s, the second one of 4 cameras recording at 1 Gbit/s and the last one of 2 cameras recording at 4.5 Gbit/s. To further investigate the customization capabilities of our framework, we decided to categorize the devices in ICT devices and non-ICT devices.

ICT devices are the OLT and the ONUs while non-ICT devices are the robot arm, the conveyor belt and the cameras. In this section we extend the results obtained over the period of one hour to one year and to multiple contemporary-running VIS. The OLT available in the testbed can support up to 40 XGS-PON ports which means that we can scale up our computations to three different scenarios based on one OLT. For each scenario we compute the total amount of traffic generated, the energy required to run, the Network Carbon Intensity energy (NCIe) and the total Kg of Emitted CO_2 (ECO_2) [ITUT22]. Specifically, the last two metrics are also compared to the expected emissions in 2028 when, e.g., Germany plans to expand the use of renewable energies. The results are shown in Figure 34. Scenario 1 leads the way as the most energy hungry and polluting scenario, which makes sense given the much higher number of devices involved with proportionally not as much traffic flowing. In fact, scenario one has ~77% higher energy consumption than scenario 3 and only ~10% more traffic, which also justifies the worse performance in terms of NCIe. It is interesting to notice that in every scenario the highest energy consumption, hence emissions, is due to non-ICT equipment. The results also show that when using more renewable energies, the emissions decrease substantially for every scenario by up to ~75%. If instead we considered an extreme scenario, such as all the ICT equipment running on renewable energy, then all the emissions (NCIe, ECO_2) will be zeroed leaving us with only the emissions of the non-ICT devices. By considering the opposite scenario we would be left with the emissions of the ICT devices only.

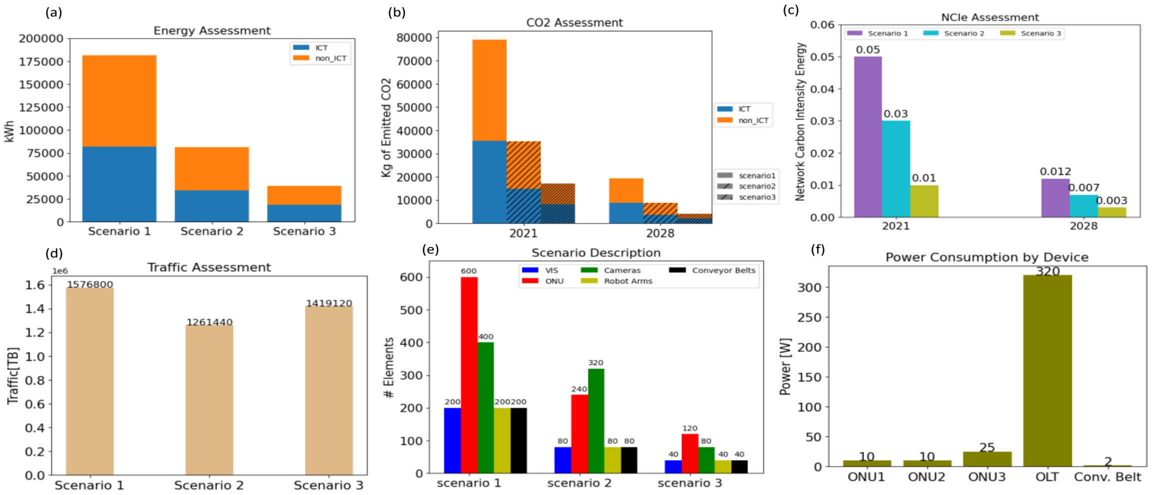


Figure 34: (a) energy assessment; (b) CO_2 assessment; (c) NCIe assessment; (d) traffic assessment; (e) scenario description (f) power consumption of other devices.

Major findings

The following insights were gained when setting up and executing the PoC:

- Latency sensitive and bandwidth hungry industrial use cases can be successfully realized in a scenario where PON is used as the base broadband connectivity solution.
- It has been challenging to set up an E2E precision time protocol to accurately measure the E2E latency between the vision Inspection station and the cloud as the multitude of networking devices in the middle have compliancy issues with the protocol, which has to be improved.
- The power consumption monitoring has been realized using smart power meters. There is a need from component manufacturer to incorporate real-time monitoring of power consumption of their networking components.
- This use case imposes a significant upstream bandwidth requirement compared to a negligible downstream amount. This is totally in contrast to the home users, where the downstream is larger in capacity. This may require modifications of the PONs for taking into account different asymmetric bandwidth flows.

References

[BESH23]	Behnam Shariati, et al. "Telemetry Framework with Data Sovereignty Features." Optical Fiber Communication Conference. Optica Publishing Group,2023.
[ETSIGR]	Standardization Document ETSI GR F5G 008 V1.1.1
[POSA22]	Pooyan Safari, et al. "Edge Cloud Based Visual Inspection for Automatic Quality Assurance in Production." 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP). IEEE, 2022.
[ITUT22]	Recommendation ITU-T L.1333 (2022), "Carbon Data Intensity for Network Energy Performance Monitoring".

10. PoC report: Edge/Cloud-based control of automated guided vehicles

Overview

This PoC demonstrates the concept of cloud-based control of automated guided vehicles (AGVs) and robots in a factory workshop environment, based on an underlying PON infrastructure. The PON is set up between the Fraunhofer HHI (F5G OpenLab) and Fraunhofer IPK [MBA23] (see Figure 35). In order to demonstrate the performance of the PON, parts of the AGV and robot control software are migrated to an edge cloud at HHI, while the components/hardware to be controlled (AGVs and robots) are located at IPK [PSA22]. The cloud-based AGV and robot controls are shown in two simplified scenarios. Here, a simple pick & place task is demonstrated with a marker-based localization of the robot arm on the AGV relative to a stationary object. In this case, the marker-based localization functionality is executed on the edge cloud. The features of this PoC include an E2E AGV control loop and cloud-based control of robots, powered by a PON-based fibre connectivity between the edge cloud and production site, where the factory shop floor is served by WiFi6-enabled ONUs.

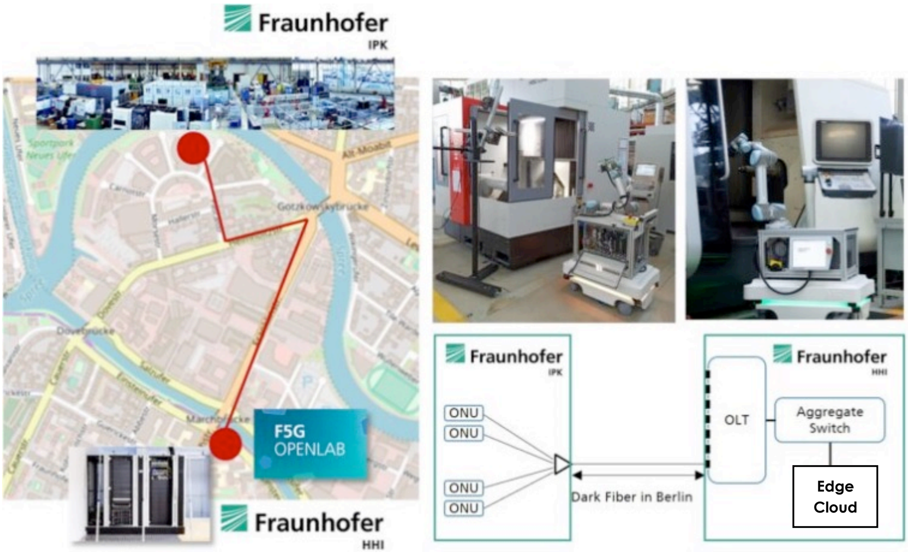


Figure 35: Setup of the use case in the city of Berlin.

Topics of Investigation

In this PoC we focus on a mobile manipulation scenario. An Autonomous Mobile Robot (AMR) equipped with a robotic arm is tending a machine tool where image analytics and control logic are moved to the edge/cloud. The set-up system is based on the Data-Distribution Service (DDS) for real-time systems. This middleware realizes a broker-less publish-subscribe architecture to link the individual services with each other.

There is a cloud computing deployed as edge DC which is connected to the shop floor via a PON. At the shop floor WiFi6 ONUs are installed, supporting the roaming of clients.

The PoC consists of the following services based on the ROS2 framework (see Figure 36):

- AGV interface: running on the AGV-edge
- Arm interface: running on AGV-edge
- Camera interface: running on the AGV-edge
- Gripper interface: running on the AGV-edge

- Navigation: running on cloud (edge DC)
- Image recognition and analysis: running on cloud (edge DC)
- Motion planning for robotic arm: running on cloud (edge DC)
- Visualization of planning scene: running on HMI-edge at shop floor
- Orchestration of the scenario: running on cloud (edge DC)

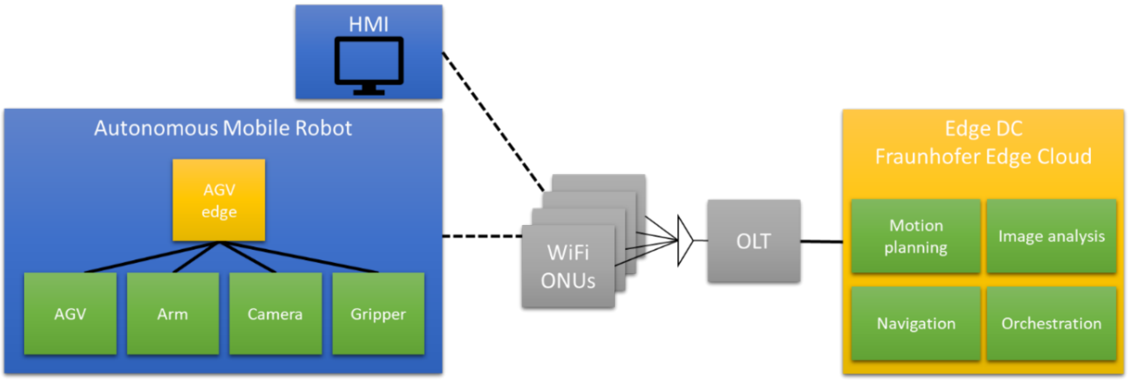


Figure 36: Components involved in the PoC.

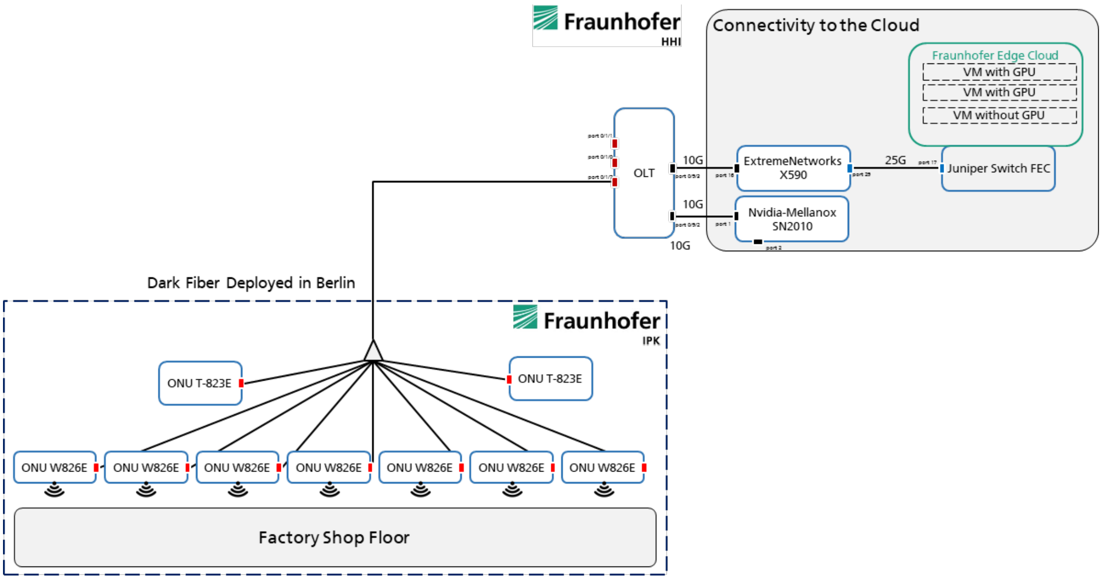


Figure 37: Networking setup.

Network Architecture of the PoC

The testbed (Figure 37) for the proposed use case spans over two different sites, one at Fraunhofer HHI and one at Fraunhofer IPK. IPK runs a factory shop floor in which the AGV and the robots run. We have installed several WiFi ONUs and one industrial ONU at IPK for the purpose of this demo. The OLT however is located at HHI and the ONUs at IPK are connected to a single XGS-PON port of the OLT. The uplink of the OLT is then connected to the edge cloud through an aggregation switch. As shown in Figure 36, the setup involves an AMR comprising an AGV (MiR 100), a robot arm (UR 5), a camera (Microsoft Azure Kinect), and a gripper (Weiss IEG 76-030). As HMI in the shop floor a laptop is used. The shop floor is provided broadband connectivity through an XGS-PON testbed with multiple WiFi ONUs and industrial ONU. When it comes to the edge/cloud, there is one Virtual Machine (VM) set up. The services are distributed on this VM.

Workflow of the PoC

In this section, we describe the execution steps of the PoC. The robot first places a part in the material shelf, then picks it up again and places it into a CNC milling machine. The sequence diagram in Figure 38 explains the example process of driving to the material shelf and picking up a part.

Initially the robot is turned off and placed in a room that is shielded from all ONU-APs except ONU-6. Once the robot is turned on *(turnOn)*, the internal Industrial PC (IPC) connects through a WiFi6 network adapter to the ONU-6 in the room. Once the network connection is established, the ROS 2 driver of the robot is started *(startDriver)*. This enables all core functionalities of the robot, e. g. driving with the mobile base (AGV), manipulation with the robotic arm, the gripper and visual perception through the camera. Since the robot's IPC is supposed to act as a driver only, all path planning, navigation and perception processing is done in the VM on the edge-DC. These so-called Computation Services (CSs) are now started *(startComputationServices, startNavigation, startMotionPlanning, startVisualPerception)*, which generates a lot of traffic between the VM and the robot, because the initial state of the robot and all of its sensor data is transmitted to the VM. With the CSs operational, a visualization PC / HMI is connected to an ONU, which displays the map view, path planning status and visual perception data in a 3D view *(start Visualisation)*.

Now that the robot is operational, the orchestration of the PoC starts generating commands. First the navigation gets a request to compute a path to the material shelf *(navigateToShelf)*. When the path is generated, a feedback loop is executed between the VM and the robot over the PON network, sending velocity commands to the robot and checking the position on the map in real-time. Arrived at the material shelf, the orchestration requests the motion planning service to compute a trajectory for the robotic arm, so that the arm moves its eye-in-hand camera to a scanning position *(moveArmToScanShelf)*. Arrived in this arm pose, the machine detection service is called *(detectShelf)*, which allows the full spatial detection of the material shelfs geometry through an attached ChArUco marker. With the shelf detected, a part is picked up from the storage surface of the AGV and subsequently inserted into the shelf. With the part inserted, the robot arm retracts out of the shelf *(retractArmFromShelf)*. To demonstrate also the process of picking up pieces from a storage shelf, the aforementioned procedure is executed again, but when entering the shelf this time, the part is extracted from the holder *(moveArmToGripPosition)* in the shelf and placed *(moveToPlaceOnAGVPosition)* on top of the storage surface of the AGV.

Now the AGV is loaded with a part (blank part), which is to be loaded into a CNC milling machine. The robot first has to navigate and drive to the machine the same way as explained above. When arrived the machine's marker is scanned to detect its exact position and geometry. Then the arm picks up the blank from the AGV's storage surface and enters the machine. Inside of the machine is a fixture, into that the robot arm has to insert the part. When done, the arm is retracted from the CNC machine and finally reaches its idle state, which terminates the orchestration.

Reached the idle state, the hosts are turned off in reverse order. First the visualization PC is turned off, followed by the CSs in the VM and finally the robot. Figure 39 show the components of the mobile robot Including the mobile base, manipulator, camera, and gripper. A screenshot of the visual motion planning environment of for robot mobile is also shown in Figure 40.

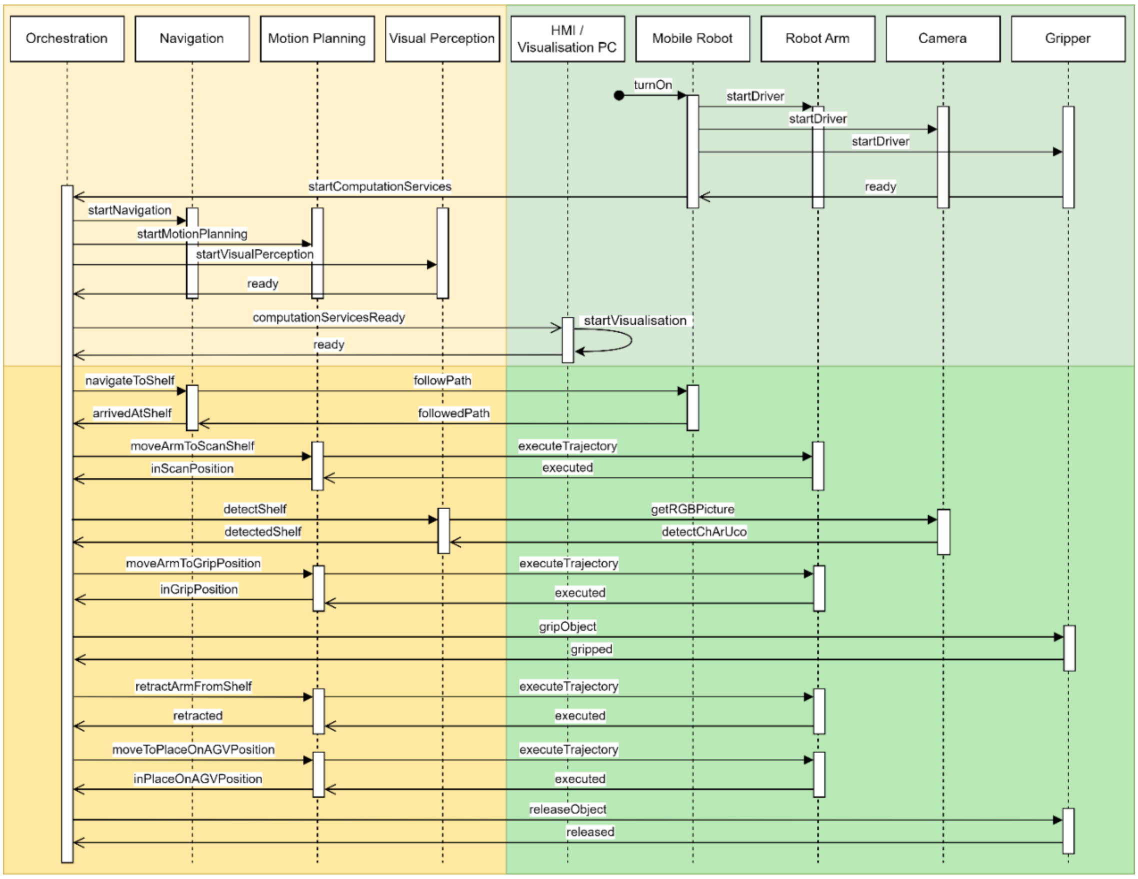


Figure 38: Sequence diagram of all used services in VM (orange shade) to control AGV and robot in shop floor (green shade) with the setup first (darker shade) and then picking up a part from a material shelf (brighter shade) as an example.

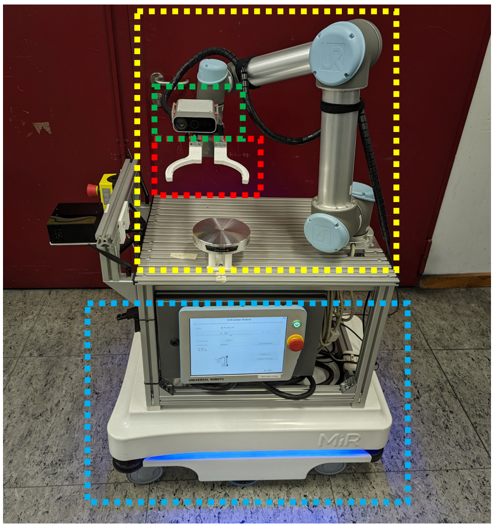


Figure 39: Components of the robot: mobile base (blue), manipulator (yellow), camera (green), gripper (red).

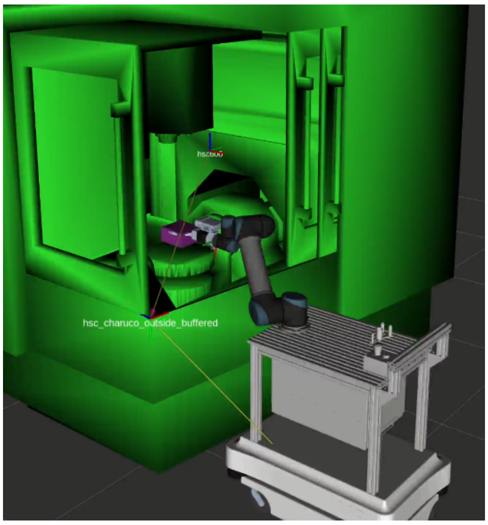


Figure 40: Visual motion planning environment on the VM.

Monitoring of the PoC

We have used our telemetry framework to record the traffic exchanges between the shop floor and the edge cloud. In this section, we look into the traffic pattern recorded while carrying out the demo. The testbed includes network slicing to guarantee the strict latency requirements associated to the use case. Only one slice is configured and it carries all the traffic associated with AGV control and visualization (shop floor map, sensor). The AGV connects to WiFi ONU 6 at startup (Figure 41), however the bulk of the traffic, both towards and from the OLT, is generated when the services are started. The sudden interruption of data transmission to the OLT in Figure 41 is due to a handover to WiFi ONU2 (Figure 42). The services are responsible to receive sensor data and compute the navigation path. The second traffic spike happens when the first visualization starts. A second visualization is later started and then quickly turned off to help the robot navigate a trickier spot. The visualization traffic is also visible in Figure 41, since the industrial ONU (Figure 43) is only connected to the laptop where they are shown. The total aggregated traffic exchange between the shop floor and the edge cloud is shown in Figure 44 resulting from monitoring the Interface on the aggregation switch between the OLT and the edge cloud. The traffic generated during the two placement tasks is clearly visible in Figure 42. Such traffic represents the set of instructions sent from the edge cloud to the AGV to properly command the placement operation.

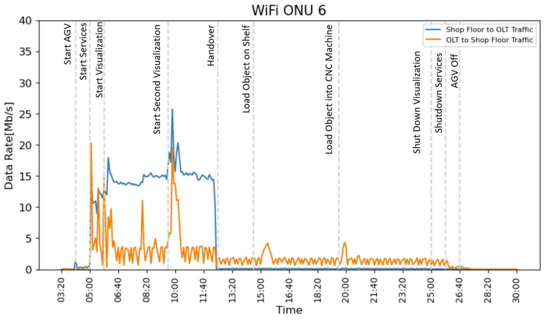


Figure 41: Traffic exchange ONU6.

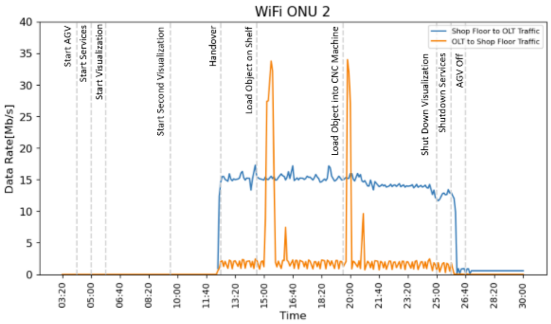


Figure 42: Traffic exchange ONU2.

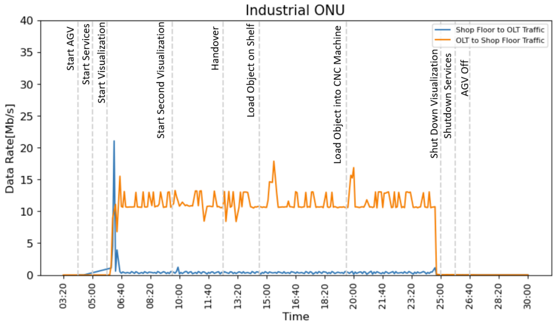


Figure 43: Traffic exchange for industrial ONU.

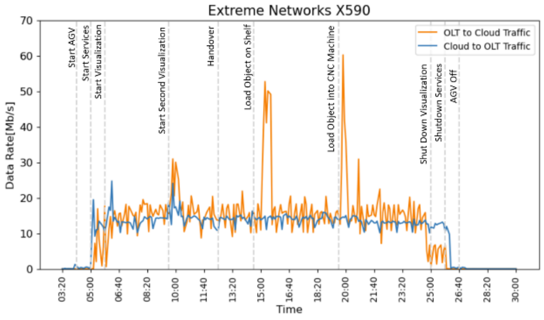


Figure 44: Traffic exchange cloud.

Major findings

The following insights were gained when setting up and executing the PoC:

High probability of traffic between two ONUs, not only ONU-OLT: The shown scenario incorporates an edge cloud. This is not yet common for industry applications. If there is no central processing needed at the OLT side, the ONUs directly communicates to each other.

Multicast traffic for DDS middleware: The DDS middleware recommends the use of IP multicast at least for the service discovery, which means discovering the different nodes (participants) in the network. There is a configuration without multicast, but then all nodes have to be known in advance and statically configured. This leads to a very high configuration effort and makes the system less robust and more error-prone. Additionally, the time for starting up the systems increases.

High signal strength of WiFi APs leads to small probability of roaming: The WiFi6 ONUs have high signal strength, which is very impressive. During the experiments, this made roaming tests difficult as the hand over between different ONUs could not be tried easily.

Latency of cloud control in PON installation: The PON installation matches the latency requirements of cloud-controlled robots as shown in this proof of concept.

References

[PSA22] Pooyan Safari, Behnam Shariati, David Przewozny, Paul Chojecki, Johannes Fischer, Ronald Freund, A. Vick, M. Chemnitz. "Edge Cloud Based Visual Inspection for Automatic Quality Assurance in Production." in Proc. of 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP).

[MBA23] Mihail Balanici, Behnam Shariati, Pooyan Safari, Paul Chojecki, Moritz Chemnitz, David Przewozny, Johannes Karl Fischer, Ronald Freund, "F5G OpenLab: Enabling Twin Transition through Ubiquitous Fiber Connectivity." in Proc. of International Conference on Transparent Optical Networks (ICTON) 2023.

11. Green all optical network and green enablement by an optical network

On 11 October 2019, the European Commission published the European Green Deal presenting a list of policy initiatives aimed at driving Europe to reach net-zero global warming emissions by 2050. The goal of the European Green Deal is to improve the well-being of people by making Europe climate-neutral and protecting Europe's natural habitat for the benefit of people, planet and economy.

This section discusses possible approaches that focus on the realisation of (1) a Green all optical network, which is using the Green ICT concept by minimizing the carbon emissions in an all optical network and (2) using the optical network to reduce the carbon emissions of other Industrial sectors, using the ICT for Green concept.

Green all optical network

This scenario reflects the situation, where solutions are applied to minimize the carbon emissions in an all optical network.

Green enablement by an optical network

This scenario reflects the situation that an optical network solution is applied in an industrial scenario to reduce its carbon emissions.

However, in order to ensure that the applied optical network solution, indeed, reduces carbon emissions in an industrial scenario, a methodology and assessment need to be followed. It is recommended that the Life Cycle Assessment (LCA) methodology specified in the ITU-T L.1480 specification, i. e., (1) goal and scope, (2) LC inventory analysis, (3) LC impact assessment and (4) interpretation of results, is followed. In addition, the methodology specified in L.1480, is complemented by the quantified method described in Section 6.4 of the AIOTI "IoT and Edge Computing Carbon Footprint Measurement Methodology", report, Release 2.0.

Methodologies

A method of calculating the avoided carbon emissions in industrial sectors, when ICT is applied, is presented in (Alliance for IoT and Edge Computing Innovation 2023) and is listed in Annex II. In particular, as described in (Alliance for IoT and Edge Computing Innovation 2023) this is a quantitative method, where the avoided emissions in vertical/industrial sectors, when applying ICT, can be calculated for all LCA phases, excluding the LCA re-use and recycling phases. This equation includes as well factors, as type of service and the load that the ICT infrastructure needs to support over a period of time. In particular, for the calculation of the ICT infrastructure emissions in the operation/use LCA phase, the quantitative method specified in ITU-T L.1333 is proposed.

Energy savings using passive optical networks

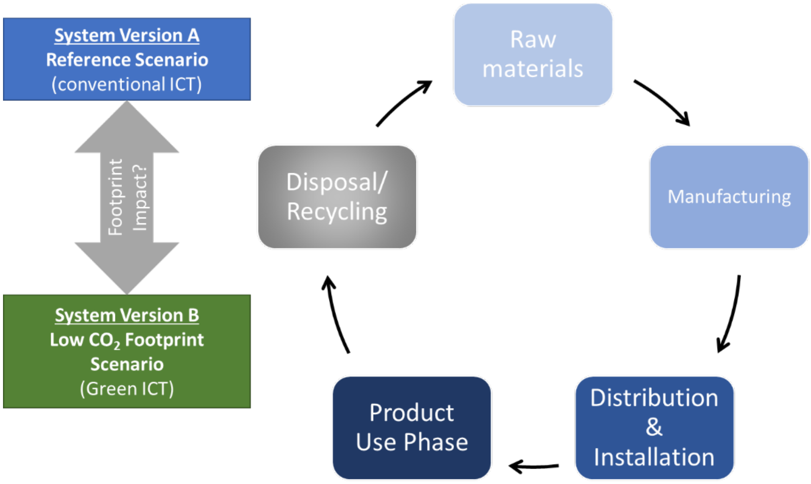


Figure 45: Modelling the carbon footprint of different system versions requires life cycle assessment including detailed technical data as well as measurements in testbeds for product use phase.

ICT plays an important role in enabling the digital transformation of vertical sectors such as the manufacturing industry. Introducing new ICT solutions to a vertical sector can thus even support decarbonization. There exist methodologies to quantify the net carbon footprint improvement that can be obtained by specific ICT solutions (see Section 0). One important part in these methodologies is the negative first order Impact of the introduced ICT solution itself. Therefore, it is imperative to also compare competing ICT solutions with each other. In order to fully quantify the carbon footprint impact of two system versions, a full lifecycle assessment would be required (Figure 45), where a green ICT solution is compared to a reference conventional ICT solution. However, in scenarios, where the total carbon footprint is dominated by the product use phase, such a comparison can be simplified by analysing the power consumption during the product use phase.

The following example considers a scenario from the manufacturing vertical sector, where a production site shall be equipped with a networking solution connecting a number of various access points to a DC. The example considers to networking solutions: (a) a network topology based on standard Ethernet switches and (b) a PON. A schematic of the network topologies is shown in Figure 46.

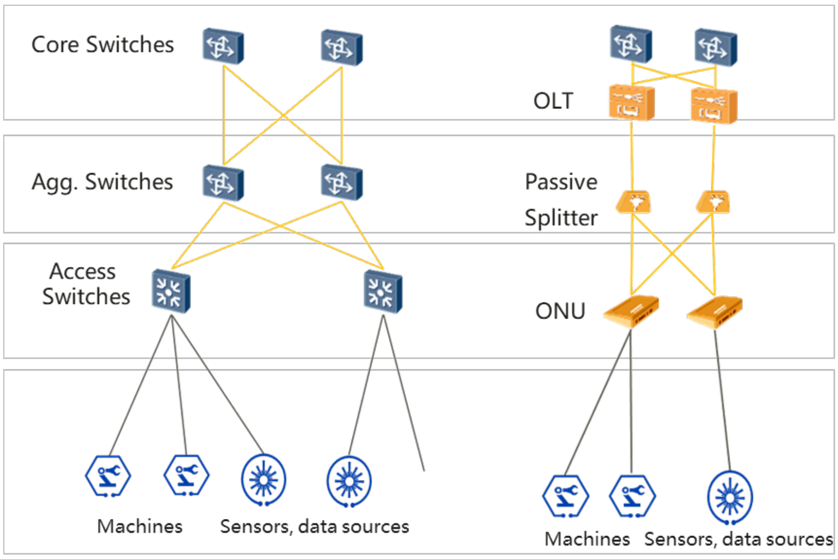


Figure 46: Switched network (left) vs. PON (right) architecture, including optical network units (ONU) and optical line terminal (OLT).

While the switched network uses a cascade of switches to aggregate the traffic from the access points to the DC, the PON uses a completely passive, fibre-based infrastructure to aggregate the traffic of the optical network units (ONU) to the optical line terminal (OLT). This eliminates the need for active, power consuming aggregation switches. Assuming a requirement of 1GE per access point, access switches with 24×1GE client ports, aggregation and core switches with 48×10GE client ports results in 167 access switches, 8 aggregation switches and 2 core switches for 4000 access points. Similarly, 2 core switches, 2 OLTs, 600 ONUs with 4×1GE client ports and 200 ONUs with 8×1GE client ports are required to connect the same amount of access points. Based on the power consumption of these units and the corresponding pluggable modules the total power consumption of both ICT solutions can be compared. Figure 47 shows the resulting power consumption for scenarios ranging from 1000×1GE access points to 6000×1GE access points. The results show that the power consumption reduction increases with an increasing number of access points that have to be served, reaching up to 35 % reduction for 6000 access points.

The results show that it is important to carefully consider the right ICT solution for each use case. In particular, passive optical technologies can play a crucial role for decarbonizing vertical sectors due to their improved footprint over conventional switch-based networking solutions in scenarios with a large number of access points.

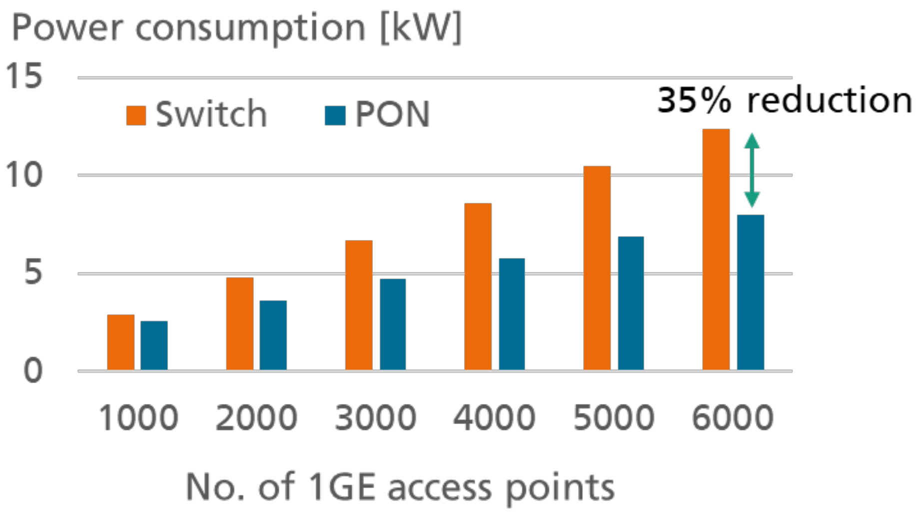


Figure 47: Reduction of power consumption enabled by PON in a scenario with a large number of access points.

12. Conclusions and recommendations

- **Computing continuum together with optical communication provides support for very high-end IoT applications** needing high bandwidth, low delay, consistent and sustained performance, and high security and isolation.
- **Computing continuum platforms with optical communication provide support for mission critical IoT applications** due to cloud native compute reliability together with well-known and proven optical network reliability.
- **Flexible placement of IoT workload without constraints in the optical network** depending on the application needs.
- **Real-time applications requirements** are most cost and energy efficiently supported by optical communication technologies.
- **It is recommended that optical network support for computing continuum** is designed and standardized.
- **It is recommended to standardize integration of optical network and cloud technologies** for a powerful computing continuum.
- **It is recommended to design more flexible optical communication systems**, e.g., for dynamic optical cut-through, on-demand provisioning, and flexible re-adjustments of the resource allocation.
- **It is recommended to evolve the F5G optical network architecture** to make it an even more scalable architecture for mass-deployment of a plethora of new IoT devices and applications.
- **It is recommended that the challenge of business models in the space of computing continuum** is studied and the administrative boundaries of those business models are defined such that interface specifications at those boundaries and the appropriate isolation technologies on network and compute level can be designed.
- **It is recommended to extend the slicing concept to cover also edge compute resources** such that joint operation and management of computing continuum and optical communication can be deployed.
- **It is recommended to standardize features** to ease the deployment and operation of optical communication enhanced computing continuum platforms.
- **It is recommended to research the use of optical communication and fibre technologies** to be used for optical sensing oriented applications.
- **It is recommended to research photonics components to be integrated into optical-oriented computing continuum platforms** for application acceleration, sensing, and display of IoT applications.
- **It is recommended that IP and optical vendors integrate more precise monitoring of power consumption in their components**, along with interfaces for real-time monitoring.
- **Industrial use cases may impose huge upstream traffic and negligible downstream traffic** (e. g. the vision inspection use case). It is recommended that such an asymmetric fashion of traffic flow be considered in equipment design.
- **Passive optical technologies can play a crucial role for decarbonizing vertical sectors** due to their improved footprint over conventional switch-based networking solutions. It is recommended to further exploit this technology for different use cases in large scale installations.

Annex I. Template for use case description

X. Title of use case

X.1 Description

- Provide motivation of having this use case, e. g., is it currently applied and successful; What are the business drivers, e. g., several stakeholder types will participate and profit from this use case
- Provide on a high level, the operation of the use case, i. e., which sequence of steps are used in this operation?

X.2 Source

- Provide reference to project, SDO, alliance, etc.

X.3 Roles and Actors

- Roles: Roles relating to/appearing in the use case
  - Roles and responsibilities in this use case, e. g., end user, vertical industry, Communication Network supplier/provider/operator, IoT device manufacturer, IoT platform provider, Insurance company, etc.
  - Relationships between roles
- Actors: Which are the actors with respect to played roles

Actors & Roles

X.4 Pre-conditions

What are the pre-conditions that must be valid (be in place) before the use case can become operational?

X.5 Triggers

- What are the triggers used by this use case?

X.6 Normal Flow

- What is the normal flow of exchanged data between the key entities used in this use case: devices, IoT platform, infrastructure, pedestrians, vehicles, etc?

X.7 Alternative Flow

- Is there an alternative flow?

X.8 Post-conditions

- What happens after the use case is completed?

X.9 High Level Illustration

- High level figure/picture that shows the main entities used in the use case and if possible their interaction on a high level of abstraction.

X.10 Potential Requirements

This section should provide the potential requirements and in particular the requirements imposed towards the underlying communication technology.

These requirements can be split in:

- Functional requirements

(to possibly consider them – but not limited to – with respect to the identified functions/capabilities)

- Non-functional requirements – possible consideration includes:
  - Flexibility
  - Scalability
  - Interoperability
  - Reliability
  - Safety
  - Security and privacy
  - Trust

Functional Requirements

- Real-time communication with the stakeholders in case of emergency.
- Reliable communication between the stakeholders.
- Scalable communication between systems to interconnects different critical infrastructures.
- Standard-based communication between critical infrastructure to align emergency information exchange with new and legacy systems.

Non-Functional Requirements

- Secure communication between the emergency bodies due to the information nature.
- Interoperability between communication protocols (linked also with the possibility to use standard communication protocols between the systems).

X.11 Optical Network specific Requirements

Annex II. AIOTI method calculating avoided carbon missions (Sec. 11)

A method of calculating the avoided carbon emissions in industrial sectors, when ICT is applied, is presented in (Alliance for IoT and Edge Computing Innovation 2023) and is introduced below. In particular, as described in (Alliance for IoT and Edge Computing Innovation 2023) this is a quantitative method, where the avoided emissions in vertical/industrial sectors, when applying ICT, can be calculated for all LCA phases, excluding the LCA re-use and recycling phases. This equation includes factors as well, as type of service and the load that the ICT infrastructure needs to support over a period of time. In particular, for the calculation of the ICT infrastructure emissions in the operation/use LCA phase, the quantitative method specified in ITU-T L.1333 is proposed.

The proposed Total Avoided Carbon Emissions equation is provided below and is visualized in Figure 48.

Equation 1: $TAE_{(l)(ts)} = (T\_EBs\_nict_{(l)(ts)} + T\_\ EictBs_{(l)(ts)}) - (T\_EGr\_nict_{(l)(ts)} + T\_\ EictGr_{(l)(ts)})$,

where:

- $\boldsymbol{TAE}_{(l)(ts)}$: Total Avoided Carbon Emission Scenario for: (1) the complete LCA, excluding the Reuse and Recycle phases, (2) for a certain Load and (3) for a type of service, e. g. follows the classification specified by ITU-T for 5G type of services;

- $\boldsymbol{T\_EBs\_nict}:_{(l)(ts)}$: Total Carbon Emission Scenario, for Baseline scenario (*Bs*), but excluding the carbon emission of the applied ICT infrastructure, i. e., carbon emissions of *ictBs*, for: (1) the complete LC phases, excluding the Reuse and Recycle phases, (2) for a certain Load and (3) for a type of service, e. g. follow the classification specified by ITU-T for 5G type of services, where:

$$\boldsymbol{T\_EBs\_nict_{(l)(ts)} = T\_EBs\_nict^{M}_{(l)(ts)} + T\_EBs\_nict^{P}_{(l)(ts)} + T\_EBs\_nict^{O}_{(l)(ts)} + T\_EBs\_nict^{D}_{(l)(ts)}}.$$

- $\boldsymbol{T\_EictBs}_{(l)(ts)}$: Total ICT Carbon Emission for Baseline Scenario, i. e., *ictBs,* for: (1) the complete LCA, excluding the Reuse and Recycle phases, (2) for a certain Load and (3) for a type of service, e. g. follows the classification specified by ITU-T for 5G type of services, where:

$$\boldsymbol{T\_EictBs_{(l)(ts)} = T\_EictBs^{M}_{(l)(ts)} + T\_EictBs^{P}_{(l)(ts)} + T\_EictBs^{O}_{(l)(ts)} + T\_EictBs^{D}_{(l)(ts)}}.$$

An example of calculating $T\_EictBs_{(l)(ts)}$ in the LC use/operation phase can be realized by using the approach defined in ITU-T L.1333: Carbon data intensity for network energy performance monitoring.

- $\boldsymbol{T\_EGr\_nict}_{(l)(ts)}$: Total Carbon Emission Scenario, for Green enabled scenario (*Gr*), but excluding the carbon emission of the applied ICT infrastructure, i. e., carbon emissions of *ictGr*, for: (1) the complete LCA, excluding the Reuse and Recycle phases, (2) for a certain load and (3) for a type of service, e. g. follow the classification specified by ITU-T for 5G type of services, where:

$$\boldsymbol{T\_EGr\_nict_{(l)(ts)} = T\_EGr\_nict^{M}_{(l)(ts)} + T\_EGr\_nict^{P}_{(l)(ts)} + T\_EGr\_nict^{O}_{(l)(ts)} + T\_EGr\_nict^{D}_{(l)(ts)}}.$$

- $\boldsymbol{T\_EictGr}_{(l)(ts)}$: Total ICT Carbon Emission for Green enabled Scenario, i. e., *ictGr*, for: (1) the complete LCA, excluding the Reuse and Recycle phases, (2) for a certain Load and (3) for a type of service, e. g. follow the classification specified by ITU-T for 5G type of services, where:

$$\boldsymbol{T\_EictGr_{(l)(ts)} = T\_EictGr^{M}_{(l)(ts)} + T\_EictGr^{P}_{(l)(ts)} + T\_EictGr^{O}_{(l)(ts)} + T\_EictGr^{D}_{(l)(ts)}}.$$

  - An example of calculating $T\_EictGr_{(l)(ts)}$ in the LC use/operation phase can be realized by using the approach defined in ITU-T L.1333: Carbon data intensity for network energy performance monitoring.

- Note that the superscripts **M**, **P**, **O**, **D**, shown in the equation terms introduced above and in Figure 48, denote that the carbon emissions calculations are related to the LC phases: Material, Product, Operation, Discard, respectively.

It can be derived that:

Equation 2: $T\_EBs\_nict^{M}_{(l)(ts)} = \sum_{m=1}^{LBs\_nict} EBs\_nict^{M}_{(m)(l)(ts)}$,

Equation 3: $T\_EBs\_nict^{P}_{(l)(ts)} = \sum_{m=1}^{LBs\_nict} EBs\_nict^{P}_{(m)(l)(ts)}$,

Equation 4: $T\_EBs\_nict^{O}_{(l)(ts)} = \sum_{m=1}^{LBs\_nict} EBs\_nict^{O}_{(m)(l)(ts)}$,

Equation 5: $T\_EBs\_nict^{D}_{(l)(ts)} = \sum_{m=1}^{LBs\_nict} EBs\_nict^{D}_{(m)(l)(ts)}$,

where:

- $EBs_{nict}{}^{M}_{(m)(l)(ts)}$ represents carbon emission of each product/component (m) used in in the Baseline scenario, excluding the ICT infrastructure, obtained in the LC Material phase; Note that in this case the subscripts (l) and (ts) can be discarded, since they are not relevant;
- $EBs_{nict}{}^{P}_{(m)(l)(ts)}$ represents carbon emission of each product/component (m) used in in the Baseline scenario, excluding the ICT infrastructure, obtained in the LC Production phase. Note that in this case the subscripts (l) and (ts) can be discarded, since they are not relevant;
- $EBs_{nict}{}^{O}_{(m)(l)(ts)}$ represents carbon emission of each product/component (m) used in in the Baseline scenario, excluding the ICT infrastructure, obtained in the LC Operation phase;
- $EBs_{nict}{}^{D}_{(m)(l)(ts)}$ represents carbon emission of each product/component (m) used in in the Baseline scenario, excluding the ICT infrastructure, obtained in the LC Disposal phase. Note that in this case the subscripts (l) and (ts) can be discarded, since they are not relevant;
- ***LBs_nict*** is the total number of products/components (m) used in the Baseline scenario, excluding the ICT infrastructure.

Note that the same type of equations can be derived for: $T\_EGr\_nict_{(l)(ts)}$; $T\_EictBs_{(l)(ts)}$; $T\_EictGr_{(l)(ts)}$.

Equation for Total ICT Avoided Carbon Emissions

Equation 6: $TAE\_ICT_{(l)(ts)} = T\_EictBs_{(l)(ts)} - T\_EictGr_{(l)(ts)}$,

where:

- ***TAE_ICT***$_{(l)(ts)}$: Total ICT Avoided Carbon Emission is a metric to measure the ICT carbon emission benefits, when replacing the ICT infrastructure used in the Baseline scenario, i. e., *ictBs*, with the ICT solution used in a Green enablement scenario, i. e., *ictGr*.

Note that in certain situations, e. g., including advanced ICT features, to reduce significantly $TAE_{(l)(ts)}$, it might result that $TAE\_ICT_{(l)(ts)}$ becomes to be a negative number, due to the carbon emissions additions of these advanced ICT features.

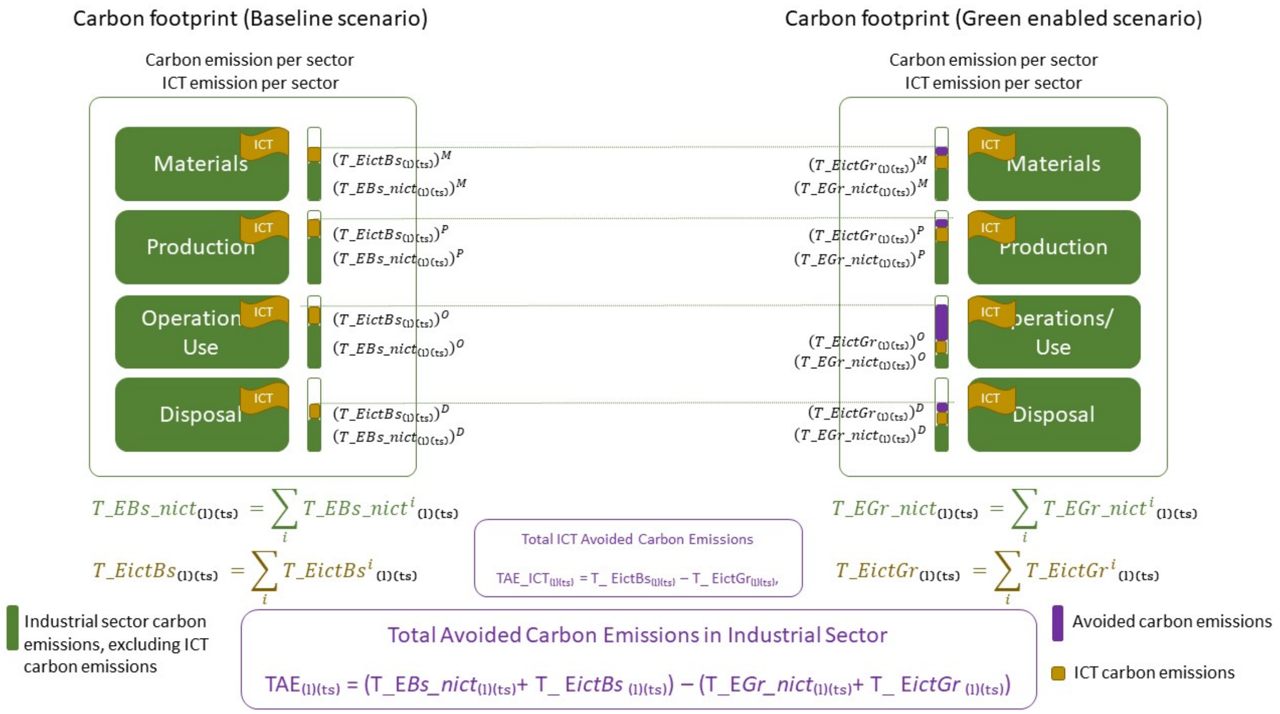


Figure 48: Visualisation of the total avoided carbon emissions, with no circularity support and when ICT is applied as an enabling technology, figure copied from "Alliance for IoT and Edge Computing Innovation 2023".

In order to derive the equation on calculating the avoided carbon emissions in an industrial sector, when ICT is used as an enabling technology, the following assumptions are considered:

- When ICT solutions are used to reduce carbon emissions in Industrial sectors, it is assumed that in the Use/Operation LC phase the carbon emissions are measured under a certain Load and for a certain type of service;
- Load = data processed by the network during a unit of time, e. g., 1 week, 1 month, 1 year; The "l" index is defined as the "percentage of average bandwidth / total bandwidth that ICT infrastructure can handle. If "l=1", it means that the applied Load equals the total bandwidth that ICT infrastructure can handle;
- TS = Type of Service (follow the 5G type of services, e. g., Ultra-Reliable Low Latency Communications (URLLC);
- LCA = Life Cycle Assessment composed by Life Cycle (LC) phases Materials, Production, Use/Operation, Disposal;
- Unit: kgCo2e.

Contributors

Editor:

Ronald Freund, Fraunhofer HHI

Reviewers:

Contributors:

Erwin Schoitsch	AIT Austrian Institute of Technology
Vasileios Karagiannis	AIT Austrian Institute of Technology
George Suciu	BEIA Consult
Anagnostis Paraskevopoulos	Fraunhofer HHI
Behnam Shariati	Fraunhofer HHI
Johannes Fischer	Fraunhofer HHI
Ronald Freund	Fraunhofer HHI
David Hillerkuss	Huawei
Georgios Karagiannis	Huawei
Jun Zhou (James)	Huawei
Liang Zhang	Huawei
Marcus Brunner	Huawei
Zbigniew Kopertowski	Orange
Giacomo Tavola	Politecnico di Milano
Ricardo Vitorino	Ubiwhere
Nikos Giannakakos	UniSystems

Acknowledgements

About AIOTI

AIOTI is the multi-stakeholder platform for stimulating IoT and Edge Computing Innovation in Europe, bringing together small and large companies, academia, policy makers and end-users and representatives of society in an end-to-end approach. We work with partners in a global context. We strive to leverage, share and promote best practices in the IoT and Edge Computing ecosystems, be a one-stop point of information on all relevant aspects of IoT Innovation to its members while proactively addressing key issues and roadblocks for economic growth, acceptance and adoption of IoT and Edge Computing Innovation in society. AIOTI's contribution goes beyond technology and addresses horizontal elements across application domains, such as matchmaking and stimulating cooperation in IoT and Edge Computing ecosystems, creating joint research roadmaps, driving convergence of standards and interoperability and defining policies.