Alliance for AI, IoT and Edge Continuum Innovation

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

Release 4.0

AIOTI WG Standardisation

2026

© AIOTI. All rights reserved.

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 increasingly 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.

© AIOTI. All rights reserved.

Table of Contents

© AIOTI. All rights reserved.

© AIOTI. All rights reserved.

Table of Figures

© AIOTI. All rights reserved.

© AIOTI. All rights reserved.

List of Tables

© AIOTI. All rights reserved.

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

© AIOTI. All rights reserved.

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
Wi-Fi	Wireless Fidelity
XGS-PON	10 Gigabit Symmetrical PON

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

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.

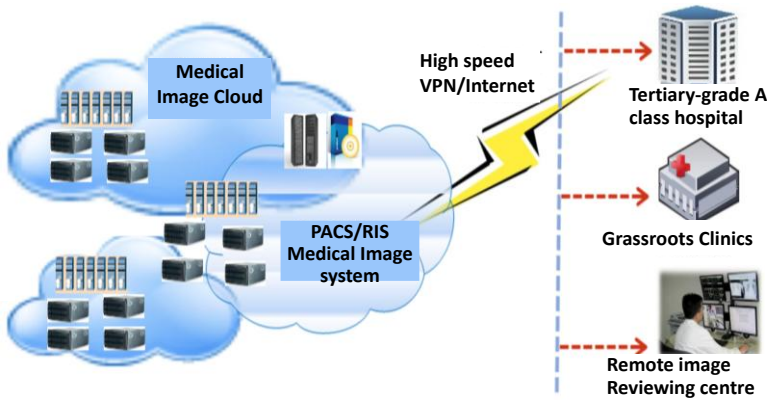


Figure 1: Medical image migration to the cloud.

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.

© AIOTI. All rights reserved.

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.

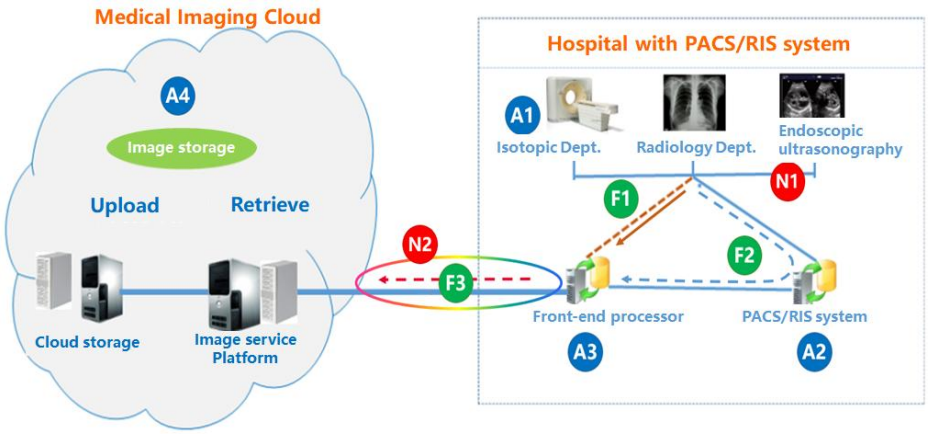


Figure 2: Key components and data flows in the cloud-based medical image migration.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

Alternative Flow

N/A.

Post-conditions

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

High Level Illustration

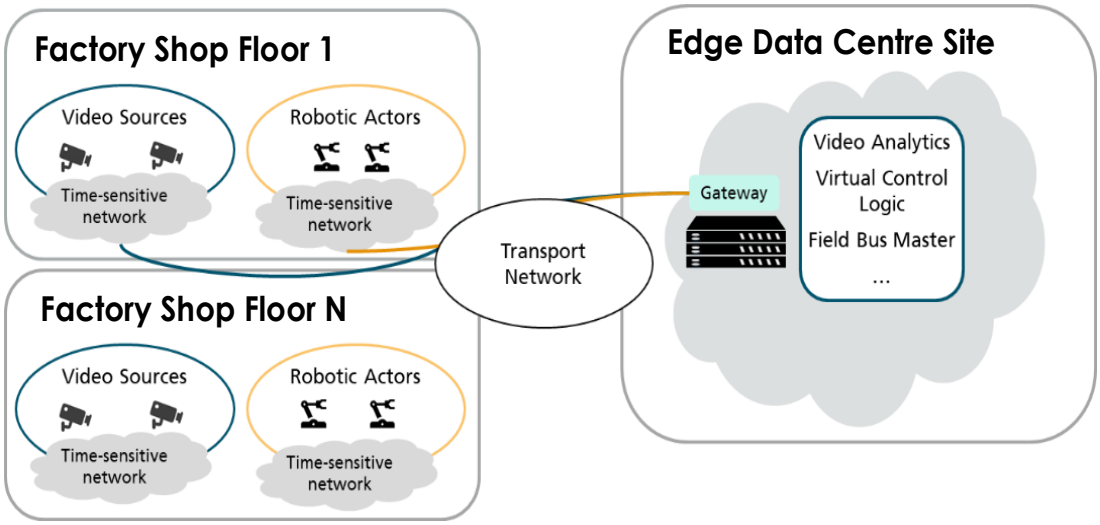


Figure 3: Schematic depiction of the visual inspection in production use case.

Potential Requirements

Functional Requirements

- 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).

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.

© AIOTI. All rights reserved.

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:

© AIOTI. All rights reserved.

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, Wi-Fi, LiFi)
  - Current and upcoming generations of Wi-Fi: Wi-Fi 6 and Wi-Fi 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.

© AIOTI. All rights reserved.

High Level Illustration

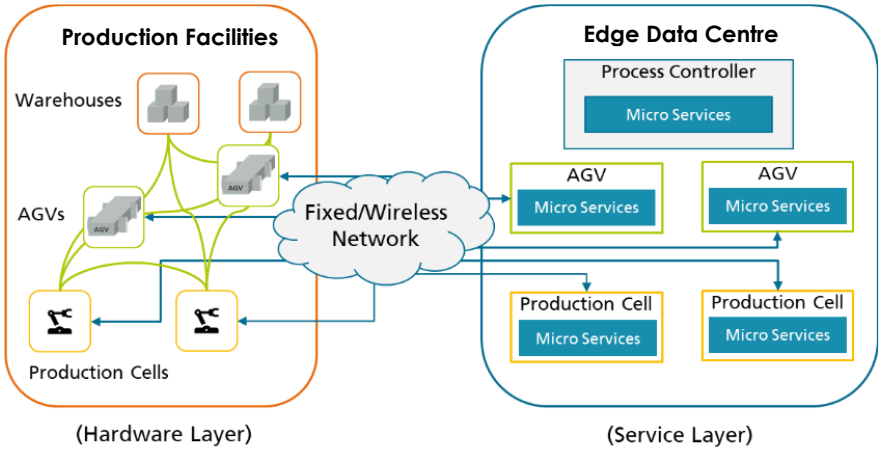


Figure 4: Schematic depiction of the automated guided vehicles use case.

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*

© AIOTI. All rights reserved.

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 Wi-Fi 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

© AIOTI. All rights reserved.

Pre-conditions

- OWC systems can be installed in parallel to existing infrastructure and exchange data with the factory network. System architecture is similar to Wi-Fi 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 Wi-Fi 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 Wi-Fi is accepted/available, a parallel OWC/Wi-Fi 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.

© AIOTI. All rights reserved.

High Level Illustration

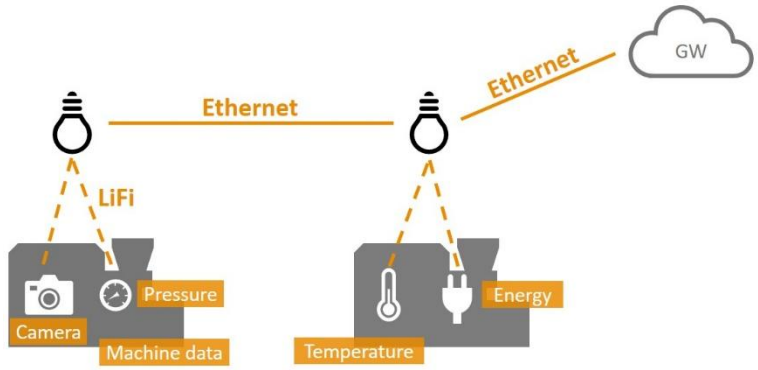


Figure 5: Schematic view of robust and reliable communication via OWC cells in IoT network: Machines are wirelessly connected via the OWC access points with the cable-based Ethernet network.

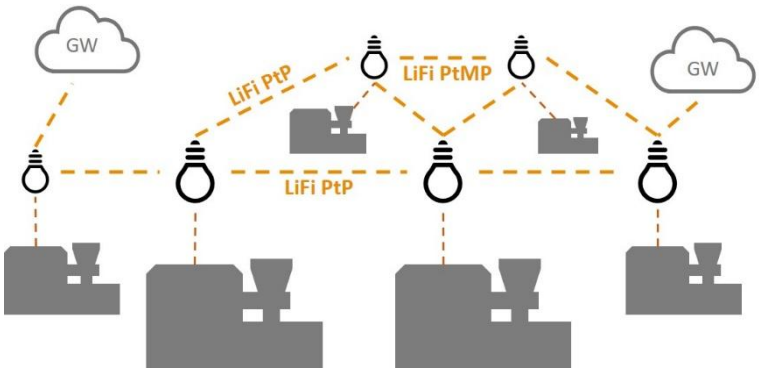


Figure 6: Schematic view of OWC system implementation for a flexible production floor: Production devices are wirelessly connected via point-to-point (P2P) and/or point-to-multipoint (P2MP) connections.

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. Wi-Fi).
- In case of parallel use of OWC with RF-based mobile communication, seamless handover between the two protocols is necessary.

© AIOTI. All rights reserved.

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).

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

Post-conditions

N/A.

High Level Illustration

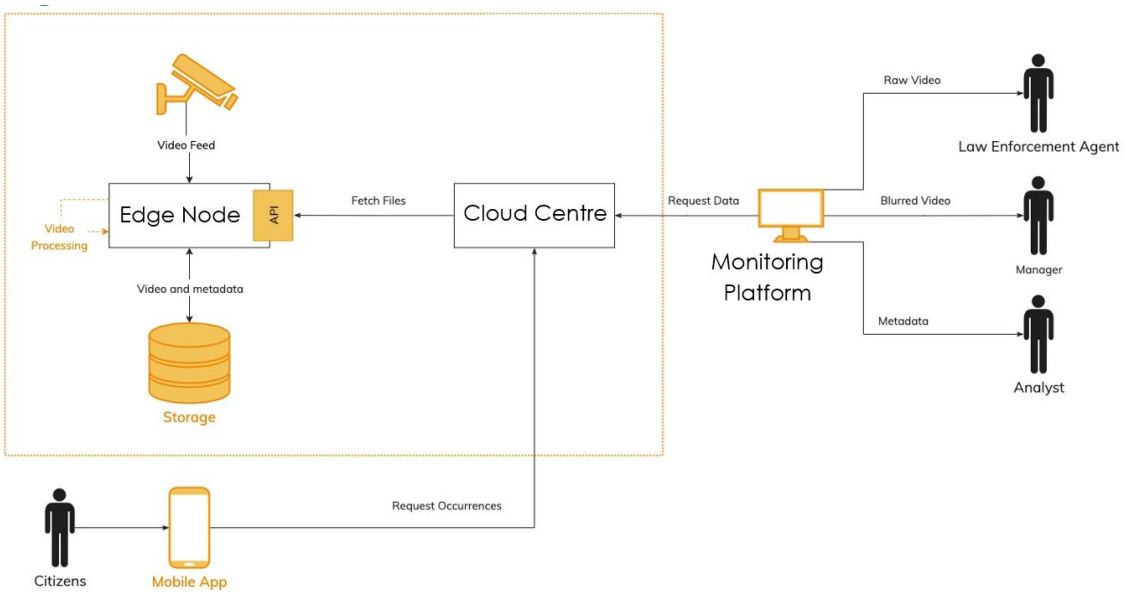


Figure 7: Schematic depiction of the protecting sensitive data within smart cities use case.

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, Wi-Fi). 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).

© AIOTI. All rights reserved.

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 Wi-Fi 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.

© AIOTI. All rights reserved.

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, Wi-Fi 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 Wi-Fi 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 Wi-Fi 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 Wi-Fi. Moreover, a coordination mechanism over fibre and Wi-Fi 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 Wi-Fi 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.

© AIOTI. All rights reserved.

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

- Wi-Fi is provided by one FTTR system, so the MFU can process the coordination function over fibre and Wi-Fi 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 Wi-Fi 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 Wi-Fi 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 Wi-Fi networking.

For example 2: The compute for the service is allocated at various places, which requires a level of coordination between the Wi-Fi 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 Wi-Fi performance.

Alternative Flow

N/A.

Post-conditions

N/A.

© AIOTI. All rights reserved.

High Level Illustration

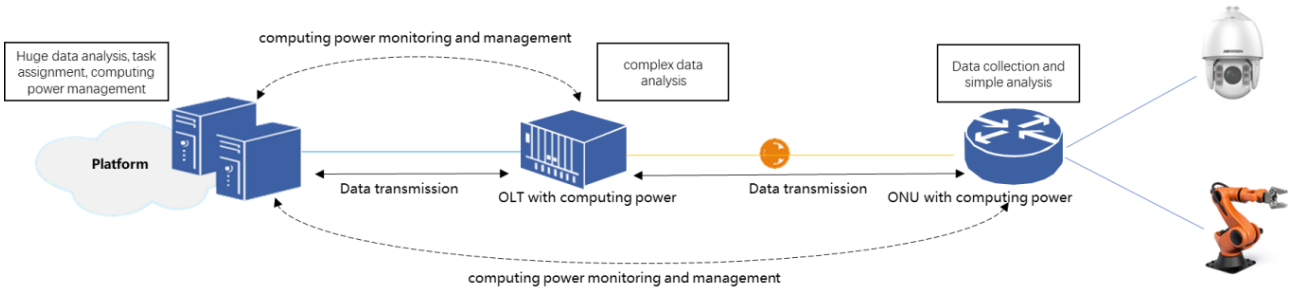


Figure 9: The basic network elements in computing power access networks.

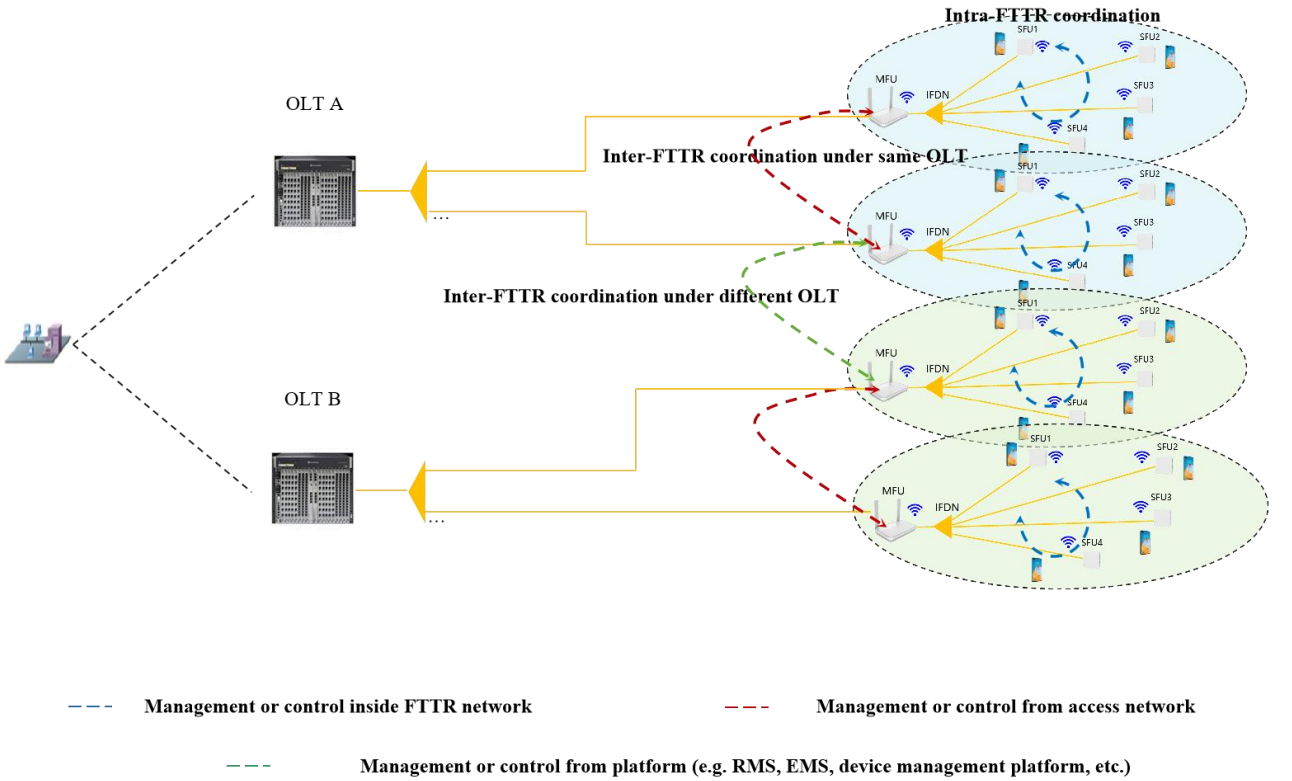


Figure 10: Coordination functions for FTTR across the network (main FTTR unit: MFU, subordinate FTTR unit: SFU, indoor fibre distributed network: IFDN).

Potential Requirements

N/A.

Optical Network specific Requirements

R1: The residential or SME network shall use optical backhaul of the Wi-Fi 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.

© AIOTI. All rights reserved.

2.7 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).

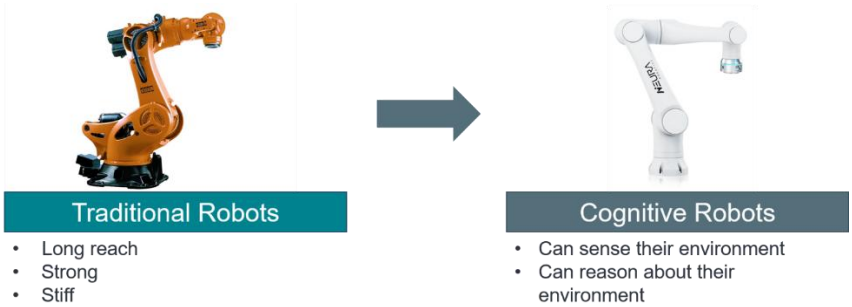


Figure 11: Move to cognitive robots.

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.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

High Level Illustration

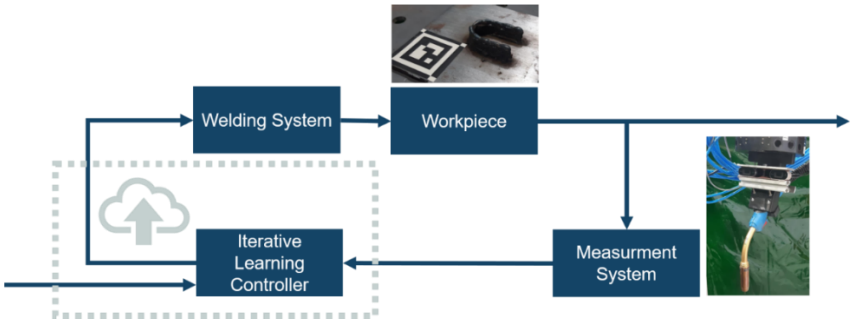


Figure 13: RaaS welding application.

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.

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

2.8 AI-Enabled Cloud Continuum Framework (COGNIT). Smart Mobility use case

Description

Our use case focuses on Connected, Cooperative, and Automated Mobility (CCAM) services within the Cloud-Edge Continuum, aiming to enable safer, more efficient, and intelligent mobility through distributed computing and seamless data sharing among vehicles, edge devices, and cloud.

Specifically, we have implemented the Public Transit Service (PTS) and the Time-to-Green (TTG) service across 114 regulated intersections in the city of Granada (Spain). A total of 38 Mobility-Hub (ACISA's Traffic light Controller) with C-V2X/DSRC capabilities have been installed to manage those intersections.

One public transit line (Line 4) has been equipped with homologated C-V2X/DSRC On-Board Units (OBUs) on all buses. Thanks to the PTS service, these buses can request priority passage at equipped intersections, effectively reducing overall transit time.

Additionally, other vehicles equipped with C-V2X OBUs may receive real-time information about the remaining time until the green light phase, enhancing driving experience.

Public bus priority systems are commonly used in the transport management of cities, but one of the key advantages of using standardized V2X equipment is that it can be reused to provide more than 30 different use cases harmonized for interoperability by the C-Roads Platform[1].

Those intersections in Granada will be gradually updated with additional V2X services using the same infrastructure.

Why this Use Case?

There's a push toward safer and smarter road transport via vehicle automation and connectivity. CCAM is being promoted across Europe (e.g., through EU-funded initiatives) as a solution to reduce accidents, improve traffic flow, and cut emissions.

Traditional centralized systems can't handle the low latency and real-time processing needed for CCAM. This use case leverages the COGNIT cloud-edge continuum framework to:

Distribute computation close to the data source (e.g., in-vehicle or roadside edge nodes).

Seamless resource provisioning by far edge devices through Function as a Service (FaaS) paradigm in the continuum.

Offload intensive tasks to Edge servers or cloud for more complex processing and learning (Intersection Digital Twin simulation functionality).

An AI based Orchestrator provided by the COGNIT framework decides where to provision resources to execute the requested FaaS.

It is foreseeable that many CCAM services will be allocated along 5G network slices[2]. 5G network slicing allows for the creation of multiple virtual networks on a shared physical infrastructure, each tailored to meet specific performance, capacity, and functionality requirements. This capability is particularly beneficial for CCAM services, which demand diverse and stringent network characteristics such as ultra-reliable low latency and, eventually, high bandwidth. As 5G technology matures, network slicing is transitioning from a theoretical concept to practical implementation across various sectors[3] such as manufacturing, healthcare, and emergency services, demonstrating its versatility and effectiveness.

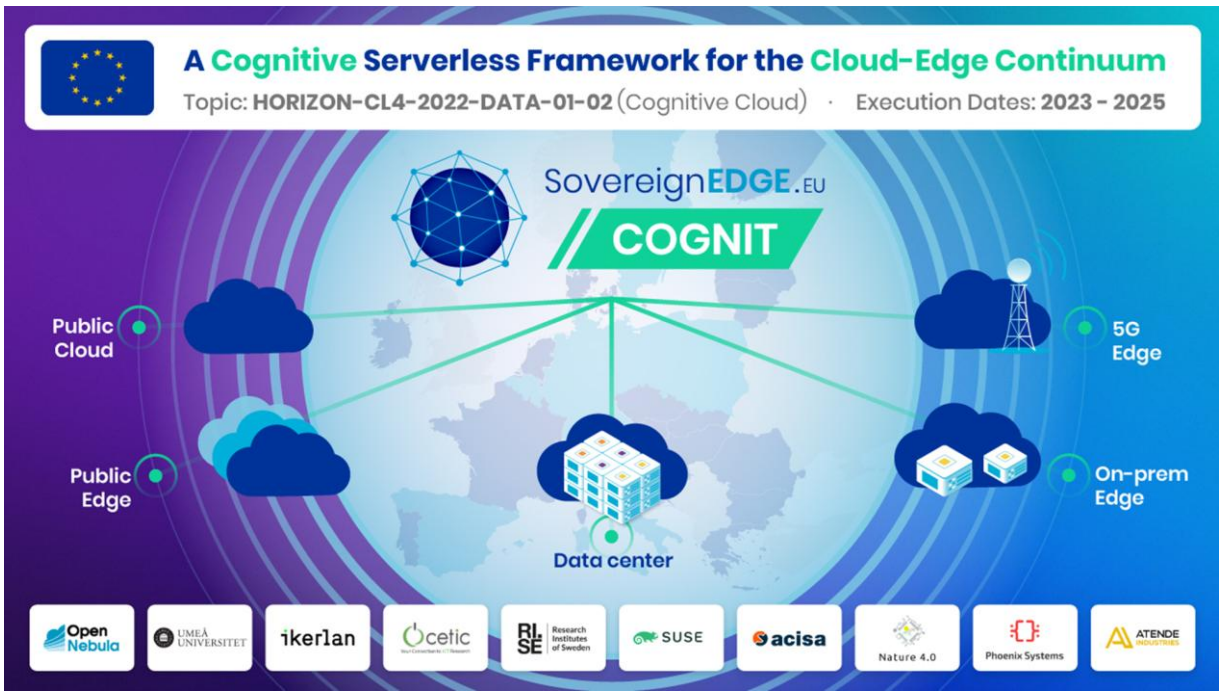
Given these advancements, it is anticipated that CCAM services will increasingly leverage 5G network slicing to ensure the required Quality of Service (QoS). This integration will facilitate the

© AIOTI. All rights reserved.

deployment of intelligent transportation systems, enhance vehicular communication, and support the development of autonomous driving technologies.

Source

Sovereign.COGNIT, H2020 European project (https://cognit.sovereignedge.eu/)



Roles and Actors

The COGNIT project involves a diverse set of roles and actors collaborating to develop an AI-enabled cognitive cloud for Europe's cloud-edge continuum. Below is an overview of these roles, their responsibilities, relationships, and associated actors:

Roles in the Use Cases:

End Users: Individuals or organizations utilizing the applications developed within the COGNIT framework, such as city residents, energy consumers, or entities responsible for wildfire detection.
Vertical Industry Partners: Organizations operating within specific sectors that implement and benefit from the COGNIT framework:
Smart Cities: ACISA coordinates this use case, focusing on urban CCAM services enhanced with continuum resources.
Wildfire Detection: Nature 4.0 leads efforts in environmental monitoring and disaster prevention.
Energy: Phoenix Systems and Atende Industries collaborate on energy management solutions.
Cybersecurity: CETIC and SUSE address security aspects within the cloud-edge continuum.
Communication Network Providers/Operators: Entities supplying the necessary networking infrastructure to support data transmission between edge devices and the cloud.
IoT Device Manufacturers: Companies producing sensors, actuators, and other hardware components deployed in various environments to collect data.
IoT Platform Providers: Organizations offering platforms that facilitate the integration, management, and analysis of data from IoT devices within the COGNIT framework, like ACISA or Phoenix.
Cloud Edge Service Providers: Entities providing cloud computing resources where certain computational tasks are executed as part of the serverless framework, such as OpenNebula.
Research Institutions: Academic and research organizations contributing to the development and validation of the COGNIT framework through scientific research and innovation, like IKERLAN, Ümea University, CETIC or RISE.

© AIOTI. All rights reserved.

Relationships Between Roles:

End Users interact with applications developed by vertical Industry Partners, utilizing data processed through the COGNIT framework.
Vertical Industry Partners collaborate with IoT Device Manufacturers to deploy necessary hardware and with IoT Platform Providers to manage and analyze data.
Communication Network Providers/Operators ensure reliable connectivity between IoT devices, edge nodes, and cloud services, facilitating seamless data flow.
Cloud Service Providers offer computational resources for tasks that cannot be handled at the edge, working in tandem with edge computing resources to optimize performance.
Research Institutions support all stakeholders by providing insights, developing innovative solutions, and validating the framework's effectiveness across different use cases.

Actors and Their Roles:

Open Nebula: Act as a project coordinator and technical lead. The initial deployments of the COGNIT framework leverages Open Nebula IaaS solutions.
Ikerlan: Act as a lead development of a distributed Function-as-a-Service (FaaS) paradigm. This paradigm aims to enable Internet of Things (IoT) and edge devices to offer compute-intensive applications through intelligent task offloading to the cloud-edge continuum
Ümea University, CETIC, RISE: Collaborate as research institutions across various roles to provide scientific expertise, such as AI and cybersecurity, and contribute to framework development, and validate use cases.
ACISA: Acts as the Vertical Industry Partner for the Smart Cities use case, coordinating efforts to integrate the COGNIT framework into urban mobility environments.
Nature 4.0: Serves as the Vertical Industry Partner for the Wildfire Detection use case, focusing on environmental monitoring applications.
Phoenix Systems and Atende Industries: Function as Vertical Industry Partners in the Energy use case, developing smart energy management solutions.
CETIC and SUSE: Operate as Vertical Industry Partners for the Cybersecurity use case, addressing security challenges within the cloud-edge continuum.

Pre-conditions

COGNIT is a RIA project that started at **TRL 2** and aims to achieve **TRL 5** by the project's conclusion. It will reach a higher TRL with the lead of Open Nebula and through further collaboration of other COGNIT partners and the Open-Source community.

Triggers

The trigger is the function sent to the Cognit framework in the form of FaaS. In the case of Smart City use case, the trigger is the request of transit priority by a public bus.

Potential Requirements

Current Smart City use case is deployed using C-V2X communication standard (3GPP PC5 Mode 4), where OBU and RSU also communicate through 4G for monitoring and maintenance services. PC5 interface was released in 3GPP release 14 and it provides direct communication between the vehicle OBU and the infrastructure RSU without the need of any cellular network.

The potential requirements consider the future evolution C-V2X communications within the 3GPP standardization body, NR-V2X, in releases 16+.

© AIOTI. All rights reserved.

Functional Requirements

- Roundtrip latency under 2ms.
- Secured communication through cryptography methods.
- The network must support millions of connected vehicles in cities, that must maintain synchronization, and transparent handover as vehicles move along the city.
- Traffic Ligh Controllers (i.e. M-Hub) should be connected to dedicated optical network or leverage metro dark fiber networks.
- Eventually, V2X services might be allocated in dedicated slices within the 5G network infrastructure.
- CCAM-related services might be classified in different 5G service categories:
- Safety use cases (real-time) → URLLC
- Entertainment/internet in the car → eMBB
- Vehicle sensors, smart traffic systems monitoring and diagnostics → mMTC

Non-Functional Requirements

- Interoperability between car manufacturers, network operators and infrastructure.
- Standard-based communication based on ETSI V2X standards.
- Guaranteed and programmatically configured Quality of Service (QoS)
- Isolation from other network traffic
- Scalable communication between systems interconnects massive volumes of vehicles with the road infrastructure and other vehicles through direct link (PC5 interface) or through the 5G network (Uu interface).
- Reliable communication between vehicles and infrastructure. There are CCAM services related to safety and security (Day 1, Day 1.5).

Normal Flow

In the Smart city use case developed by ACISA, this is a normal flow of exchanged data between key entities:

Vehicles: Equipped with On-Board Units (OBUs), vehicles send and receive data through V2X communication. Public buses may send priority request to M-Hub to reduce transit time at intersections. They may also receive information like traffic light timings, intersection layout, or information.
Infrastructure: Roadside Units (RSUs) and Traffic Light Controllers (M-Hubs) are part of the infrastructure. They communicate with vehicles, providing information about traffic conditions, signal timings, and receiving requests (e.g., for traffic signal priority). M-Hubs also exchange data with the COGNIT framework through FaaS paradigm, and Saturno, ACISA's Smart Mobility Suite.
IoT Platform: The IoT platform (like ACISA's Saturno) collects and manages data from various sources, including M-Hubs, other traffic subsystems or external information systems. It may also provide data and services to other entities, like providing historical and real-time traffic data to Digital Twins.
COGNIT Framework: This framework manages the cloud-edge continuum, handling the deployment and orchestration of applications and services. It receives function calls and data from entities like Saturno, and provides resources and services in return, such as the orchestration of resources to execute the computation demanded in FaaS. In our use case, it manages the execution of the Digital Twin traffic simulations to assist decision taking.
Digital Twins: Acisa implements a distributed Digital Twins of the intersections to better evaluate its status and evolution. It receives data from various sources, including vehicular data, traffic information, and environmental data, to create a digital representation of the intersection. Based on situational awareness might request timing phase adjustments to the Traffic Light Controller (M-Hub).

© AIOTI. All rights reserved.

Figure 14: ACISA's next-generation traffic light controller integrates edge computing capabilities to support the deployment of containerized applications (i.e., V2X services, Video analytics, etc).

Each time a priority request reaches the M-Hub, an evaluation is performed with the aim of its Digital Twin to determine whether it is appropriate to grant priority or not. Priority will be granted if the process does not significantly disrupt the intersection flow and if it improves the bus's transit time according to simulations results (see Figure below).

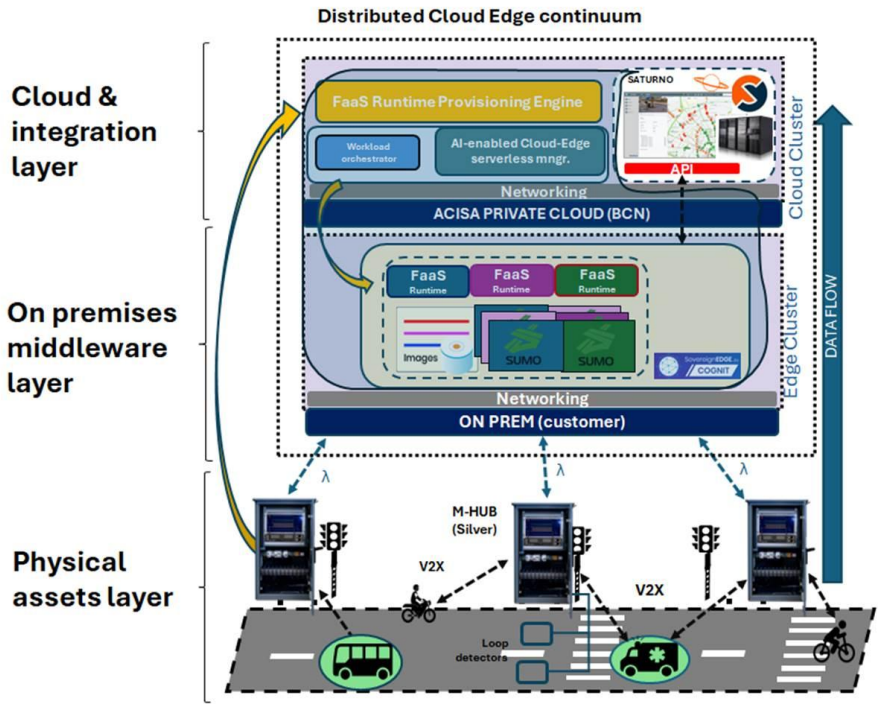


Figure 15: COGNIT Smart City use case. Priority request flow representation.

[1] https://www.c-roads.eu/platform.html
[2] https://blog.sasken.com/5g-network-slicing
[3] https://elpais.com/proyecto-tendencias/2025-01-31/network-slicing-de-concepto-teorico-a-una-realidad-gracias-a-la-red-5g.html

© AIOTI. All rights reserved.

2.9 Generative AI for F5G-A Network Management Tasks

Description

Modern and future fixed optical networks like F5G-Advanced increasingly require automation and intelligent management due to growing complexity and scalability needs driven by emerging technologies like 6G. Manual or semi-automated network management methods struggle to efficiently handle dynamic demands, especially given the increasing diversity of network elements and vendors. Generative AI techniques, specifically Large Language Models (LLMs), can play a pivotal role by translating human intent into machine-readable configurations, automating routine management tasks, and streamlining the monitoring of network conditions. This use case focuses on leveraging generative AI capabilities to automate configuration, monitoring, and management tasks across fixed and optical network infrastructures. It highlights the application of general-purpose, on-premise hosted LLM agents ensuring compliance with data privacy regulations, such as Non-Disclosure Agreements (NDAs).

The proposed generative AI-assisted solution introduces an LLM-assisted network copilot capable of interpreting high-level operator intents, automatically generating API client code aligned with standardized network data models, and performing real-time monitoring and interpretation of network device responses. This solution can be implemented using a multi-agent architecture, utilizing general purpose LLMs optimized with dynamic context retrieval and advanced prompt-engineering strategies such as few-shot, In-Context Learning (ICL), and Chain-of-Thought (CoT) prompting.

This use case is applicable across various scenarios of optical networks, including optical transport networks (OTN), Optical Access Networks such as Passive Optical Networks (PON), and IP/Ethernet-based fixed networks.

The key functions such a generative AI-assisted automation solution could include:

Intent Classifier: Uses pre-trained lightweight LLM to categorize operator requests into specific tasks (e.g., configuration, monitoring, provisioning).

Code Generator: Automatically generates standardized API client code using LLMs guided by system-provided standardized XML/YANG examples.

Code Validator: Validates generated configuration code for correctness and compliance with standardized data models.

Digital Twin Sandbox: Performs realistic simulation of network device behaviour, ensuring configurations are error-free before actual deployment.

Network Response Interpreter: Uses generative AI to parse and interpret real-time responses from network elements, translating them into actionable insights for network operators.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

The actors in this use case are the network operators who want to automate their network management tasks.

© AIOTI. All rights reserved.

Pre-conditions

- Network elements need to be programmable
- General-purpose LLMs need to be deployed on-premise to comply with data privacy regulations
- Content repository that contains validated configuration examples, standardized network data models etc.
- Adequate computational resources available
- Simulated sandbox environment available

Triggers

The operational workflow begins with a network operator submitting a natural-language request that specifies a desired network task, such as registering new ONUs in the OLT, certain service provisioning on the OLT and ONU side, retrieving performance metrics with telemetry or configuring device parameters, etc.

Normal Flow

Upon reception of the natural language request, it undergoes initial processing by an Intent Classifier Agent, which interprets and categorizes the intent, identifying relevant network devices and required operations. Following classification, the request is relayed to the Generator Agent, which dynamically retrieves standardized configuration examples, in example YANG/XML templates, and vendor-specific parameters from an on-premise content database. Leveraging these inputs, the Generator Agent automatically constructs standardized, vendor-independent API client code, such as NETCONF XML RPC commands, ensuring adherence to established data models and standards. Once generated, the configuration code proceeds to the Validator Agent, which meticulously validates the structure and content against the relevant standardized data models and schemas. This ensures accuracy and compliance with network standards and avoids potential misconfigurations. Upon successful validation, the proposed configurations are forwarded to a digital twin sandbox—an isolated, simulated environment that precisely mimics the target network devices and their interfaces. This sandbox provides a risk-free verification step, where configurations are tested for operational viability without affecting live network equipment. If any discrepancies or errors arise during sandbox testing, the feedback loop triggers automatic corrections, returning to the Generator and Validator Agents for revisions until the configuration passes validation successfully. After sandbox validation confirms the correctness and reliability of the generated configurations, the approved commands are deployed by the API Client to the actual optical network elements. Post-deployment, real-time responses from network devices – such as confirmations, status updates, or performance metrics – are gathered and forwarded to the Interpreter Agent. This agent uses generative AI capabilities to translate complex XML or protocol responses into concise, human-readable insights, effectively communicating the results back to the network operator. This detailed, modular workflow ensures end-to-end automation, significantly reducing manual effort, enhancing operational agility, and streamlining the interaction with complex multi-vendor network infrastructures. Moreover, by executing all operations on-premise, the system maintains strict compliance with data privacy and confidentiality regulations, enabling operators to securely automate network management at scale.

Alternative Flow

N/A.

Post-conditions

There are no post-conditions.

© AIOTI. All rights reserved.

High Level Illustration

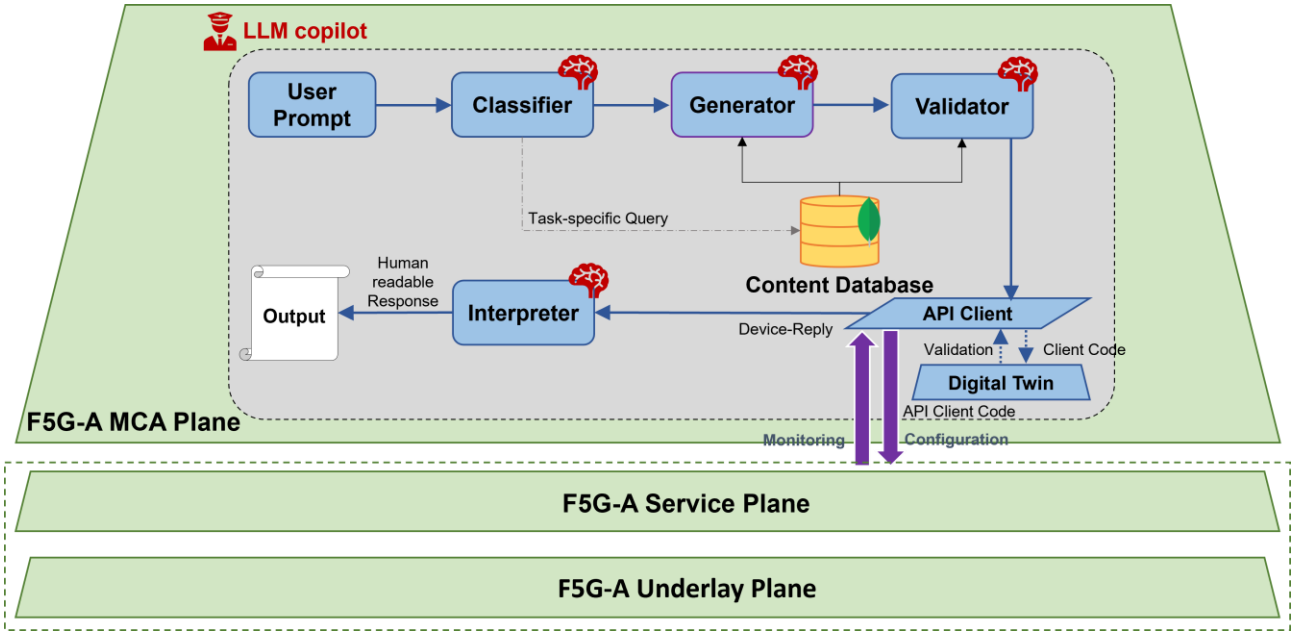


Figure 16: Envisioned Architecture of the LLM-assisted Networking Copilot in the context of F5G-Advanced Network architecture.

Potential Requirements

To effectively implement this use case, operators need to ensure that several key requirements are satisfied:

Programmable Network Elements: Network devices are programmable and support standardized protocols enabling automated and programmable operations.

On-premise Deployment of General-purpose LLMs: To comply with data privacy regulations and NDA requirements, LLMs are deployable within an on-premises environment, eliminating reliance on cloud-hosted services and ensuring sensitive configuration data remains confidential.

Content Repository (Content Database): A comprehensive repository containing validated configuration examples, standardized network data models (e.g., YANG models), and task-specific instructional templates are also available. This repository requires to be continuously maintained, updated, and indexed effectively to ensure rapid and dynamic retrieval of relevant data by generative AI agents.

Computational Infrastructure: Operators need to provision adequate computational resources – whether through local data centers, edge computing platforms, or hybrid cloud environments – to effectively host multiple generative AI agents and digital twin sandbox simulations. Infrastructure will account in this case for AI processing demands, real-time responsiveness requirements, and scalability to accommodate future growth.

Digital Twin Sandbox Environment: A simulated environment replicating real network devices and their interfaces (using standardized data models) will be established to safely test and validate AI-generated configurations before deployment. This significantly reduces the risk of errors impacting live network operation and performance.

© AIOTI. All rights reserved.

2.10 Distributed Intelligence with Privacy-Preserving Features for FTTR

Description

The Fiber-to-the-Room (FTTR) technology represents the next step in providing ultra-high-speed connectivity within residential and enterprise environments by delivering optical fiber links directly to individual rooms. Traditional centralized network architectures struggle to meet the growing demands for adaptive Quality of Service (QoS), energy efficiency, and real-time network intelligence. The integration of Distributed Intelligence (DI) into FTTR systems seeks to address these challenges and limitations by embedding computational capabilities across various network elements, thereby decentralizing network intelligence across the entire optical access domain. This use case focuses on a DI-enabled FTTR architecture that features intelligent functionalities distributed across the Sub FTTR Units (SFU), Main FTTR Unit (MFU), Optical Line Terminal (OLT) and the Edge Cloud, and includes the optimization of power consumption and QoS on the subscriber side, driving better service quality, privacy preservation, reduced energy costs, and higher customer satisfaction.

This use case details the use of distributed intelligence for optimizing QoS and power consumption in FTTR systems while adhering to General Data Protection Regulation (GDPR) compliance. Specifically, intelligence is embedded in the SFU and MFU assuming that computing capability is supported, enabling local monitoring and processing. The SFU interacts directly with user devices through integrated Wi-Fi and/or Ethernet ports, and is responsible for local QoS estimation, optimization and power consumption monitoring. The MFU aggregates data from multiple SFUs and communicates with the OLT and Edge Cloud. GDPR compliance is ensured by processing subscriber-related data locally at the SFU and MFU levels, without sending sensitive information to the operator's infrastructure (i.e., OLT and Edge Cloud).

The objective of this use-case is to design and implement a distributed intelligence system aligned with the FTTR architecture, supporting the three key phases of closed-loop operation: observation, analysis, and action. The DI system aims to enhance monitoring and operational efficiency within FTTR solutions, contributing to End-to-End (E2E) management that ensures a consistently high-quality user experience through collaborative computing across the FTTR components.

The implementation of distributed intelligence in the FTTR scenario is highly dependent on the computing capabilities of each FTTR component, including the SFUs, the MFU, and the OLT. Depending on the equipment vendor, these network elements can need to possess some limited, or more advanced computing capabilities, and support different protocols, APIs, and data formats. In case of limited or non-existent computing capabilities, external computing nodes can be connected to these FTTR components using their native GE interfaces, such that the DI software components can interact with the modules' APIs.

The DI service functions and capabilities are defined through the DI pipeline components' functionalities and can be generally distinguished between the client-side DI and the operator-side DI. The main components of the client-side DI pipeline for MFU and SFU control include:

HTTP-based Telemetry Agent implementing the "observation phase" of the closed-loop operation, and which interacts with the two types of ONU modules using HTTP-requests, querying traffic-, QoS- and power consumption-related telemetry data, and converting it into popular encoding formats, such as e.g., JSON, for subsequent writing into the Data Lake. The agent also acts as a Data Lake client, converting the format of telemetry messages into Data Lake-native formats;

Data Lake represents a data storage instance, such as a Time-Series Database (TSDB), in which the different types of telemetry data are sorted accordingly and stored into dedicated measurements/fields;

© AIOTI. All rights reserved.

Analytics Module implements the "analysis functionality" of the closed-loop operation model, performing the different telemetry data processing tasks, including the analysis of current QoS, traffic and power consumption metrics, and takes decisions such as when the ONU is in idle operation and can be switched into sleep mode, or vice-versa – activated;

HTTP-based ONU Controller employs the "action step" of the closed-loop operation, getting the decisions from the analytics module, and interacting with the ONU modules through HTTP-queries to perform the corresponding hardware configuration changes on MFU and SFUs.

Visualization Dashboard based on open-source software used to assess the performance of the DI pipeline operation.

When it comes to the operator-side DI pipeline integration for OLT monitoring and control, it can be deployed on the Edge Cloud infrastructure, if the OLT possesses limited or no computing capabilities, or integrated into the OLT directly, if on-device computing power is available. In this case, both components (OLT and Edge Cloud) are managed by the operator, and the DI pipeline includes similar pipeline modules as for MFU/SFU. The main difference lies within the distinct Telemetry Agent, which can use other types of APIs and data encoding formats. A popular example is a gRPC-based Telemetry Agent, which subscribes to the OLT streaming telemetry server, and receives the Protocol Buffers (Protobuf) binary-encoded telemetry data, delivered in real-time using UDP or other transport protocols. The agent performs the telemetry data format decoding and conversions, for its subsequent storage into the Data Lake.

It is worth noting that the OLT gets aggregated QoS-, traffic- or power consumption-related telemetry data, which refers to the entire FTTR domain, which can be further used for real-time monitoring and analytics.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

The actors in this use case are the network operators who want to shift toward an intelligent and energy-aware FTTR solution. The operators can position their network offerings as eco-friendly and energy-efficient, appealing to a growing base of environmentally conscious consumers and enterprise clients.

Pre-conditions

- All DI pipeline components should be able to run inside dedicated Docker containers, making the entire solution cloud-native and easily deployable.

Triggers

The network operator sets up a closed loop operation that performs the monitoring of network components, analyses the current status and then takes actions.

Normal Flow

The DI pipeline can run either on the client-side or the operator-side. For the client side, the operation of the DI pipeline on the SFU/MFU side consists of five main steps shown in Figure 17: **Distributed Intelligence pipeline deployed on the SFU and MFU ONUs for DI functionality.**:

- QoS-, traffic intensity and power consumption-related telemetry data acquisition from the ONU modules through HTTP-requests, with a frequency chosen heuristically, depending on the refreshing frequency of the HTTP server running on the ONU.

© AIOTI. All rights reserved.

- Telemetry data conversion from plain text strings, into JSON structured messages, and subsequent translation to Data Lake-native formats for writing and storage into dedicated measurements/fields of the database. A TSDB is a good, widely deployed example of such a Data Lake instance.

- Querying of telemetry data from the Data Lake by the Analytics Module, which analyses the current traffic intensity on the user side, optical signal quality, and power consumption, and depending on the user activity, decides on the most appropriate ONU's Wi-Fi operation mode (e.g., sleep/, standard/, or strong/through-wall). The module is also capable to raise an alarm if the optical signal quality is considered to have degraded.

- The results of the analytics module are further communicated to the ONU Controller, with the outputs also being written back to the Data Lake for storage and dashboard visualization purposes. The ONU Controller subsequently creates HTTP-queries to be sent to the ONUs for reconfiguration, if required.

- The HTTP-request is sent for MFU/SFU module reconfiguration through dedicated TCP sessions.

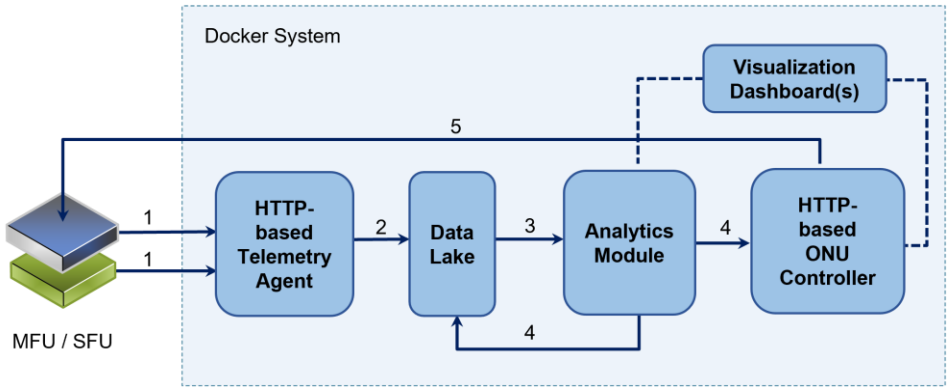


Figure 17: Distributed Intelligence pipeline deployed on the SFU and MFU ONUs for DI functionality.

Alternative Flow

The DI operation on the operator-side consists of a few distinct steps (Figure 18), which are deployed on the edge cloud:

- QoS-, traffic intensity and power consumption-related telemetry metrics are acquired through binary, gRPC/Protobuf streaming telemetry, where relevant telemetry data describing these metrics are collected from the corresponding OLT sensors. The binary encoded Protobuf messages are further decoded and converted into JSON format for readability and structure.

- In the second step, the gRPC-based Telemetry Agent translates the JSON telemetry messages into Data Lake-native formats, writing each corresponding metric into a dedicated measurement/field, similarly to the previous workflow.

- The Analytics Module further queries the FTTR-related telemetry data, where similar analytical tasks, such as optical signal quality assessment, traffic classification (e.g., depending on the running applications), or overall power consumption measurements, are executed.

- The outputs of the Analytics Module are further stored in the Data Lake, with a subsequent display on Visualization dashboards.

© AIOTI. All rights reserved.

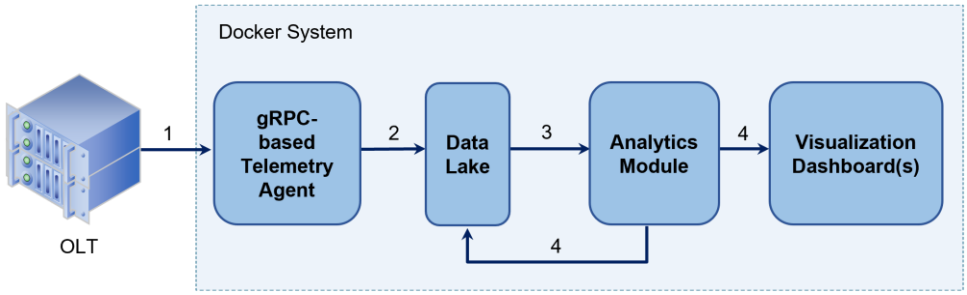


Figure 18: Distributed Intelligence pipeline deployed on the Edge Cloud for OLT Monitoring.

Post-conditions

There are no post-conditions.

High Level Illustration

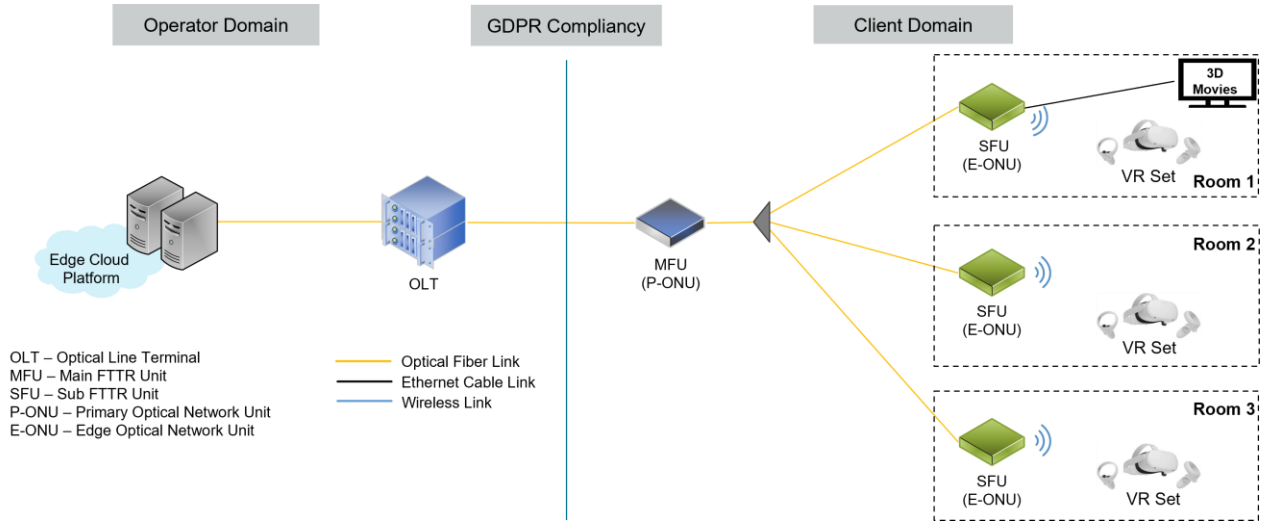


Figure 19: Distributed intelligence in an FTTR system with DI functions carried out on the Edge Cloud (operator domain), and on the MFU and SFU nodes featuring computing capabilities (client domain) of the FTTR.

Potential Requirements

- GDPR compliance must be met by not sending the sensitive information to the operator's infrastructure (i.e., OLT and Edge Cloud).
- Power consumption and traffic need to be monitored from the network components in real time.
- It should be possible to dynamically adjust the power setting of network components.
- The network components need to provide interfaces for monitoring and control.
- The network components need to possess computing capabilities, or it should be possible to connect them to external computing nodes.

© AIOTI. All rights reserved.

2.11 Smart Sensor Cloud for AI in Industrial Manufacturing

Description

Traditional industrial systems have heavily relied on edge computing for the processing of sensor data and the execution of AI models. This paradigm has necessitated localized computational power, often resulting in the deployment of small server racks near production systems. However, with the rapid growth of artificial intelligence (AI) models in terms of complexity and size, maintaining such distributed compute infrastructure has become increasingly inefficient and costly.

The Smart Sensor Cloud presents a transformative approach by leveraging fifth-generation fixed networks (F5G and F5G Advanced) and advanced optical fibre technologies to eliminate the need for edge devices running AI models. Instead, all sensor data—including high-resolution visual feeds from industrial cameras, environmental metrics, and other inputs—are streamed directly to the cloud for processing. This centralized model shifts computational demands away from edge devices to the cloud, enabling more efficient use of advanced AI-optimized hardware and centralized resources. Note that centralized might either refer to the manufacturing company's data centre optimized for AI workloads or external data centres in an industrial AI as a Service model.

This approach is motivated by the increasing size of AI models and their growing demand for computational resources. Centralizing the compute infrastructure reduces the overhead of maintaining and upgrading edge devices. The Smart Sensor Cloud utilizes F5G-A's high bandwidth and low-latency capabilities, enabling real-time processing, decision-making, and control from a centralized location.

This distributed approach often leads to scalability challenges and inefficiencies, particularly in systems where AI models need frequent updates or retraining. With the Smart Sensor Cloud, data can be streamed directly to the cloud, where advanced AI algorithms process it and generate actionable insights (see Figure 20).

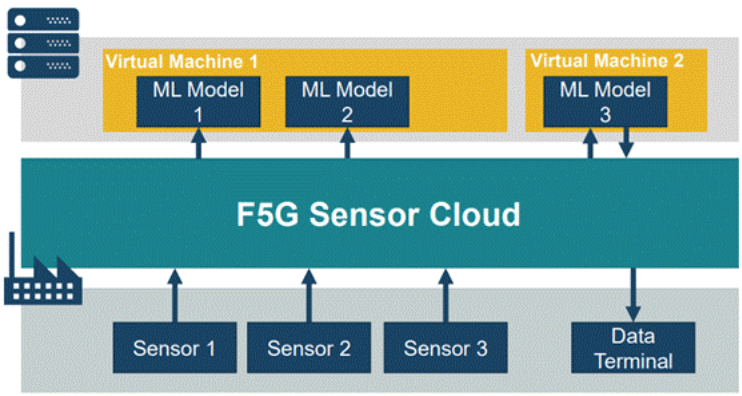


Figure 20: Smart Sensor Cloud Overview (based on F5G Use Case).

By utilizing F5G-A networks, edge devices are eliminated through the Smart Sensor Cloud, which addresses several critical challenges. The centralized model allows advanced AI algorithms to process data in real time, ensuring uniformity across systems and reducing hardware costs. For example, a production line monitoring system can now rely on high-resolution cameras that stream real-time video feeds directly to the cloud. Advanced AI models analyse this data to detect defects, optimize operations, or predict equipment failures, all without requiring localized compute resources. This approach also facilitates rapid scaling of AI capabilities across multiple facilities, unifying data analytics while enhancing efficiency and reliability. The architecture of the Smart Sensor Cloud consists of several key components, as illustrated in the proposed solution diagrams. At its core is the F5G Sensor Cloud, which integrates a Sensor Lake for storing unstructured sensor data and an Asset Administration Shell to ensure interoperability.

© AIOTI. All rights reserved.

The Sensor Lake functions similarly to a data lake, aggregating and structuring data from various sensors. This structured data is then processed by machine learning (ML) models running on virtual machines in the cloud, enabling advanced analytics and decision-making. The system's Network Configurator optimizes latency and bandwidth for each sensor, ensuring seamless and efficient communication between the factory floor and the cloud. By enabling sensors to communicate with multiple systems simultaneously, the F5G-A Sensor Cloud ensures that each sensor receives the required network performance for its specific tasks.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

N/A

Pre-conditions

Network Infrastructure Without Edge Computing Devices:

- The communication network directly transmits all raw and processed data from sensors, machines, and systems to the on-premises cloud in real time.
- A fibre-optic backbone capable of supporting between 10 Gbit/s and 10 Tbit/s depending on the size and sensing equipment of the factory is essential, with high-bandwidth last-mile connectivity to each sensor or device.
- Sub-millisecond latency across the entire network must be ensured to handle time-sensitive operations seamlessly.
- Redundancy and fault tolerance are critical to prevent downtime, requiring redundant fibre paths and network devices with rapid failover capabilities.

Triggers

N/A.

Normal Flow

N/A.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

N/A.

© AIOTI. All rights reserved.

Potential Requirements

- **Direct Data Management in the On-premises Cloud:**
  - All data streams need to flow directly into the on-premises cloud for processing, storage, and analysis, with sufficient ingestion bandwidth to handle real-time streams from potentially thousands of endpoints.
  - Real-time data handling capabilities are essential to ensure immediate processing, filtering, and decision-making without delay.
  - The centralized infrastructure needs to scale dynamically to meet increasing data volumes through horizontal scaling for compute and storage nodes.
- **Security:**
  - Sensitive data needs to be encrypted from source to destination to ensure confidentiality without intermediate filtering.
  - Secure access to the network needs to be maintained with authentication protocols for all connected devices.
  - Anomaly detection and intrusion prevention systems are essential to safeguard the network and data integrity.
- **Centralized Redundancy and Reliability:**
  - The centralized cloud needs to have built-in redundancy, including active-active clusters and fast recovery storage technologies, to compensate for the lack of edge devices.
  - Load balancing mechanisms need to dynamically manage traffic and ensure scalability.
- **Compliance Without Edge Devices:**
  - All raw data need to be handled in compliance with relevant regulations, with direct tagging and classification at the source for compliance tracking.
  - Auditability needs to be ensured with centralized log management systems to track every data packet from origin to cloud.
- **Time-sensitive Applications:**
  - Critical control systems need to operate seamlessly without edge devices, requiring deterministic performance and traffic prioritization for time-sensitive data.
- **Performance Monitoring and Optimization:**
  - Real-time network health monitoring is crucial to predict and resolve bottlenecks.
  - Continuous tracking of bandwidth utilization and storage capacity ensures sufficient resources are available for operations.

© AIOTI. All rights reserved.

2.12 Integrated NAS over FTTR

Description

On-premises network attached storage (NAS) enable various applications for end users including data backup & classification, photo sharing, smart search, etc. Traditionally, the NAS device is separated from the network, acting as an independent and accessible network device. Therefore, it suffers from a number of issues including poor reliability, unstable remote access, difficult to manage for home users.

To address the issues mentioned above, the integrated NAS in FTTR devices (e.g. in the MFU) will make use of the network functions and resources of the FTTR network. An FTTR-based solution constructs a distributed storage network, ensuring data storage reliability via both system level and network level. The FTTR solution, in general, is offered by the F5G-A network service provider, connecting to the F5G-A access network. Thus, E2E communication is guaranteed since it is delivered by a single managed entity and the F5G-A network service provider is responsible for the reliable operation.

To overcome the drawbacks of the traditional independent NAS device, the NAS device is integrated with the FTTR devices, i.e. MFU or SFU, shown in Figure 21. Such a case will eliminate the interfaces between FTTR device and NAS device and unify the management methodology. The data from/to integrated NAS can be identified by the FTTR network, ensuring data transmission priority and protection (such as encryption). The coordination with cloud storage as redundant backups provides additional stability and reliability.

Data transmission priority is reflected by assigned bandwidth and transmission opportunities, resulting in higher network performance to support data transmission for NAS services. For example, as to local video unicasting or video recording, the system enables video tap-to-use.

To facilitate better implementing AI functions for smart home, the AI needs to collect real-time information of the network, sensor and also make use of the historical data, which is stored in NAS. The seamless coordination between network, storage and computing power is important for AI deployment. The local NAS in FTTR can be scaled with cloud storage for extension. The frequent use data can be stored in local while the non-frequent use data can be pushed to cloud. The extension then also be ensured to provide sufficient storage space according to the needs of the end users.

The service provider also intends to use a single protocol to manage different devices. Integrated NAS will be as a module of FTTR so that it can be easily managed for remote visualization and management. On the other side, the network management system is able to quickly identify faults and notify end users of quality risks in advance.

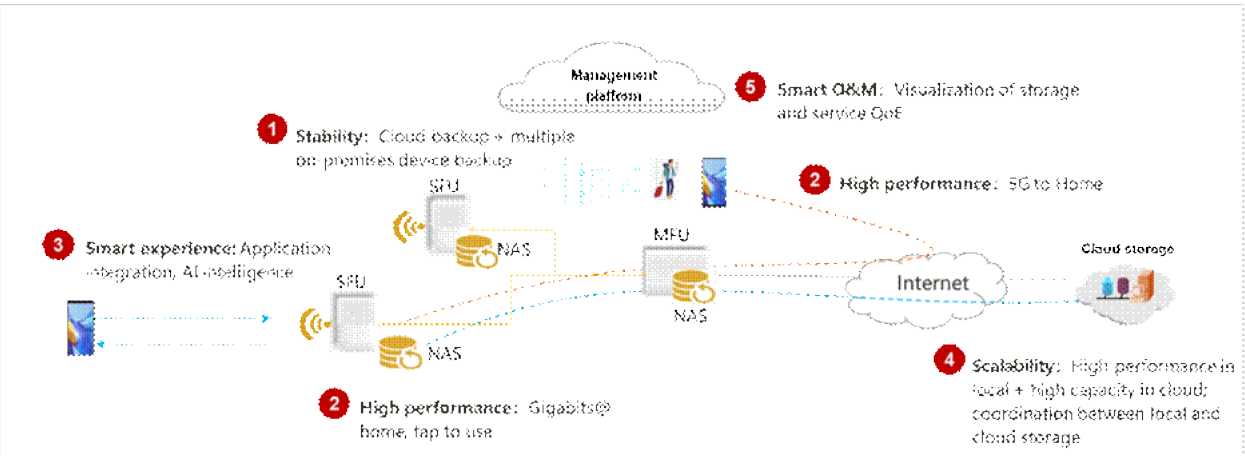


Figure 21: The scenario of integrated NAS over FTTR.

© AIOTI. All rights reserved.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

N/A

Pre-conditions

N/A.

Triggers

N/A.

Normal Flow

N/A.

Alternative Flow

N/A.

Post-conditions

N/A.

Potential Requirements

The integrated NAS over FTTR requires the FTTR network to provide the following functions:

- **Hardware reliability:**
  - Component reliability:
  - The system should support fault diagnostic and repair of storage component. Controller in the system should monitor working temperature of component and provide necessary protection and warning indication.
  - Memory reliability:
  - The system should guarantee the memory working after power-up. The system should support system-level fault tolerance, i.e. system-level fault recovery and fault isolation.
  - System reliability:
  - The system should support background inspection, bad block isolation, online self-repair, DIE failure handling, and power-off protection.

- **Smart applications based on NAS functions:**
  - Mobile apps: support simple and smooth operation process, flexible and intelligent photo albums, and convenient and intelligent search
  - Component-based storage module: with small size and no connection is required, therefore, avoiding the risks of large size, difficult deployment, and reliability of connection.
  - Remote access: no relayed server is required, implementing a dedicated fast channel for P2P VPN.
  - High-speed and reliable backup: local high-speed backup + cloud disk dual backup, achieving both high speed and reliability.
  - Intelligent search: the AI computing engine provides personalized search results based on user search content and behaviour, improving search efficiency and accuracy.

© AIOTI. All rights reserved.

- **High performance of FTTR network: Integrated NAS over FTTR is a new device, reusing FTTR** hardware and intelligent capabilities. The performance requirements in local access, remote access and cloud access are as following:

  - Local access:
    - High throughput: Without affecting network services, FTTR system should provide a throughput larger than 1000 Mbit/s to ensure that end user can quickly upload and download large files, such as HD photos and videos, when accessing the NAS device. This means that end user can enjoy an almost instantaneous data transfer experience without waiting for a long upload or download process (1 GB video upload < 10 seconds).
    - Low latency: FTTR system should provide a latency of less than 15 ms so that end user can hardly experience any latency during operations, improving real-time interaction experience, such as online photo or video editing.
  - Remote access:
    - High throughput: When end users are not at home, FTTR system provides a data access capability of large throughput. FTTR system helps end user quickly upload content from mobile phones to NAS systems or download content from NAS systems to mobile phones, improving work efficiency and data availability.
  - Cloud access:
    - High throughput: the system can use intelligent traffic steering to build dedicated acceleration channels with edge/central cloud disks, enabling end user to quickly back up content from home storage systems to cloud disks or restore data from cloud disks to FTTR NAS system. The data synchronization time is greatly shortened.

- **Scalability:**

The integrated NAS system needs to support scalability functions, including cloud scalability (e.g. support NAS device connected to the cloud disks of service provider and third-party web disks), local scalability (e.g. support flexible pluggable storage units) and new node expansion (e.g. support a distributed storage system based on the distributed architecture components of MFU and SFU).

- **Smart O&M:**

In addition to remote management of FTTR network services, service provider should also manage end user's NAS services. Pre-warning and proactive services are provided to prevent data loss caused by hardware aging and expiration, protecting user data security to the maximum extent.

© AIOTI. All rights reserved.

2.13 Railway Flat wheel detection using Distributed Acoustic Sensing (DAS)

Description

Flat wheels can develop due to various factors related to the interaction between the wheel and rail. Key factors include wheel-rail interaction irregularities, excessive braking, and heat generation during braking. Flat wheels increase derailment risks and damage rail infrastructure, necessitating effective detection and maintenance practices to ensure safe and efficient railway operations. The implementation of Distributed Acoustic Sensing (DAS) for flat wheel detection is driven by the need for enhanced safety, operational efficiency, and cost savings. DAS offers continuous monitoring, comprehensive coverage, and non-intrusive installation, making it a practical solution for both new and established railway systems.

Distributed Acoustic Sensing (DAS) offers a solution for flat wheel detection by using optical fibers installed along rail tracks. These fibers serve as both communication data transmitters and acoustic sensors, enabling continuous and real-time monitoring of rail infrastructure. This technology enhances safety by reducing derailment risks, improves operational efficiency through real-time monitoring and prompt maintenance actions, and leads to cost savings by avoiding extensive repairs and minimizing downtime.

Optical fibers are installed along the rail tracks, serving as both communication lines and acoustic sensors. The DAS system is housed in a technical room located near the tracks, typically positioned at intervals of 20 to 40 km along the railway (see Figure 22a and Figure 22b).

(a) **(b)**

Figure 22: (a) Vibration mapping of rail activities can be achieved using fibre-optic acoustic sensing equipment (DAS) and real-time data processing; (b) Proposed layout for DAS system installation for a balanced and dense coverage.

The DAS system continuously monitors the rail infrastructure, providing uninterrupted data collection and analysis. Operators can visualize acoustic signatures in real-time, aiding in the early detection of flat wheels, as shown in Figure 23.

© AIOTI. All rights reserved.

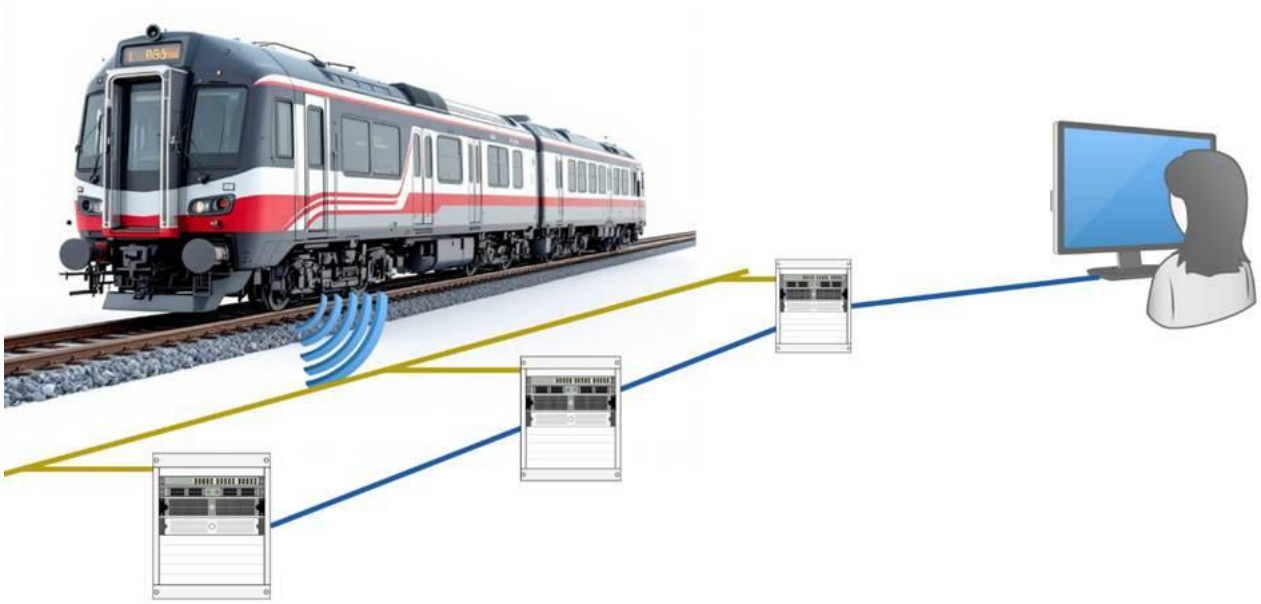


Figure 23: Principle of flat wheel detection and information transmission chain.

Upon detecting potential flat wheels, the DAS system generates immediate alerts, enabling prompt maintenance to minimize risks and enhance operational efficiency. Characteristic periodic peaks in DAS signals indicate wheel flats. By determining the train velocity from DAS data or other sources, we can verify whether these peaks align with the expected frequency based on the train speed and wheel diameter, confirming the alerts.

A field experiment was conducted on a 400-meter track with optical cables buried at a depth of 15 cm. The train, consisting of five carriages, had one wheel intentionally fitted with flat spots. Tests were conducted with optical pulse widths ranging from 20 to 50 nanoseconds. In the experiment, the train was traveling at a speed of 7.0 m/s (25.2 km/h). Assuming there is only one flat spot on the wheel, with a wheel circumference of 2.64 meters, the characteristic frequency can be calculated as 7.0/2.64 = 2.65 Hz. However, since the wheel exhibits multiple wheel defects forming a hexagonal pattern, we observe multiple frequency peaks in Figure 3. The smallest frequency peak corresponds to 2.65 Hz, while the highest frequency peak reaches 16 Hz, which is approximately six times the smallest frequency. This finding suggests a hexagonal geometric distribution of flat spots on the wheel, as confirmed by the experimental condition.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

N/A

Pre-conditions

N/A.

Triggers

N/A.

Normal Flow

N/A.

© AIOTI. All rights reserved.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

N/A.

Potential Requirements

N/A.

© AIOTI. All rights reserved.

2.14 Integrated RFID over FTTR

Description

Passive Internet of Things (IoT) plays an important role in industrial upgrading and digital transformation, especially in industrial automation, intelligent transportation, health care, and smart home applications. Passive UHF RFID technology is one of the widely used technologies of passive IoT in asset management, logistics and warehousing scenarios.

The traditional RFID solution uses devices, like handheld or checkpoint readers to identify items attached with passive RFID tags, which is widely used in retails, logistics, and warehousing. However, the current solution faces critical limitations:

- Limited coverage via a single device
- One use case is that the RFID read-out relies on manual walking to inventory items to achieve item statistics and management, which is inefficient
- Another case is that the deployment of the checkpoint reader in the logistic management scenarios to identify the moving items with passive tags is very complex. The reader needs to be closed to the items for accurate checking
- Multi-device interference:
- Due to lack of coordination between devices, when devices are used at the same time, the devices interfere with each other, resulting in performance deterioration of each device.
- Difficult to manage and no inter-networking:
    - IoT devices and communication networks are independent. Different vendors have different network access modes. There is no unified standard and management for interconnection with networks. The collaborative access capability is poor.

To address the issues mentioned above, integrated RFID over FTTR network enables seamless, comprehensive surveillance on things through constructing a distributed and synergistic RFID network, eliminating the interference problem between multiple independent devices. In addition, single FTTR network can cover Wi-Fi and RFID signals in a certain area. RFID signals can be used to manage items with passive RFID tags based on user-defined detecting profile (such as the cycle period or triggering). This automatic identification and management which does not require users to walk and inspect, greatly improvs management efficiency. It is also possible to locate objects through distributed multipoint RFID collaborative manner. Compared with traditional IoT, integrated RFID in FTTR network brings many advantages, including, such as automated operation, full coverage, and high efficiency.

To overcome the drawbacks of traditional independent RFID devices, the RFID reader is integrated in the FTTR device. For example, the RFID reader is integrated in the FTTR SFU device, as shown in Figure X. This technical architecture can implement collaborative management and control between FTTR data communication and RFID detecting. In this system, similar with Wi-Fi, the RFID inventory can be controlled by OLT using two-level optical-layer OAM to implement collaborative management, control, and scheduling for FTTR MFU and SFUs. The MFU is responsible for managing and controlling the SFUs and ensuring high-bandwidth and low-latency scheduling for internal local area transmission. The RFID integrated over SFU realizes awareness of people, objects and environment by identifying passive RFID tags attached to them.

Furthermore, the tag data obtained by each SFU may be aggregated on MFU, OLT, or cloud for further processing based on service and privacy requirement or regulation, implementing full lifecycle management of assets and materials for customers.

An entire distributed RFID architecture endowed with centralized coordinated scheduling by FTTR network can avoid intra-band disordered interference between devices. The traditional independent RFID readers consist of the protocol chip, emitter, receiver, and antenna. The detection distance of traditional RFID reader with monostatic configuration always limited by non-negligible self-interference, because the received signal is severely interfered by the leaking signal from the emitter. However, the RFID system integrated on FTTR SFUs, can use one SFU to

© AIOTI. All rights reserved.

transmit signals and the other SFU to receive signals to avoid self-interference, which greatly improves coverage and efficiency. Based on the management and control capabilities of the MFU and the AI computing capabilities of the OLT and MFU, the system can implement automatic RFID counting on the entire network without manual operations. Big data analysis using AI computing power can further explore data value and provide basis for customers' procurement and decision-making.

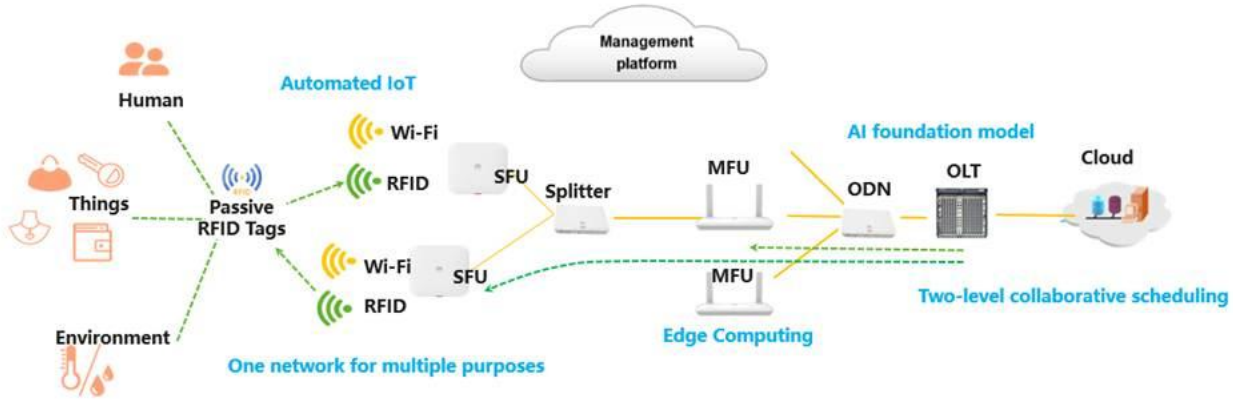


Figure 24: The scenario of integrated RFID over FTTR.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

N/A

Pre-conditions

N/A.

Triggers

N/A.

Normal Flow

N/A.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

N/A.

© AIOTI. All rights reserved.

Potential Requirements

The integrated RFID over FTTR requires the FTTR network to provide the following functions:

- **Hardware reliability:**
  - Component reliability: FTTR MFU or SFU should support fault diagnostic and repair of RFID component. Controller in the system should monitor working temperature of component and provide necessary protection and warning indication.
  - System reliability: MFU should support monitor the status of all the RFID component of different SFUs, and control them co-ordinately, flexibly and in an agile way for the minimum mutual interference and maximum whole performance of RFID network.

- **Smart applications based on RFID functions:**
  - Standard interfaces: FTTR should provide standard interfaces of RFID and plug-in development permission for general developers. Developers can build various and personalized applications based on the massive RFID data to directly provide data analysis, item management and so on for end users.
  - Mobile APP LAN access and control: FTTR should provide the interface for mobile terminals to control RFID function, such as starting, inventory policies, and data transmission.
  - Remote cloud access and control: FTTR should provide the interface for users to remotely control RFID function, such as starting, inventory policies, and data transmission, via Cloud.
  - Edge computing power: the FTTR AI edge computing engine provides abnormal data diagnosis and optimize the inventory strategy in real time, improving efficiency and accuracy.
  - Cloud computing power: operators provide cloud AI platform and service to users for massive RFID data analysis to achieve the purpose of wisdom recommendation and operation optimizing.

- **High performance of FTTR network:**
  - Stable, low-latency network: the protocol interaction of FTTR and RFID tag rely on stable and low-latency network between different SFUs. A high inventory efficiency requires extremely low latency <30 us.
  - Precise clock synchronization: the SFUs in the FTTR network should have precise clock synchronization to support joint transmission and receiving for RFID signals, which can significantly enhance the downlink budget and improve the coverage.

- **Smart O&M:**
  - In addition to remote management of FTTR network services, service provider should also manage the RFID services. Service provider should be aware of users' data privacy concerns and provide data storage and processing at different levels.

© AIOTI. All rights reserved.

2.15 3D video enabling via FTTR

Description

As home users increasingly demand immersive audio-visual experiences, traditional 2D viewing no longer meets the growing need for spatial realism, interactivity, and depth perception. While 3D display technology has seen significant advancements and widespread adoption in cinemas (e.g., polarized glasses systems), its integration into home environments faces critical barriers. These include limited native 3D contents, expensive display devices for 3D viewing, suboptimal 3D experience, and inadequate interactivity, all of which hinder the mainstream adoption of home-based 3D experiences.

Recent breakthroughs in AI technology have revolutionized content creation, shifting from conventional "capture-based" methods to more efficient "generative" approaches, particularly in 3D content production. AI-powered processing now enhances video clarity, immersion, and adaptability to user preferences, significantly improving user engagement. Concurrently, the evolution of 10 Gbps all-optical networks provides the technological backbone for seamless 3D experiences. High bandwidth enables uncompressed or lightly compressed transmission of 3D video, ensuring exceptional resolution and smooth frame rates. Ultra-low latency supports real-time interactivity and AI-driven rendering processes, further elevating immersion. Additionally, these networks facilitate cloud-based 3D rendering, reducing costs and technical barriers for home terminals while unlocking new market opportunities.

To address these challenges, the FTTR 3D solution proposes a home 3D viewing system leveraging 10 Gbps all-optical networks. By integrating cloud computing, high-speed transmission, and terminal playback capabilities, the solution achieves end-to-end coordination from content generation to home delivery, dramatically enhancing the 3D viewing experience.

As illustrated in Figure 1, home playback terminals utilize next-gen FTTR devices with embedded video processing to encode/decode cloud or locally stored 3D content into alternating high-frame-rate formats (e.g., >120Hz) for left- and right-eye views. This will transfer the 3D stream to a series of 2D stream. Processed signals are transmitted to high-refresh-rate displays. The 3D glass enables left and right eye switch. Here needs an accurate synchronization between FTTR devices (handing the framing of 2D imaging stream) and 3D glasses. The synchronization could be done via any wireless technologies with low-latency and high-reliability, such as Bluetooth, Wi-Fi or infrared. Such alignment between glasses and screen content delivers accurate stereoscopic visuals for premium home 3D experiences.

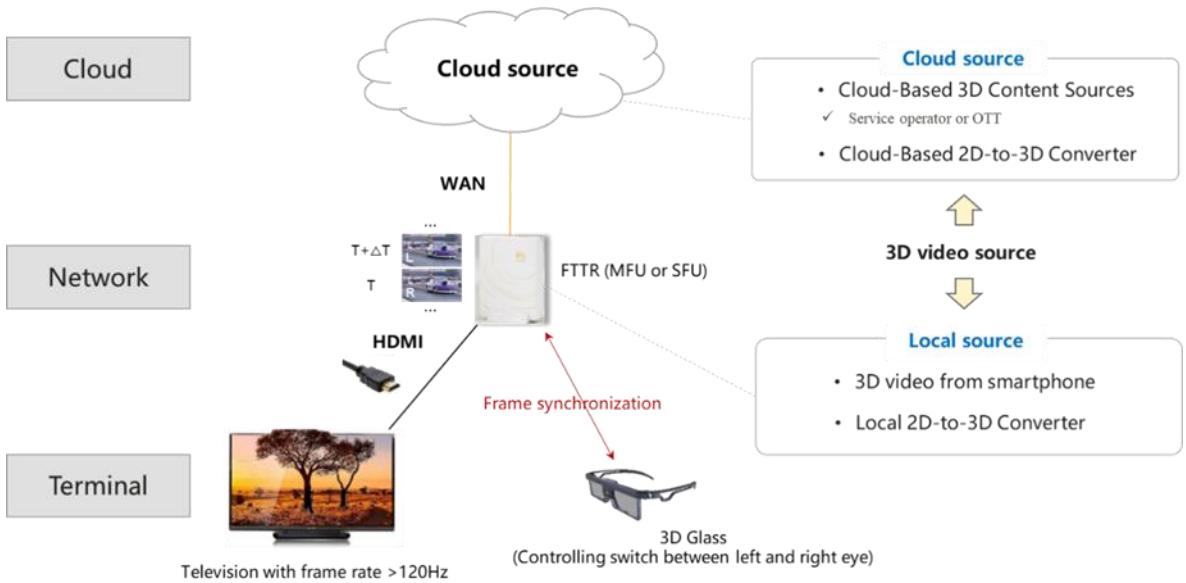


Figure 25: The scenario of 3D video enabling via FTTR.

© AIOTI. All rights reserved.

By leveraging 10 Gbps all-optical networks, the solution enables seamless access to cloud-hosted native 3D resources, including movies, games, and rendered models. Users may also convert 2D content to 3D through cloud services, significantly expanding content diversity and enriching user engagement.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 4 (work in progress).

Roles and Actors

N/A

Pre-conditions

N/A.

Triggers

N/A.

Normal Flow

N/A.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

N/A.

Potential Requirements

N/A.

© AIOTI. All rights reserved.

2.16 Intelligent Interconnection Bus for Optical Access Network

Description

With the popularization and development of AI applications, optical access networks and home networks should also evolve to meet the requirements of AI inference and training services. Access networks and home networks are relatively sensitive to device costs, so low-cost is an important factor for success. In addition, there are a huge number of ONU and FTTR devices in the existing network, and it is also necessary to consider how to enable the deployed devices in the network to support AI applications.

This use case descripts an intelligent interconnection bus for OAN to support collaboration between access-edge computing resources (deployed at OLT side) and FTTR or ONU. The intelligent interconnection bus provides a unified abstraction layer for network resources and computing resources, connecting OLTs (Optical Line Terminals), ONUs (Optical Network Units), and intelligent terminals (Figure 26), enabling unified management and coordinated scheduling of resources within the OAN.

With the in-depth integration of AI technologies into diverse services and applications, massive terminal devices including network devices, ordinary PCs and cameras are in urgent demand of an AI upgrade. Most of these terminals are inherently constrained by insufficient local computing power. Hardware renovation to locally deploy large AI models will incur excessive costs and form high investment thresholds for large-scale deployment.

To tackle this pain point, the Intelligent Interconnection Bus adopts a core design of unified management and intelligent scheduling of edge computing resources, delivering efficient, flexible remote computing services for terminals with weak or zero local computing capability in the optical access network. This case enables existing end devices to be quickly empowered with AI functions without the need for expensive dedicated hardware, substantially lowering the cost and technical barriers for widespread AI adoption in home networks and access networks.

AI Plugins can be deployed at the terminal application side. It serves as a unified entry point for applications to invoke remote computing power and provides standardized APIs for inference/training data from the terminal.

A management and control agent is deployed at operator's management platform which acts as the "brain" of the entire system. It is responsible for the global monitoring, management, and scheduling decision-making of computing resources.

The computing power is deployed on the edge side (OLT) equipped with computing resources such as NPUs and GPUs, which undertakes and executes specific AI inference tasks.

A data channel connects the AI plugin and the computing resources, carrying AI inference and training requests and result data. A control channel is used for computing and other network resource application, release, and status synchronization. This control channel can also be used for computing node registration, load reporting, and task distribution.

Following is the example of request procedure of the Intelligent Interconnection Bus:

(1) The AI application in the terminal initiates an AI inference and training resource request to the management and control platform through the unified API of the AI Plugin via the user control channel.

(2) Based on the global resource status, the platform matches and assigns the optimal computing resource for the incoming request.

(3) Subsequently, a dedicated data channel is established between the AI Terminal and the allocated computing node. The control platform is responsible for performing the network resource computation, that is, allocate the required resources for the dedicated data channel based on the target AI service.

© AIOTI. All rights reserved.

(4) Upon channel setup, the terminal application continuously delivers inference requests and receives corresponding results through the established data channel.

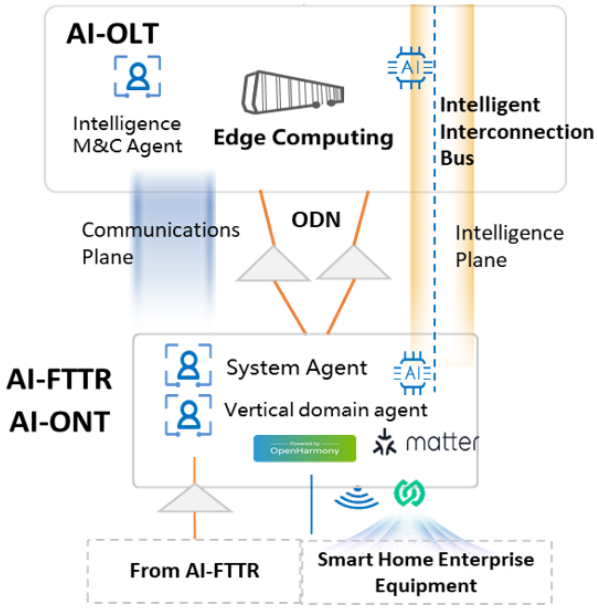


Figure 26: The scenario with an Intelligent Connection Bus between the various Network Elements at the Edge of the Network.

Source

ETSI ISG F5G, Fifth Generation Fixed Network (F5G); F5G Advanced use cases, Release 5 (work in progress).

Roles and Actors

N/A.

Pre-conditions

N/A.

Triggers

N/A.

Normal Flow

N/A.

Alternative Flow

N/A.

Post-conditions

N/A.

High Level Illustration

N/A.

Potential Requirements

N/A.

© AIOTI. All rights reserved.

2.17 Generative AI Supported Analysis and Evaluation of SmartX Data

Description

SmartX environments, like smart city or smart health, generate huge amounts of data that need to be analysed in order to obtain valuable information out of it to help monitoring and/or optimizing their underlying processes. In a smart city these processes can be pretty complex with multiple parameters and relations, some maybe even not clearly identified at first. The smart health scenarios can be less complex, but still some diversity in that respect is possible.

This analysis of data and interpretation of the situation has several implications. First, it may require data from different sources, so it needs to be handled in a privacy protecting and ownership protecting way. Second, depending on the real-time requirements of the specific scenario data throughput and QoS aspects may play a crucial role. Additionally, the distribution of data storage and processing elements of the system can support both these aspects.

In order to support the split of data sharing and processing planes in distributed SmartX systems, a middleware can be applied that allows to connect individual computational or data management units (see Figure 27). These units (represented by red shapes in the figure) can belong to independent stakeholders and the middleware allows them to share data in privacy and ownership protecting way and to run their individual data processing services that read and write data into the middleware allowing generating knowledge based on own and foreign data.

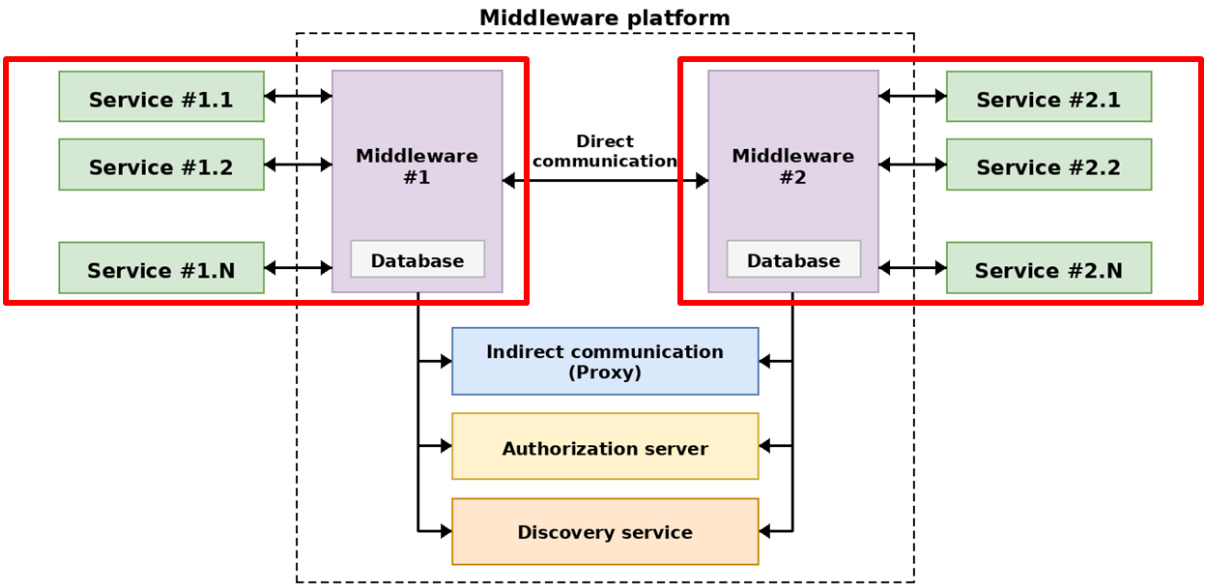


Figure 27: A middleware platform connecting two computational units (red shapes) running their data processing services.

Such middleware platform allows constructing interoperable distributed and multi-stakeholder systems that may even combine scenarios from different domains, like energy, smart city or smart health together. But even focusing on a single scenario the split between data sharing and processing simplifies the construction of SmartX systems as it clearly modularises the components and allows, for instance, some of the services to take the role of adapters that connect the middleware platform with external data sources, like sensor networks (see Figure 28). By that it is possible to implement the complete data chain, from data acquisition within the environment, data transport, via the data collecting and processing, to the final step where the results (or knowledge) is presented to the end users.

© AIOTI. All rights reserved.

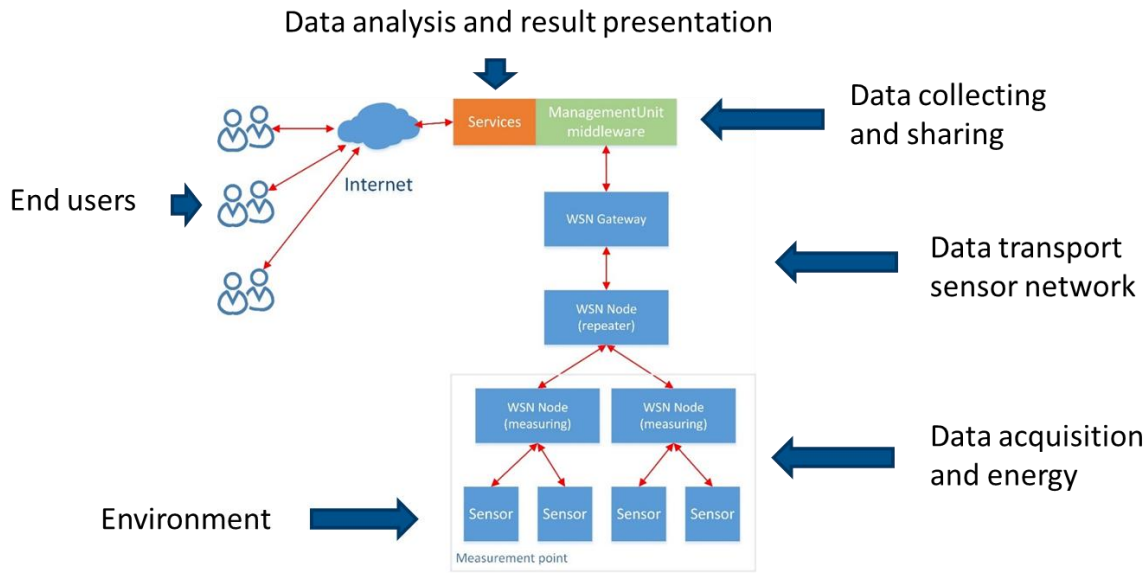


Figure 28: A sensor network connected to the middleware layer to form the complete data chain.

However, this simple and straight data path benefits a lot when connected with other scenario implementations, as synergies can be exploited. Combining multiple data collecting sensor networks and multiple (data) management units allows to implement complex scenarios. Data from multiple processes and multiple application domains can be shared within a common data plane (and even globally using connectors to form a federated data space) and a hierarchical (or any other) topology of connected data management units allows to distribute the data processing application logic optimally to reduce delays and unnecessary data exchange (see Figure 29).

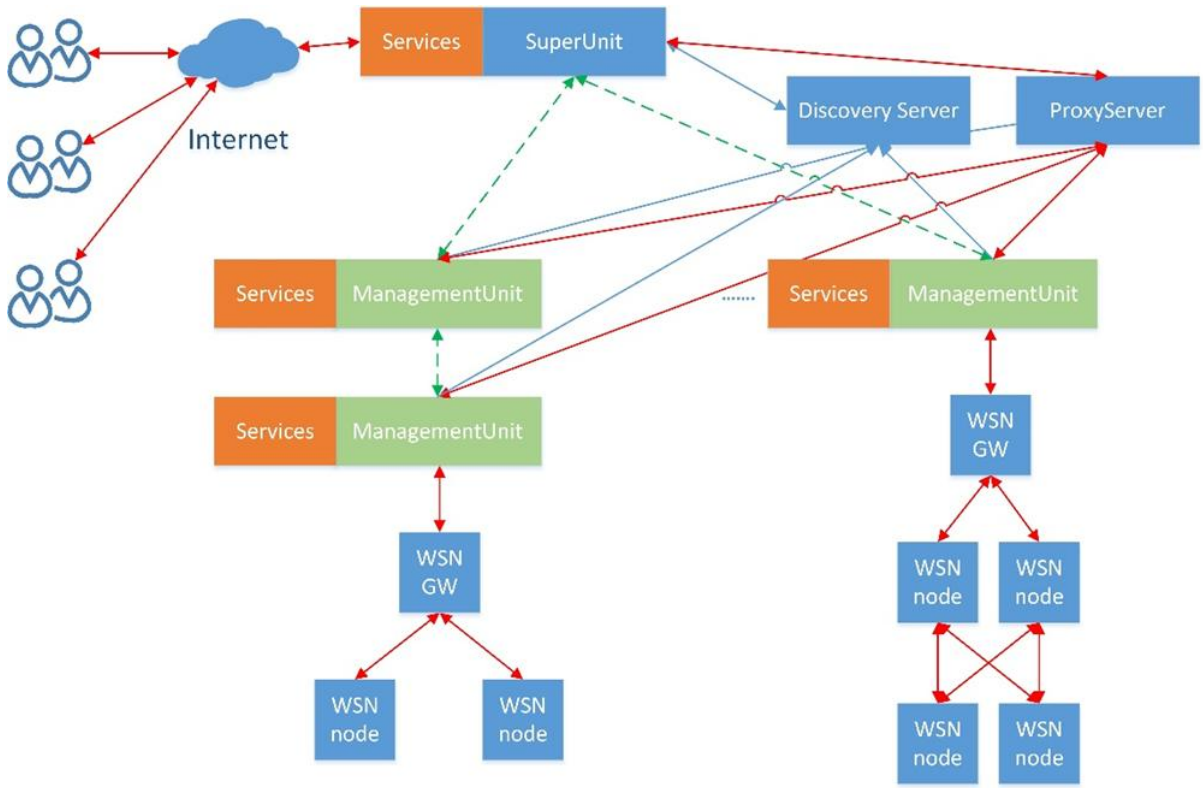


Figure 29: A combination of multiple units to implement more complex data sharing and processing scenarios.

The interoperability between the different applications is provided by the way the data is shared within the middleware platform – the data is represented by *variables*. The meaning (and type) of a variables is shared in the system, and the individual *values* (or *instances*) of a variable form a time-series. Services representing different stakeholders (their owners) access the variables

© AIOTI. All rights reserved.

available in the shared data space within the middleware via the data interface. Finally, the owner of the service who wrote a value (an instance) defines who else can access that specific value, thus, we can experience a time-series of values for a given variable belonging to different stakeholders (or end-users). A service can also subscribe to new values of a variable - it will get all the values that it is allowed to access.

Within the above-mentioned context, a diversity of applications can be realized. Specifically, the data processing services can be implemented in many ways. The basic way are simple services that know exactly what they process, they read and write the specific variables and process them in a simple algorithmic way. This can involve simple (or not) operations or detection of passing a threshold. For the smart city and the smart health applications it might be an analysis of the values of a parameter, like heart rate or air temperature and its presentation on a graphical diagram in time domain to show the recent changes and especially those beyond given thresholds. Even supporting some scripting or other means to provide flexibility these services are fixed. But, on the other hand, algorithmic data processing approaches are usually reliable and repeatable. They simply provide their results to the personal (end-users) as they were requested to do.

But the services can go far beyond that. Under the current developments in generative AI the services can also obtain a new level of data processing or handling abilities. They could provide flexibility in terms of applying different data processing steps, in terms of generating graphical output in different ways, or in terms of allowing human-like reasoning and pattern matching. In order to support the end-users, the GenAI-based services can (e.g. in parallel to the algorithmic ones) run alternative analytics or alternative on-demand analytics. They can also collect the knowledge on the relations between the processes and data, helping to create digital twins or running and fine tuning these. Further, the GenAI based services can be deployed at different levels, e.g. as services supporting individual end-users at their premises and on their data management units, but also as cloud-based multi-user services.

To support the generative AI in reasoning it is needed to extend the semantic and ontology support in the middleware platform to help describing the data and relations more precise. It is also possible to make this plane bidirectional, allowing the generative AI based services to reason on that as well and to express it in a human readable way for human feedback or for further processing.

For obvious reasons the management units need to be connected, and optical networks provide the required throughput and QoS levels.

Operation of the use case:

- Use case vital and environment parameters: These parameters related to health and safety (e.g. AAL) are collected in the middleware platform and their analysis and evaluation is supported by an LLM-based service (patient and doctor focused, speech in and out).

- Use case smart city parameters: Parameters related to the city processes (e.g. traffic, air quality, water levels) are collected in the middleware and their analysis and evaluation is supported by an LLM-based service (multi-user, cloud based).

Source

- EU INTERREG Project SensorNet

- EU INTERREG Project SmartRiver2

© AIOTI. All rights reserved.

Roles and Actors

- End-user: citizens, local government officers, people with disabilities, elderly people, doctors, caregivers, families
- System providers: Developers, Infrastructure providers

Pre-conditions

- The middleware instances are deployed in the end-user premises or devices and are connected
- The middleware platform infrastructure is available and allows sharing the data and meta data among the connected middleware instances, providing the security and information infrastructure
- Connection to the data sources (sensor networks) is available, allowing data collection from the environments using the proper adapters (services translating data)
- Services offering the data processing are available for local as well as cloud deployments to implement the desired application services

Triggers

- Evolution of data sharing and market for applications for that
- Wish for better and more flexible monitoring and analysis abilities in different domains

Normal Flow

- On a general level, there is a bidirectional data and knowledge exchange between the end-users and the distributed intelligence in form of data processing services in the system.
- Use case vital and environment parameters: The vital parameters of the monitored persons (special case: sport activity monitoring) and the environment parameters are collected in the data sharing middleware. The basic services can detect events and compute information out of the parameter values. The LLM-based services available to the patients and to the doctors allow them to formulate free processing requests with speech. These requests can include analysis or statistics steps, but also suggestions on how to improve the sport activity in the current environmental conditions (too hot, rainy, etc.).
- Use case smart city parameters: The parameters of the monitored city processes are collected in the data sharing middleware. The simple services can analyse these for individual scenarios or in conjunction. The LLM-based services available for the citizens (cloud-based service usage via Internet) and for the city officers (cloud-based or locally deployed) allow to analyse the current situation in the city and to present the processing outcome in an attractive way.

Alternative Flow

- A combination of scenarios: citizens, officers and all other kinds of people (private or professional) have the data sharing middleware deployed on their devices (smartphones and PCs) and define the data they are sharing and who is allowed to access their data. A market of services exists that allows the people to find and install on their devices the right services they need for monitoring and optimizing their respective processes related to sport activity, health in general, energy efficiency, etc. Additionally. a market of data

© AIOTI. All rights reserved.

sources exists, where people can buy reliable data they need for their services, like weather predictions, etc. People can also sell their data to others.

Post-conditions

The data sharing infrastructure is provided and maintained. The markets are provided and maintained.

High Level Illustration

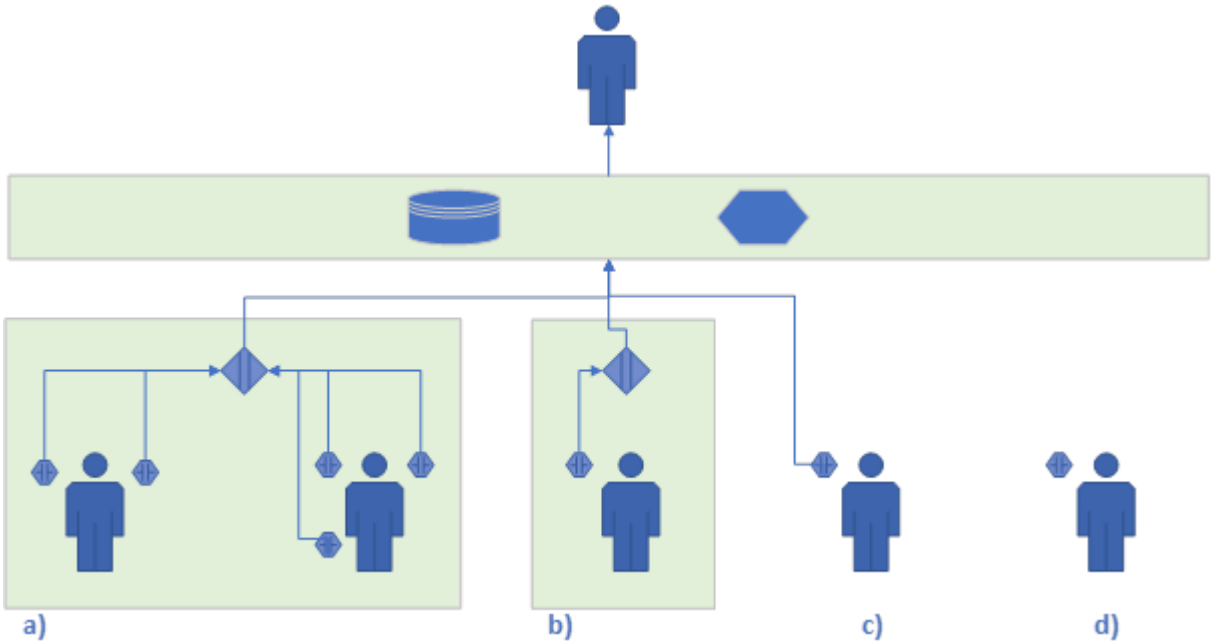


Figure 30: Different complexity levels for vital and environmental parameter use case.

Potential Requirements

Functional requirements

- The semantic and ontology meta data needs to be shared and maintained
- Achievable data rates and QoS need to fulfil scenario requirements
- Enough processing power for the right LLM service at the right deployment

Non-Functional requirements

- Infrastructure and interests for data sharing and markets needed
- Security, data ownership and privacy needed

© AIOTI. All rights reserved.

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 2.1 Cloud-based medical imaging)

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

© AIOTI. All rights reserved.

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 2.2 Cloud-based visual inspection in production)

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 2.3 Cloud-based control of automated guided vehicles)

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 2.4 Cloud-based control of production via optical wireless communication)

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

© AIOTI. All rights reserved.

Use case: Protecting sensitive data within smart cities (Section 2.5 Protecting sensitive data within smart cities)

- 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: Computing collaboration in optical access networks (Section 2.6 Computing collaboration in optical access networks)

- The residential or SME network shall use optical backhaul of the Wi-Fi APs such that control and management traffic for an intelligent optimization is not getting too 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 2.7 Robotics as a service)

- 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.

Use case: Intelligent Interconnection Bus for Optical Access Network (Section 2.16 Intelligent Interconnection Bus for Optical Access Network)

- The use case requires implementations with AI compute capabilities on OLTs, ONU, and FTTR network elements. For the interaction and allocation of resources where the compute for AI workloads is happening, the intelligent interconnection bus used and needs to meet the appropriate performance depending on the AI applications and the AI user service requirements.

© AIOTI. All rights reserved.

- The intelligent interconnection bus dealing with coordination and allocation of AI workloads requires standards to be developed, based on the larger computing continuum approach.

Other requirements

In Figure 31, 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 marketplaces.

- **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.

© AIOTI. All rights reserved.

- **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.

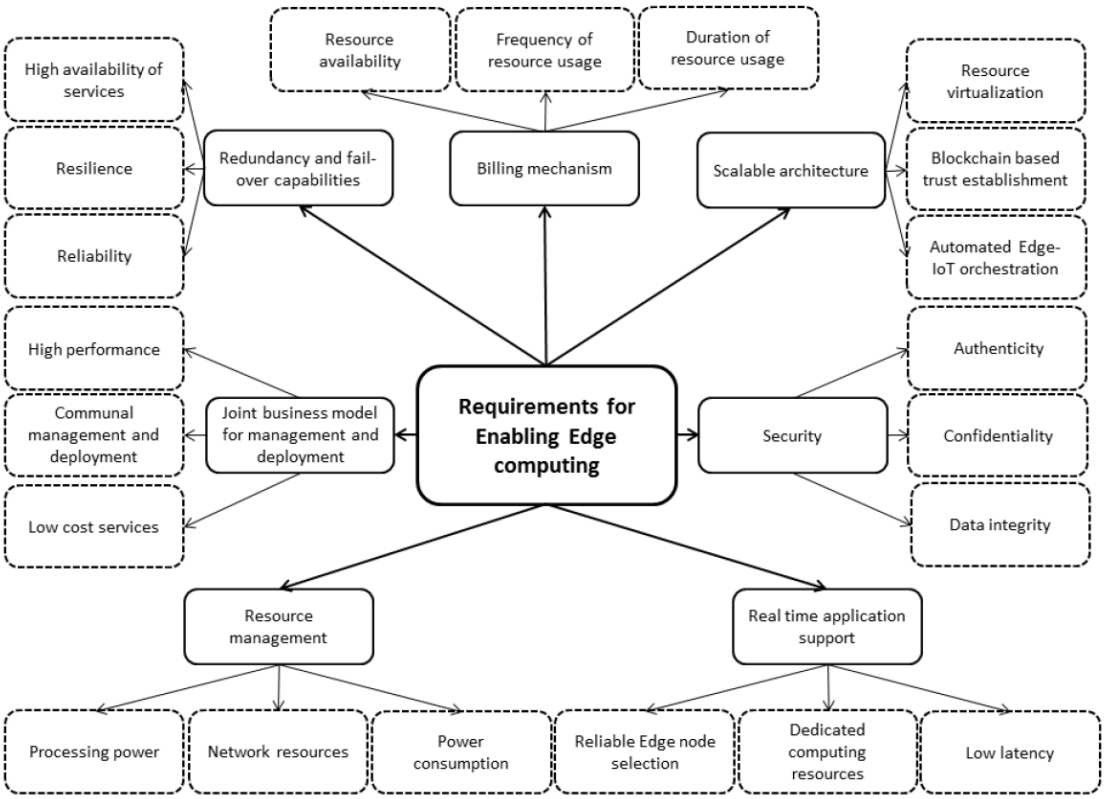


Figure 31: Requirements for enabling edge computing and computing continuum [ZaAh19].

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

© AIOTI. All rights reserved.

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

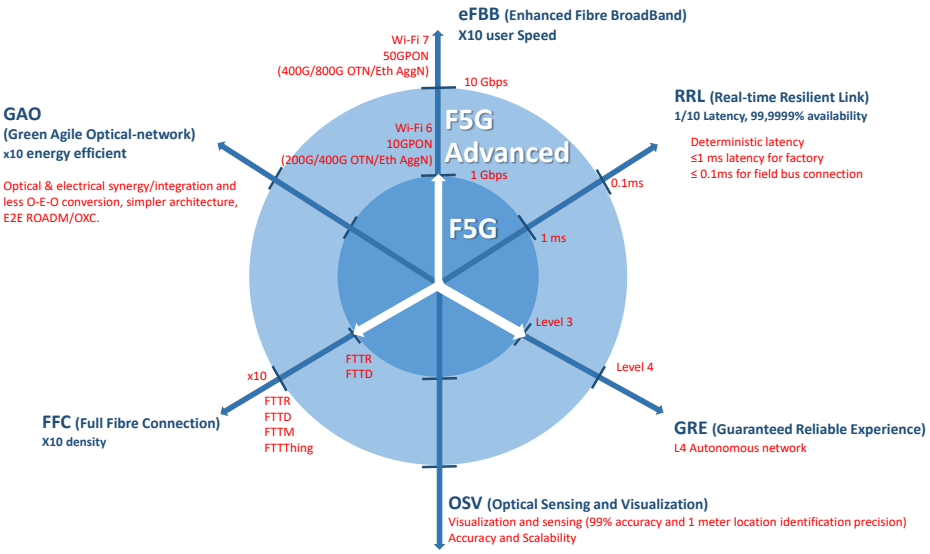


Figure 32: F5G Advanced generation dimensions and KPIs.

References:

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

© AIOTI. All rights reserved.

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.

© AIOTI. All rights reserved.

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).

© AIOTI. All rights reserved.

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 33, 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 Wi-Fi-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).

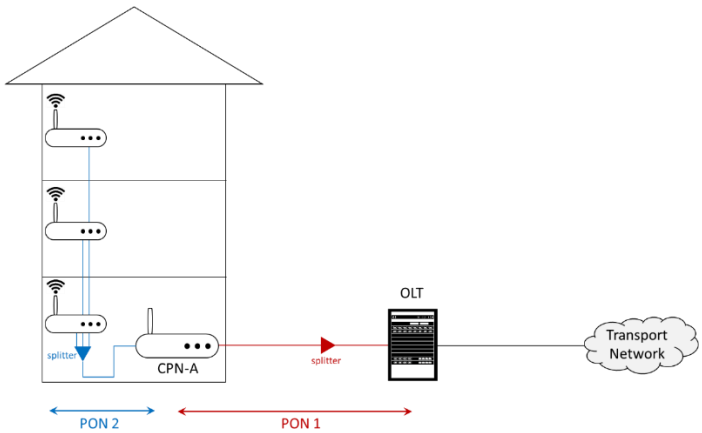


Figure 33: Cascading PON architecture.

✓ **Energy efficiency**

That means to migrate from today's architectures to increasingly 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).

✓ **AI Computing collaboration**

With the availability of AI computing capabilities on various network nodes and at different places in the network, collaboration between the different AI compute locations is needed to achieve an optimal AI computing continuum. Also, the network elements are expected to provide access to the intelligent computing bus to access the right AI compute node (route AI messages to the appropriate compute node, taking the application requirements into account. See below for different edge computing scenarios.

© AIOTI. All rights reserved.

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 34. 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: 1) CPN gateway, 2) access node (OLT), 3) aggregation edge, or in the cloud. In general, edge computing functionality can be part of the CPN gateway, OLT or Aggregation Edge (the combination of 1, 2 and 3). 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.

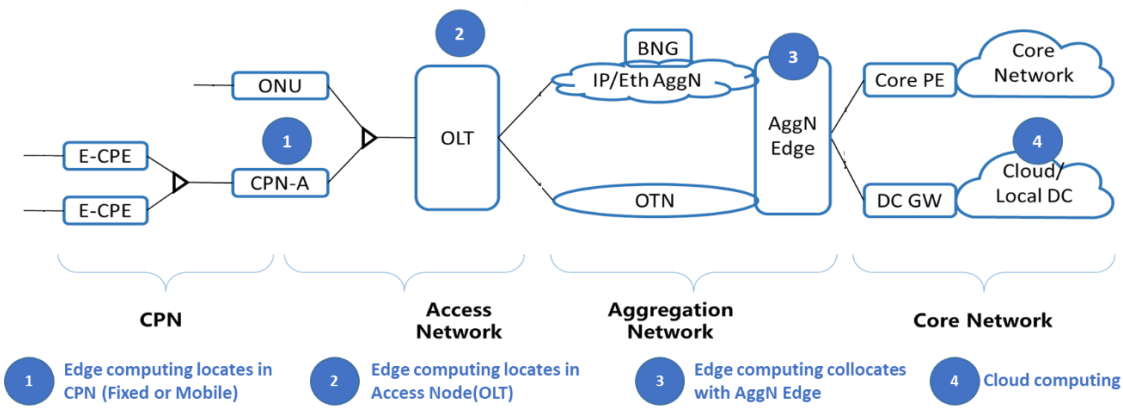


Figure 34: Network topology for edge and cloud computing.

O 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.

O 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 O). 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 O. In this case, the requirements for the link between edge and central cloud will be even looser than for case O, as more power full edge computing can handle more pre-processing works. Case O raises stricter requirement to the access network between CPN and OLT, which sometimes exceeds the threshold of PON technology.

© AIOTI. All rights reserved.

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 is 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.

© AIOTI. All rights reserved.

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 35 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.

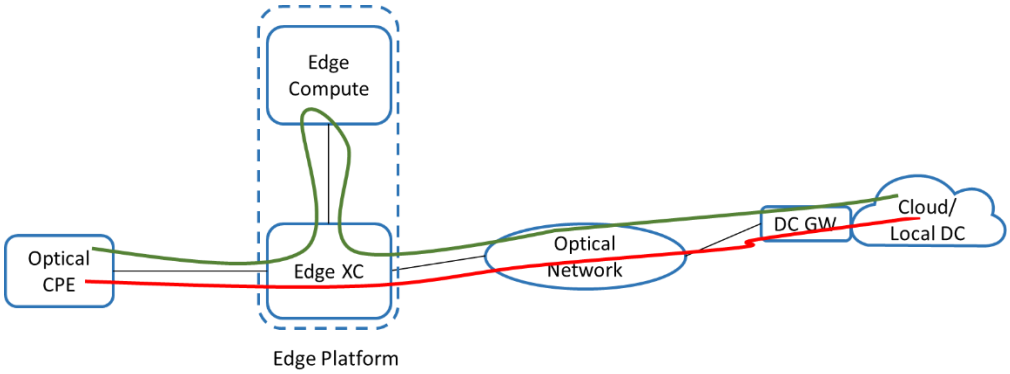


Figure 35: Illustration of the optical cut-through approach (read line).

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.

© AIOTI. All rights reserved.

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 workloads, 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.

© AIOTI. All rights reserved.

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).

© AIOTI. All rights reserved.

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 36 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 37 and Figure 38 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 36. 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 42.

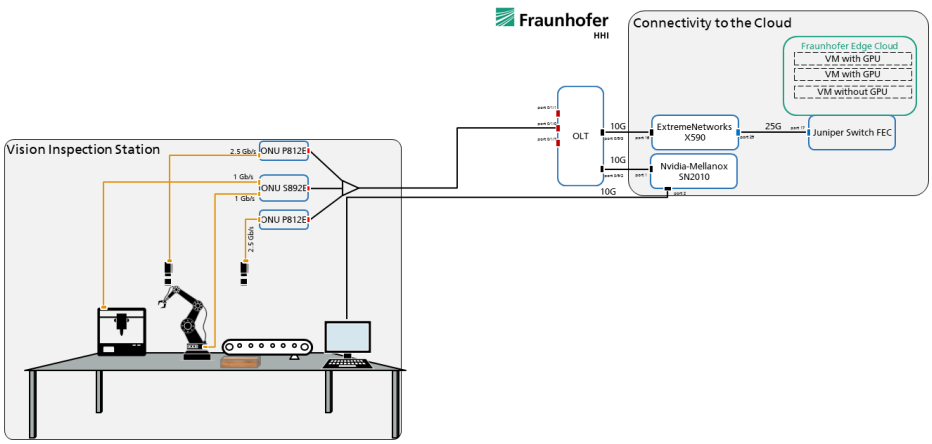


Figure 36: Testbed architecture and network slicing configuration.

© AIOTI. All rights reserved.

Figure 37: Basler Camera.

Figure 38: COBOTTA IP30.

The sequence diagram in Figure 39 and Figure 40 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 41), 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).

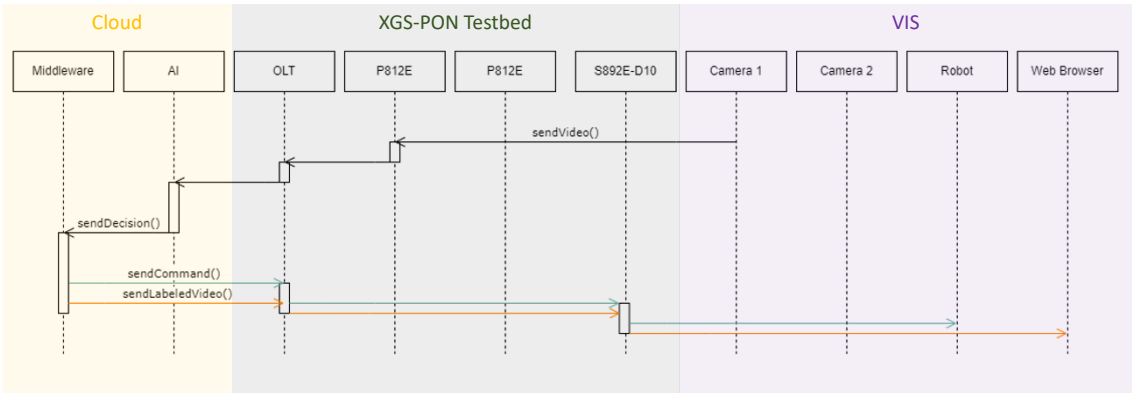


Figure 39: Testbed architecture and network slicing configuration.

© AIOTI. All rights reserved.

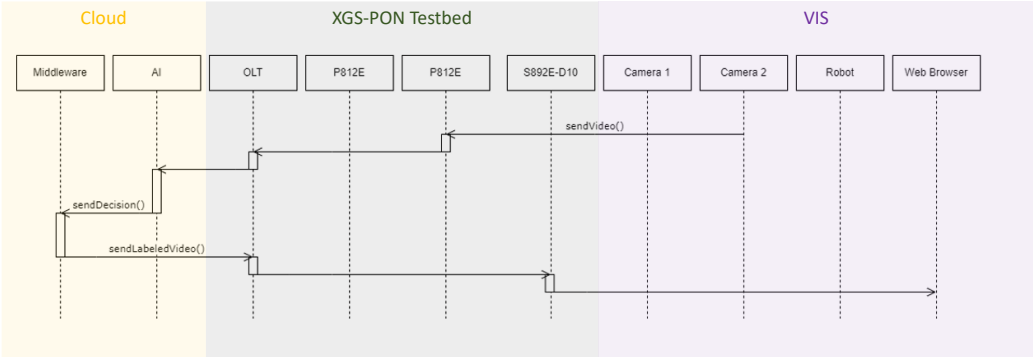


Figure 40: Sequence diagram for camera 2 operation.

Figure 41: Faulty(left), non-faulty (right) objects.

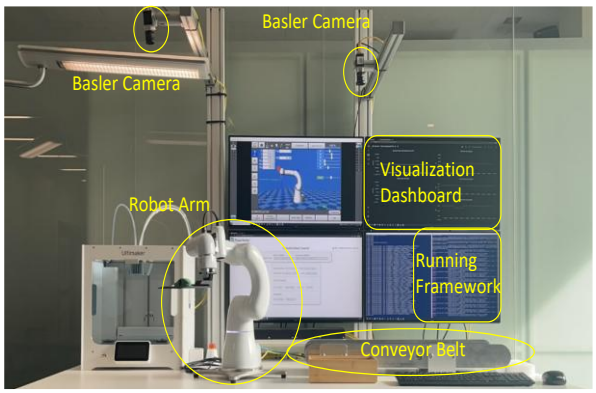


Figure 42: Physical setup.

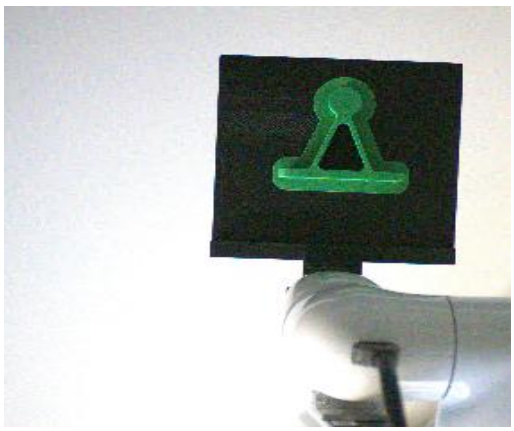


Figure 43: Camera 1 view from PylonViewer.

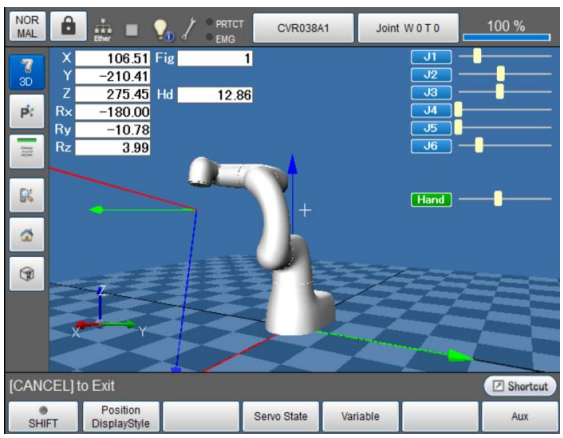


Figure 44: VirtualTP's main screen.

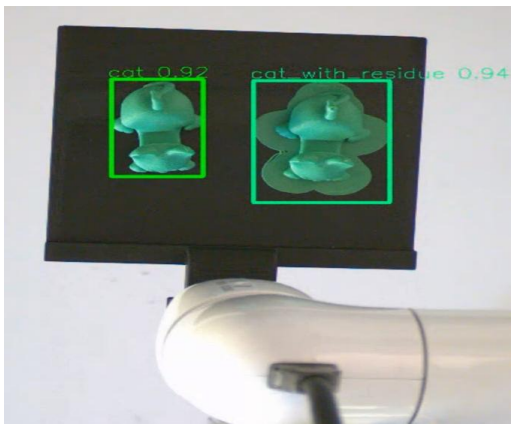


Figure 45: Classification output first camera.

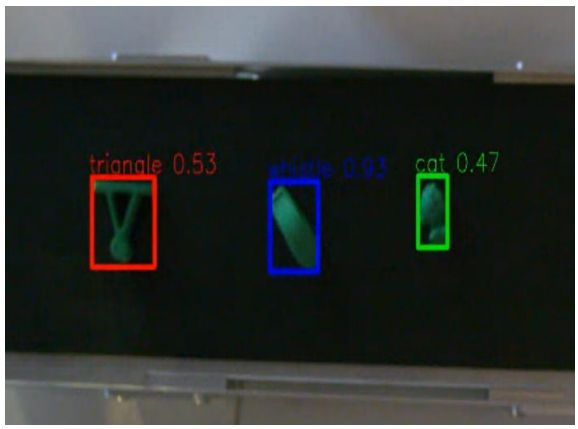


Figure 46: Classification output second camera.

© AIOTI. All rights reserved.

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 42).The sequence diagram in Figure 40 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 43). It is also capable of managing the robot autonomously through a Graphical User Interface (GUI) which can be used for testing and programming (Figure 44). The outcome of the vision inspection models based on the captures from camera1 and camera2 are provided in Figure 45 and Figure 46, 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 47 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 48). 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 49 and Figure 50 show the telemetry retrieval process for traffic and energy data, respectively, as a sequence diagram.

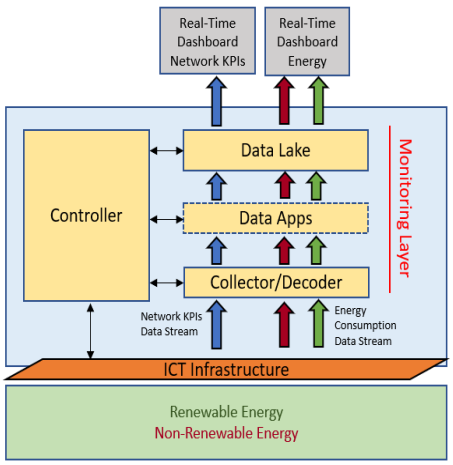


Figure 47: Framework architecture.

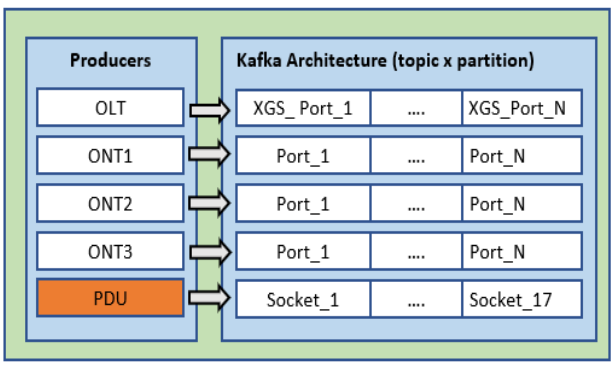


Figure 48: Kafka architecture.

The code running in the edge/cloud starts the traffic monitoring process (Figure 49) 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.

© AIOTI. All rights reserved.

The code running in the cloud starts the energy collection process as well (Figure 50). 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.

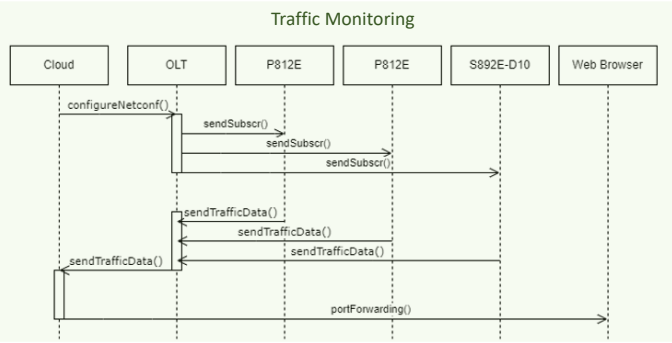


Figure 49: Sequence diagram for traffic monitoring.

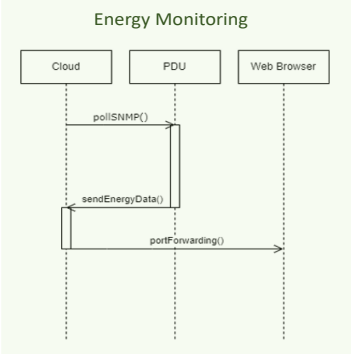


Figure 50: Sequence diagram for energy monitoring.

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 36. 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 51. 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.

© AIOTI. All rights reserved.

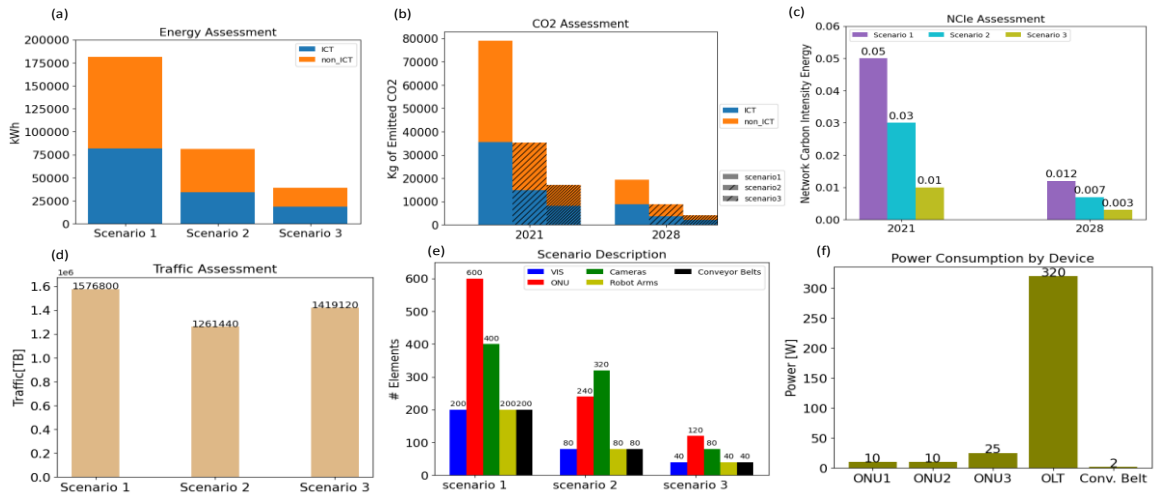


Figure 51: (a) energy assessment; (b) CO_2 assessment; (c) NCIe assessment; (d) traffic assessment; (e) scenario description (f) power consumption of other devices.

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".

© AIOTI. All rights reserved.

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 52). 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 Wi-Fi6-enabled ONUs.

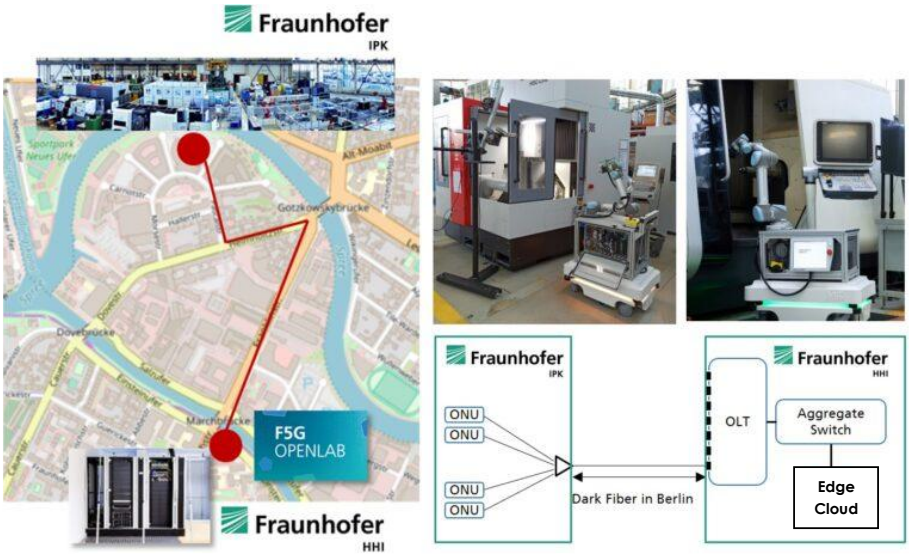


Figure 52: Setup of the use case in the city of Berlin.

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 Wi-Fi6 ONUs are installed, supporting the roaming of clients.

The PoC consists of the following services based on the ROS2 framework (see Figure 53):

- 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

© AIOTI. All rights reserved.

- 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)
- Visualisation of planning scene: running on HMI-edge at shop floor
- Orchestration of the scenario: running on cloud (edge DC)

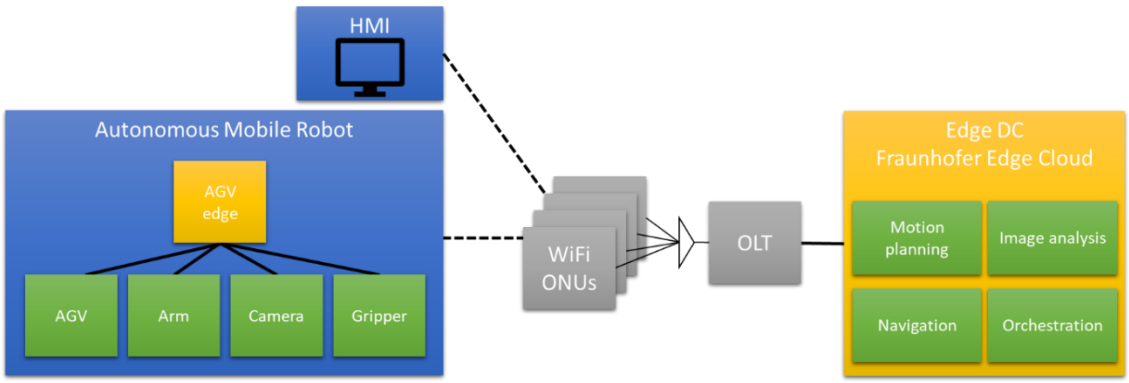


Figure 53: Components involved in the PoC.

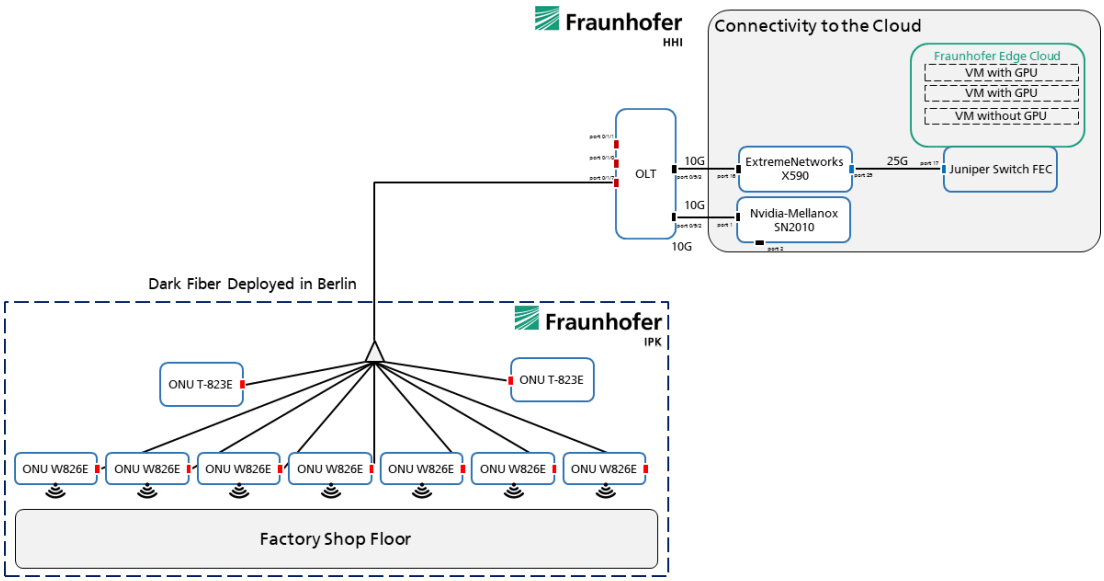


Figure 54: Networking setup.

Network Architecture of the PoC

The testbed (Figure 54) 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 Wi-Fi 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 53, 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 Wi-Fi 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.

© AIOTI. All rights reserved.

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 55 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 Wi-Fi6 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 56 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 57.

© AIOTI. All rights reserved.

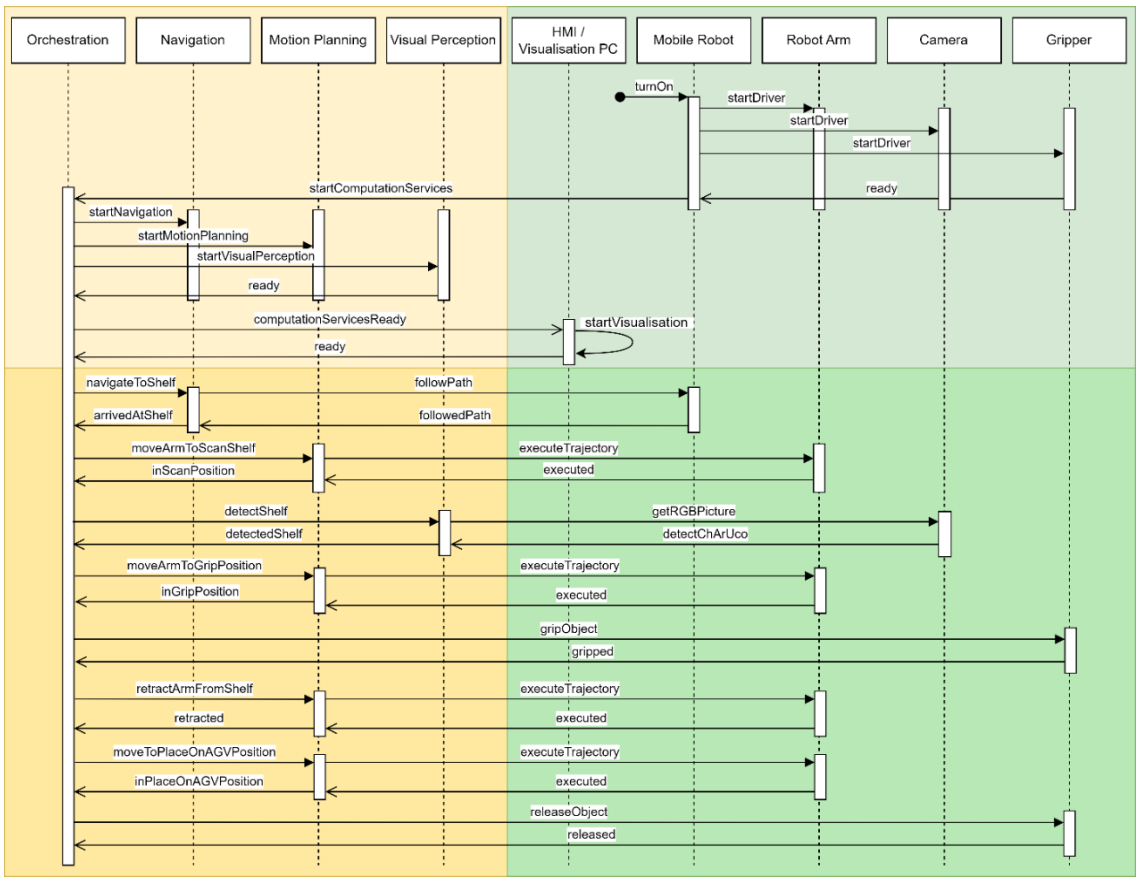


Figure 55: Sequence diagram of all used services in VM (orange shade) to control AGV and robot in shop floor (green shade) with the setup first (darker shade) and then picking up a part from a material shelf (brighter shade) as an example.

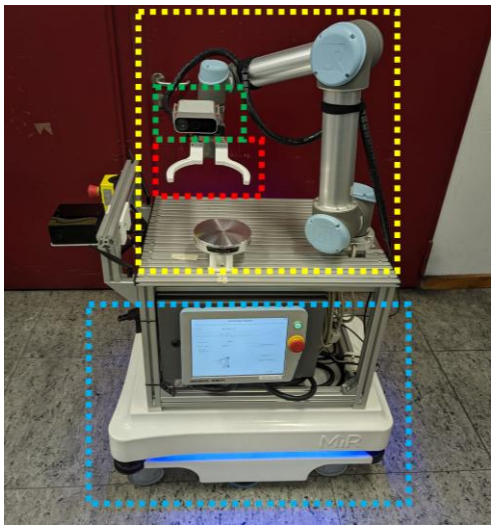


Figure 56: Components of the robot: mobile base (blue), manipulator (yellow), camera (green), gripper (red).

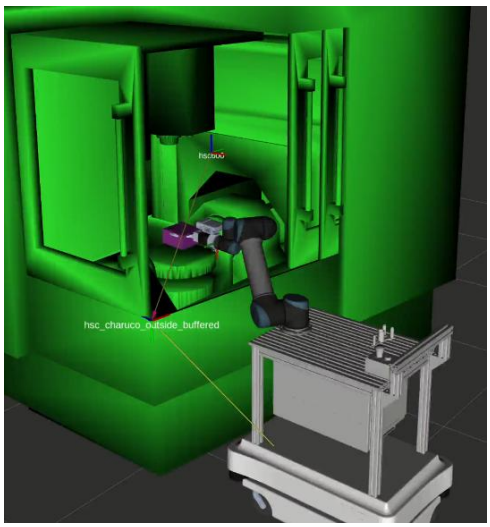


Figure 57: Visual motion planning environment on the VM.

© AIOTI. All rights reserved.

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 Wi-Fi ONU 6 at startup (Figure 58), 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 58 is due to a handover to Wi-Fi ONU2 (Figure 59). 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 58, since the industrial ONU (Figure 60) 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 61 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 59. Such traffic represents the set of instructions sent from the edge cloud to the AGV to properly command the placement operation.

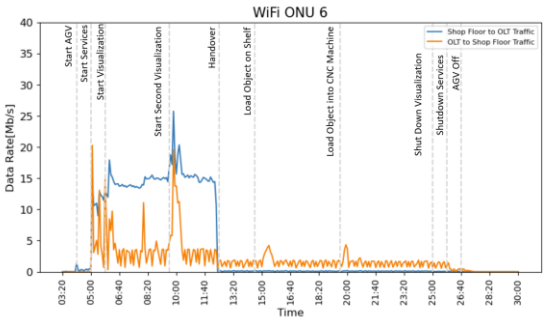


Figure 58: Traffic exchange ONU6.

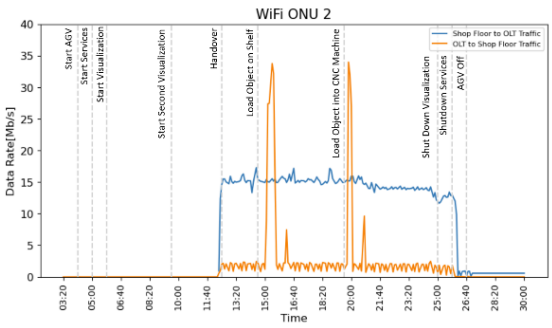


Figure 59: Traffic exchange ONU2.

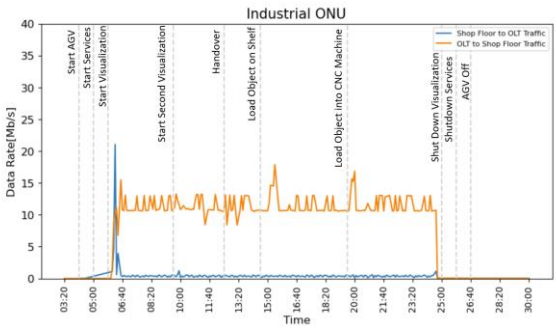


Figure 60: Traffic exchange for industrial ONU.

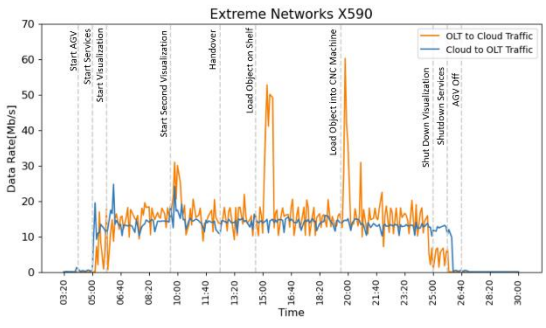


Figure 61: Traffic exchange cloud.

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.

© AIOTI. All rights reserved.

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 Wi-Fi APs leads to small probability of roaming: The Wi-Fi6 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.

© AIOTI. All rights reserved.

11. PoC report: Generative AI for F5G-A Network Management Tasks

Overview

Efficient management of modern optical access networks requires interaction with multiple independent management systems, including configuration interfaces, telemetry databases, analytics dashboards, and inventory systems. Performing such operations often demands expert knowledge and the use of specialized tools, which raises the barrier to efficient network operation. This PoC demonstrates an LLM-based AI assistant that enables users to manage the F5G Network Fabric through a single, natural-language interface. The assistant is deployed on the F5G OpenLab testbed at Fraunhofer HHI, where it connects to live management and monitoring components of the network.

The system allows operators to query the network inventory, retrieve connectivity and topology information, perform configuration actions on devices via NETCONF, query telemetry data from InfluxDB, generate Grafana dashboards for visualizing performance metrics, and access power-monitoring and energy-consumption information, all through conversational interaction. All LLM components run fully on-premise within the OpenLab infrastructure to preserve data privacy and guarantee local processing without any cloud exposure [ZAIDOFC26].

Networking Architecture of the PoC

The PoC is implemented on the F5G OpenLab testbed operated by Fraunhofer HHI. The testbed includes an optical access network with a single OLT and multiple ONUs, interconnected via a PON segment, and a management and analytics environment consisting of InfluxDB, Grafana, NetBox, and power-monitoring systems. The software stack of the assistant follows a modular, layered architecture comprising the following components.

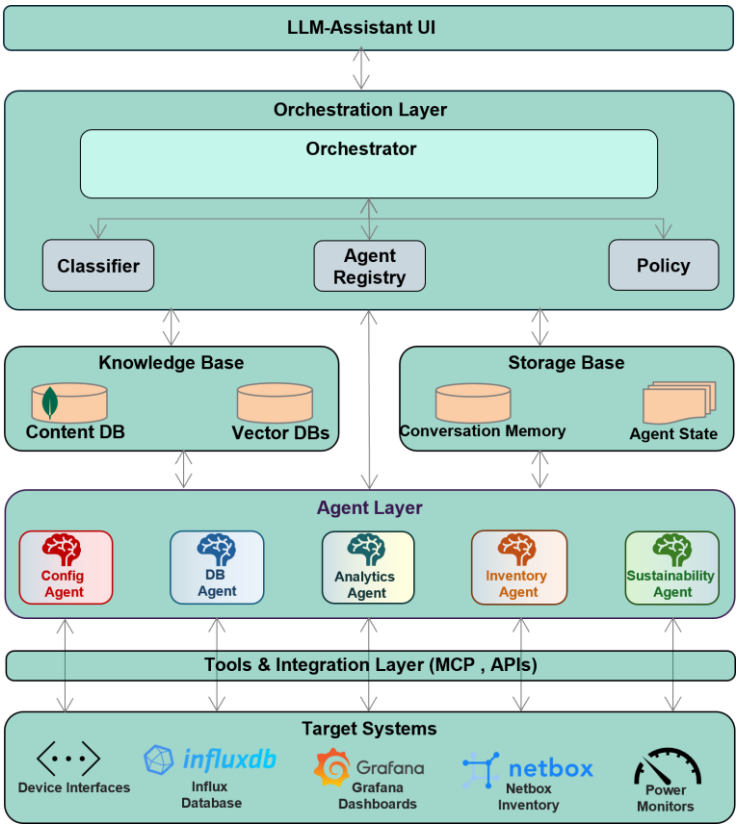


Figure 62: Software Architecture of the LLM-based AI Assistant.

© AIOTI. All rights reserved.

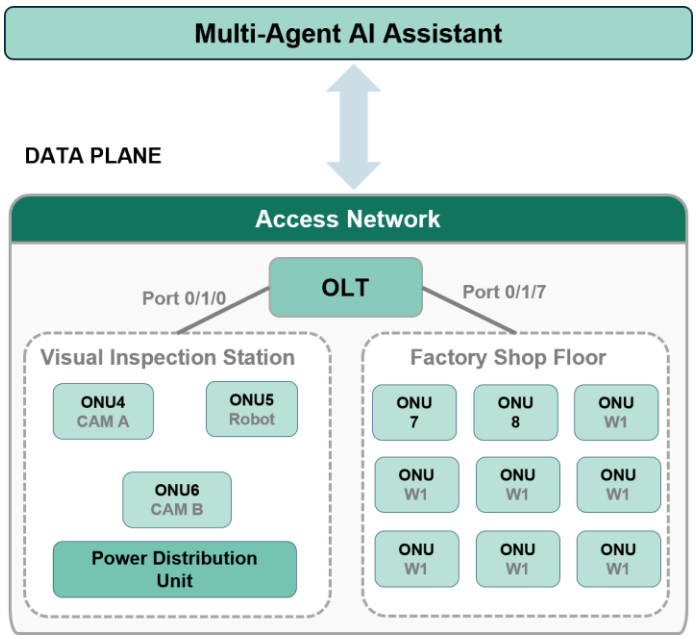


Figure 63: PoC Testbed.

User Interface

The user interface is a web-based chat application that allows operators to issue natural-language commands and receive responses in real time. The GUI features a conversation panel, agent status indicators showing which agent is currently handling a request (e.g., CONFIG, INVENTORY, DB, ANALYTICS, SUSTAINABILITY), a suggestion bar with example queries, and a debug panel for monitoring system-level events such as WebSocket connections and agent routing decisions.

Orchestration Layer

The orchestration layer is based on an open-source multi-agent orchestrator responsible for interpreting user inputs, enforcing policies, and dispatching requests to the appropriate specialized agents. It performs request classification to determine whether an incoming message is a new query or a follow-up to an ongoing conversation. A router agent resolves ambiguity when a request could be handled by multiple agents, and the orchestrator maintains conversation state and task summaries across interactions. An agent registry and policy module govern agent selection and execution.

Knowledge Base and Storage Base

The system maintains shared knowledge and storage resources including a content database with configuration examples and reference data, a conversation memory for maintaining multi-turn dialogue context, and an agent state store for tracking the internal state of each agent across sessions.

Agent Layer

The agent layer consists of five independent LLM-powered agents, each dedicated to a specific management domain:

- **Configuration Agent:** Interfaces with the OLT and ONUs via NETCONF/YANG to retrieve and modify device configurations, including telemetry subscription parameters such as sensor groups, sample intervals, and destination addresses.

© AIOTI. All rights reserved.

- **Inventory Agent:** Queries NetBox via its REST API to retrieve topology and device information. It supports context-aware follow-up queries, for example first asking what is connected to a specific OLT port and then drilling down to the end devices connected to the listed ONUs.
- **Database Agent:** Queries the InfluxDB time-series database to discover available measurements and their fields. It translates natural-language requests into InfluxDB queries and returns structured results, such as listing all available measurement types or enumerating the fields within a specific measurement.
- **Analytics Agent:** Generates Grafana dashboard panels based on user requests by combining the Grafana API and InfluxDB queries. It can create combined visualizations showing traffic from multiple network elements, apply time-range filters, and respond to follow-up requests such as zooming into a specific time window.
- **Sustainability Agent:** Accesses power-monitoring and energy-consumption data from the smart PDU and other sources. It can generate combined Grafana panels correlating network traffic with power consumption of specific devices, such as showing the relationship between ONT receive traffic and camera power consumption.

Integration Layer

The integration layer provides the connectivity between the agent layer and the underlying target systems via APIs and management protocols. The target systems include the OLT NETCONF server for device configuration, the InfluxDB database for telemetry storage, the Grafana server for dashboard management, the NetBox inventory system for topology data, and the power monitoring infrastructure for energy metrics. All communication occurs through well-defined APIs, enabling the architecture to support easy extension by registering new agents for additional management functions.

Demonstrated Capabilities

The PoC successfully demonstrated the following capabilities:

Inventory and Topology Queries (Inventory Agent)

The assistant was able to process natural-language queries to retrieve inventory and topology information from the NetBox system. In the demonstrated scenario, a user asked "what is connected to the OLT on port 0.1.0?" and the assistant, routed to the Inventory Agent, returned a structured response listing the connected ONUs (ONU 4, ONU 5, ONU 6) along with a human-readable summary. The user then issued a context-aware follow-up query "which end devices are connected to this port?" and the assistant correctly resolved the reference to the previously mentioned port and returned the end-device mappings: Inspection Cam A on ONU 4, Robot Arm on ONU 5, and Inspection Cam B on ONU 6.

Telemetry Database Queries (Database Agent)

The Database Agent demonstrated the ability to query the InfluxDB time-series database using natural language. When asked "Which measurements do we have in the database?", the agent returned a complete listing of available measurements including pm_olt_traffic (OLT port traffic), pm_ont_traffic (ONT traffic), pdu_telemetry_cloud (PDU power from cloud), and pdu_telemetry_vis (PDU power from visualization). A follow-up query requesting the fields of a specific measurement returned the full set of telemetry fields such as portRxBytes, portTxRate, portRxPeakRate, and others.

© AIOTI. All rights reserved.

Grafana Dashboard Generation (Analytics Agent)

The Analytics Agent successfully created Grafana dashboard panels based on natural-language requests. In the demonstrated example, a user requested "create a combined panel for received (Rx) traffic of the OLT on port 0.1.0 and the Rx traffic of the ONTs on port 0.1.0.1 and 0.1.0.3 over the last 24 hours." The agent generated the appropriate InfluxDB queries and configured a Grafana panel showing the combined receive traffic for the specified OLT port and ONTs over the requested time window. The panel was made available for direct viewing via an embedded link. A subsequent follow-up request "Zoom to 16:00 and 17:00" was handled in context, and the assistant re-rendered the panel with the narrowed time range, demonstrating context-aware request handling across conversational turns.

Device Configuration via NETCONF (Configuration Agent)

The Configuration Agent demonstrated both read and write operations on the OLT via NETCONF/YANG. When asked "show me the current telemetry subscriptions of the OLT," the agent retrieved the complete list of six telemetry subscriptions, including their sensor groups, sample intervals, and destination addresses. The user then issued a configuration change request: "change the sub01 sample interval to 10000 ms." The agent sent the corresponding NETCONF edit-config operation to the OLT and confirmed the successful update. The effect of the configuration change was independently verified by plotting the OLT port receive traffic before and after the change, where the altered sampling interval was clearly visible in the time-series data.

Power and Energy Monitoring (Sustainability Agent)

The Sustainability Agent accessed power-monitoring data from the smart PDU infrastructure. In the demonstrated scenario, the user asked to "create a combined plot for the Rx traffic of the ONT at port 0.1.0.1 and the power consumption of camera A over the last 24 hours." The agent generated a Grafana panel correlating the ONT receive traffic with the camera power consumption, showing both metrics on a shared time axis. This capability validates the extensibility claim of the PoC, as the Sustainability Agent was added as a new agent to the existing framework to demonstrate that additional management domains can be integrated without modifying the core architecture.

On-Premise Deployment and Data Privacy

All LLM components, including the orchestration layer, individual agents, and the knowledge base, are hosted on dedicated servers within the F5G OpenLab environment at Fraunhofer HHI in Berlin. No data leaves the premises and no external cloud services are used for inference or model hosting. This fully on-premise deployment ensures complete data privacy and local processing, meeting the privacy requirements.

Multi-Agent Modularity and Extensibility

The multi-agent architecture demonstrated its modularity through the independent operation of five specialized agents, each handling a distinct management domain [ZAIDOFC25, ZAIDJOCN26]. The extensibility of the architecture was validated by the successful addition of the Sustainability Agent and the Database Agent to the existing framework, which initially comprised only the Configuration, Inventory, and Analytics agents. New agents were integrated by registering them with the orchestration layer and providing the necessary tool definitions. No changes to existing agents or the core orchestration logic were required.

© AIOTI. All rights reserved.

Conclusions

This PoC has demonstrated a working implementation of an LLM-based AI assistant for the unified management of the F5G Network Fabric, deployed on the F5G OpenLab testbed at Fraunhofer HHI in Berlin. The demonstration showed that it is feasible to provide a single natural-language interface to a heterogeneous set of management systems, including NETCONF-based device configuration, telemetry, analytics, inventory, and power monitoring, while keeping all AI processing fully on-premise for data privacy.

The multi-agent architecture proved both modular and extensible: new management domains were integrated by simply registering additional agents without requiring changes to the existing system. This approach has the potential to significantly lower the barrier to network operations by replacing specialized tooling and expert knowledge with conversational interaction.

We invite AIOTI members to consider this AI assistant framework for developing more advanced LLM based PoCs in the area of intelligent and autonomous network management.

References:

[ZAIDOFC25] H. Zaid, P. Safari, B. Shariati, A. Jafari, M. Balanici, and J. K. Fischer, "Multi-Agent Design for LLM-assisted Network Management," in *Optical Fiber Communication Conference and Exhibition (OFC 2025)*, San Francisco, CA, USA, pp. 1–3, 2025.

[ZAIDJOCN26] H. Zaid, B. Shariati, P. Safari, J. K. Fischer, and R. Freund, "Empowering network management with a multi-agent LLM architecture," in *Journal of Optical Communications and Networking (JOCN)*, 18, 568–580 (2026). DOI: 10.1364/JOCN.585043.

[ZAIDOFC26] H. Zaid, B. Shariati, P. Safari, A. Jafari, J. K. Fischer, and R. Freund, "Demonstration of an On-Prem Conversational AI Assistant for Unified Network Operations and Observability," in *Optical Fiber Communication Conference (OFC 2026)*, Los Angeles, CA, USA, paper M3Z.4, 2026. DOI: 10.1364/OFC.2026.M3Z.4.

© AIOTI. All rights reserved.

12. PoC report: Distributed Intelligence with Privacy-Preserving Features for FTTR

Overview

The telecommunications sector has undergone a major shift in recent years, driven by the increasing demand for high-bandwidth and low-latency applications such as real-time video streaming, telemedicine, and online gaming. As the internet transitions toward a more experience-driven era, the convergence of physical and digital environments is enabling a new generation of immersive services, including virtual and augmented reality, tactile technologies, and digital twins. In this context, traditional fiber deployments combined with in-home wireless access are no longer sufficient to meet the stringent quality requirements of modern residential and enterprise users. To address these limitations, the FTTR concept extends conventional architectures by delivering optical connectivity directly into individual rooms within a building. This approach ensures end-to-end optical access and supports high-throughput wireless performance through next-generation technologies such as Wi-Fi 7. An FTTR system consists of two main hardware components: an MFU, which serves as the central control unit, and multiple SFUs, which operate under its coordination. The MFU manages the local network domain, orchestrating the SFUs to bring fiber connectivity as close as possible to end-user devices. This architecture enables ultra-low latency, high reliability, and consistent performance throughout the premises. However, these advantages come with increased operational complexity and higher energy consumption due to the greater number of active components. In addition, data privacy becomes a critical concern, particularly as operators incorporate AI-driven techniques for network automation and optimization, where access to sensitive in-home data is often restricted. Integrating distributed intelligence into FTTR networks provides a promising solution by embedding computational and AI-based decision-making capabilities across network elements and cloud infrastructure. This decentralized approach supports autonomous operation, real-time performance optimization, and energy-efficient management, while maintaining user privacy through localized data processing. In the proposed architecture, monitoring data used for automated reconfiguration remains within the home environment, ensuring compliance with data protection regulations such as the European general data privacy regulations (GDPR). Nevertheless, several challenges must be addressed, including limited monitoring capabilities in current equipment, restricted access to telemetry through available interfaces, and the constrained computational resources of customer-premises devices. This proof of concept demonstrates how these challenges can be overcome by introducing and validating a distributed intelligence framework tailored for FTTR deployments. The resulting architecture represents a step toward more intelligent, energy-efficient, and privacy-aware broadband systems. The proof of concept is implemented on a live FTTR testbed at Fraunhofer HHI, consisting of one MFU and three SFUs distributed across four rooms. To compensate for limited onboard processing capabilities, external computing nodes are connected via Ethernet: Raspberry Pi devices are used for the SFUs, while a laptop is assigned to the MFU to reflect its relatively higher computational capacity. Figure 64 provides a generalized representation of the overall system architecture [SICAOFC26].

© AIOTI. All rights reserved.

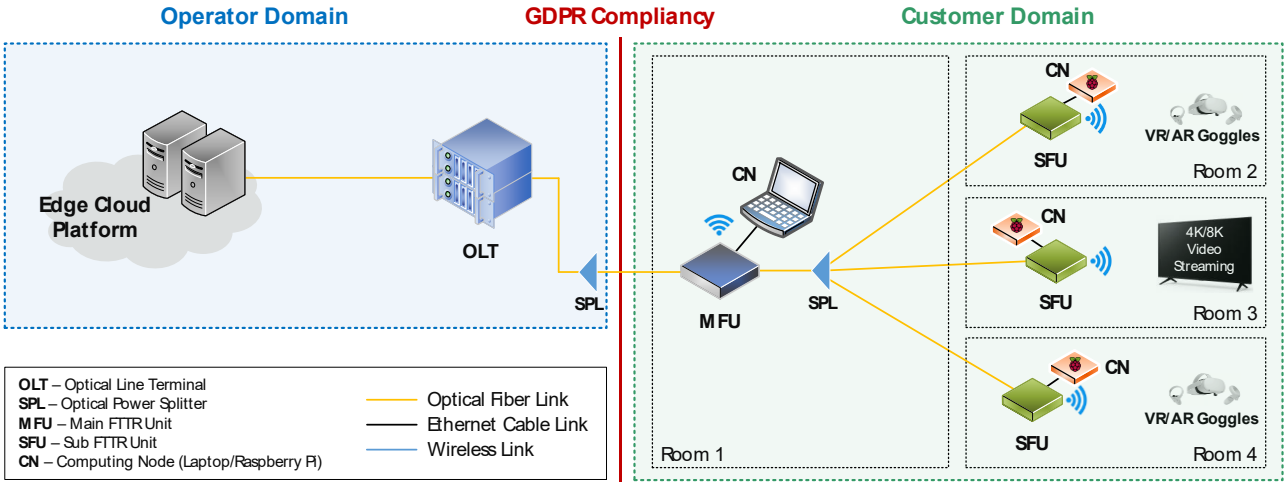


Figure 64: Overall FTTR deployment architecture at HHI.

Description of the Telemetry Framework

The integration of distributed intelligence within an FTTR network relies on efficient telemetry collection, processing, and control mechanisms, all of which are unified through a modular web-based framework. In this PoC, a comprehensive telemetry system is implemented to gather a wide range of key performance indicators, including throughput per interface, resource utilization, connected devices, optical signal levels, and energy consumption, with a temporal resolution of approximately 30 seconds [SICAOFC26]. Data extraction is performed using HTTP-based agents built on Selenium, which parse device graphical interfaces due to the lack of standardized telemetry support in current equipment. In parallel, energy-related metrics are collected via SNMP from a power distribution unit (PDU). Collected telemetry data is stored in distributed data lakes tailored to device capabilities, enabling efficient handling of time-series information. A visual description is available in Figure 65. Controller modules then retrieve relevant metrics and apply closed-loop decision-making to determine whether network elements require reconfiguration.

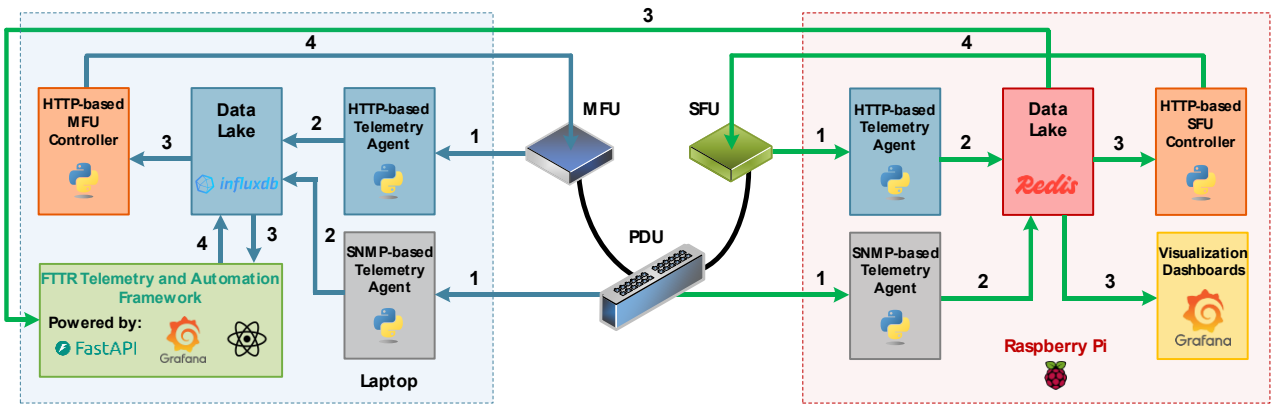


Figure 65: FTTR telemetry framework description. The numbers indicate the sequence of data flow.

These controllers act as distributed intelligence instances, interacting with devices via HTTP requests to dynamically adapt system behaviour. Monitoring is further supported through visualization tools, providing real-time insights into network performance while distinguishing between synchronous control interactions and asynchronous data flows. Building on this telemetry and control pipeline, the DI Framework Web Application serves as the central orchestration and management platform for the entire FTTR domain. The system is implemented as a modular full-stack architecture deployed on the MFU, enabling unified monitoring,

© AIOTI. All rights reserved.

visualization, and configuration of all connected units through a single interface. Its design follows a strict layered approach that separates presentation, application logic, data management, monitoring, and data acquisition, ensuring scalability, maintainability, and robustness. The architecture is composed of five main layers. The presentation layer is realized as a React-based single-page application, responsible solely for user interaction and visualization. The application layer is implemented using FastAPI, which acts as the core system logic, handling data preprocessing, storage operations, API communication, and device configuration. A reverse proxy layer based on NGINX provides a unified access point by exposing a single domain while internally routing requests to backend services and monitoring tools. This approach simplifies deployment, enhances security, and enables seamless embedding of monitoring dashboards without cross-origin constraints. The monitoring and storage layer consists of a time-series database for efficient metric storage and a visualization platform for rendering dashboards. Telemetry data is first processed by the backend and stored, after which the visualization layer directly queries the database to generate dashboards that are integrated into the web interface. This loose coupling between backend logic and visualization improves scalability and reduces system overhead. At the foundation of the system lies the data acquisition layer, implemented through standalone scripts that extract device metrics using Selenium-based GUI parsing. These scripts run on dedicated computing nodes associated with each network element, ensuring that data collection remains independent from core application logic. This separation allows for flexible adaptation to device-specific changes without impacting the overall system architecture. The web application coordinates all components through well-defined interfaces, primarily interacting with centralized and distributed data storage systems. It provides users with a unified view of the entire network, enabling both historical analysis and real-time monitoring without requiring direct access to individual devices. At the same time, sensitive data remains within the local environment, preserving privacy. Beyond monitoring, the framework incorporates autonomous control mechanisms that optimize network operation. For example, underutilized units can be dynamically transitioned into low-power states by reducing wireless transmission power, improving energy efficiency while maintaining service availability. Overall, the combined telemetry pipeline and web-based DI framework establish a cohesive, scalable, and privacy-aware solution for FTTR networks. By integrating distributed intelligence, modular system design, and closed-loop automation, the architecture enables efficient, adaptive, and secure management of next-generation broadband environments. Figure 66 provides an example workflow of the observe-analyse-act framework operation.

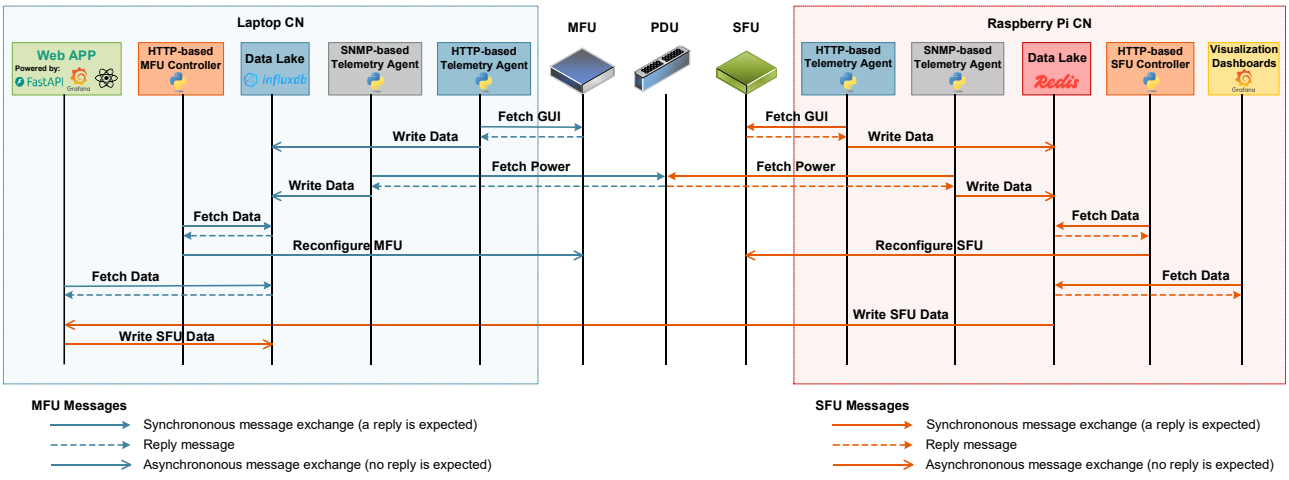


Figure 66: FTTR framework example workflow.

© AIOTI. All rights reserved.

Results and Analysis

The DI framework's web application, presented in Figure 67, provides shortcut tabs that enable quick access to the FTTR monitoring and configuration dashboards, eliminating the need to log in separately onto each FTTR unit. By selecting any of the FTTR Monitoring tabs, multiple Grafana dashboards can be launched directly within the web application. These dashboards present both real-time and historical trends of key operational parameters, including Tx and Rx optical power, Wi-Fi and Ethernet traffic, energy and power consumption, operating temperature, CPU and memory usage (an example is shown in Figure 68). The DI Framework also provides a configurable set of operational parameters for both the MFU and the SFUs. This parameter set is extensible and can be tailored to specific use cases or end-user requirements, as shown in Figure 69 and Figure 70. Such flexibility enables targeted configuration and fine-grained control of the FTTR infrastructure.

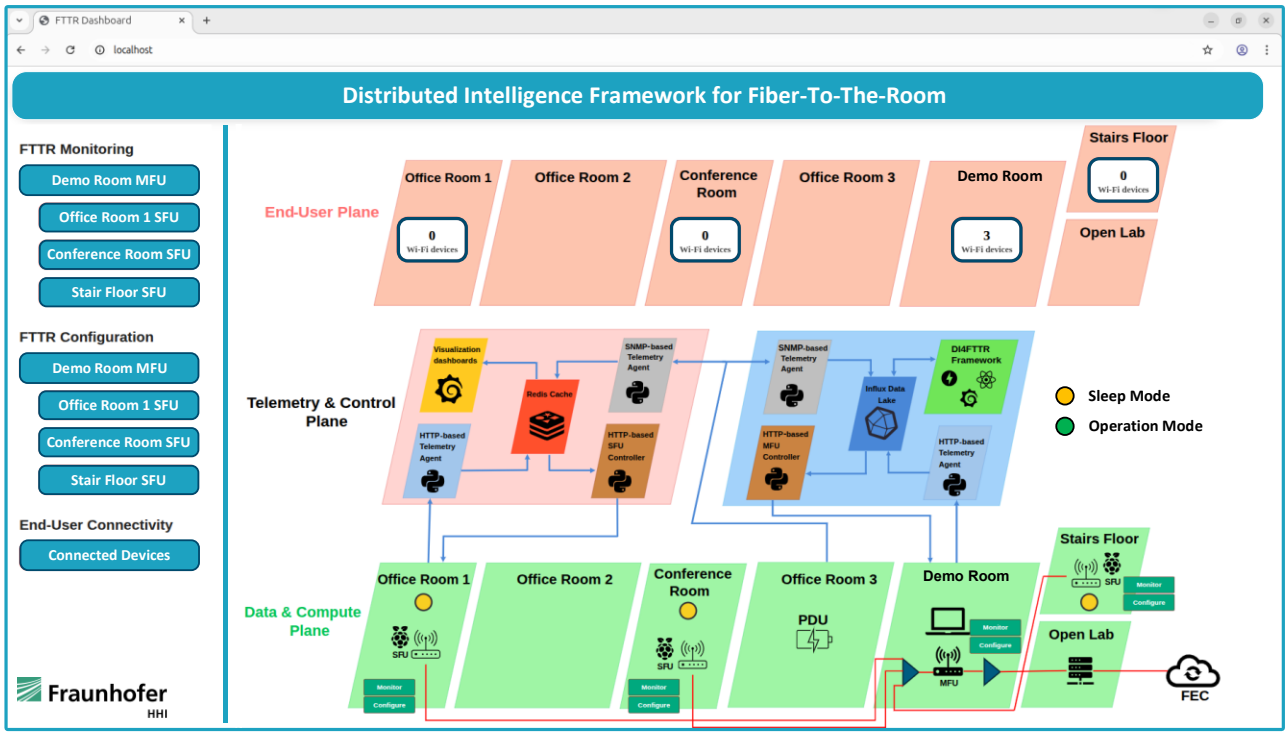


Figure 67: Developed distributed intelligence framework for FTTR monitoring and automation.

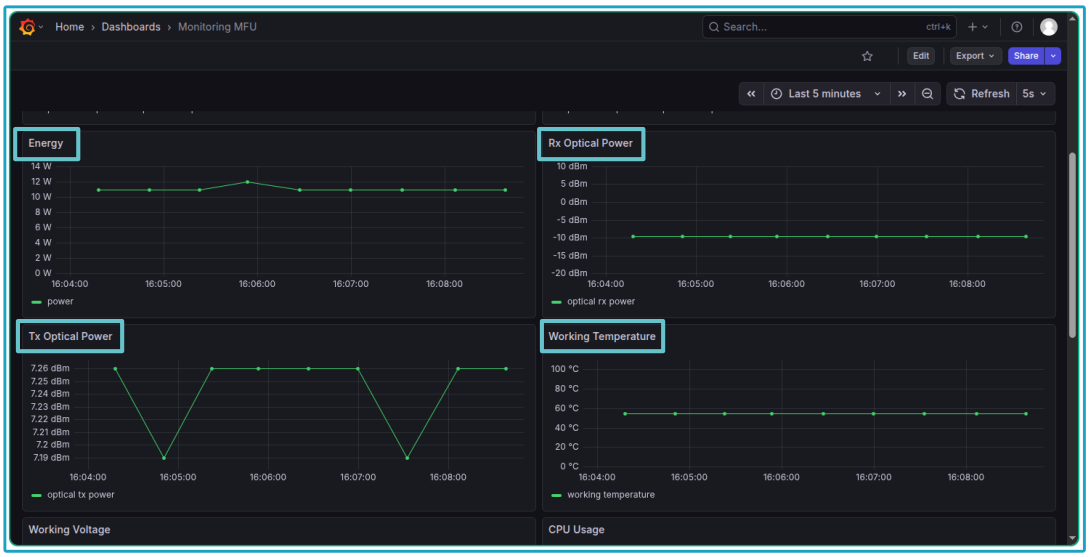


Figure 68: A set of Grafana dashboards monitoring the operation of the SFU.

© AIOTI. All rights reserved.

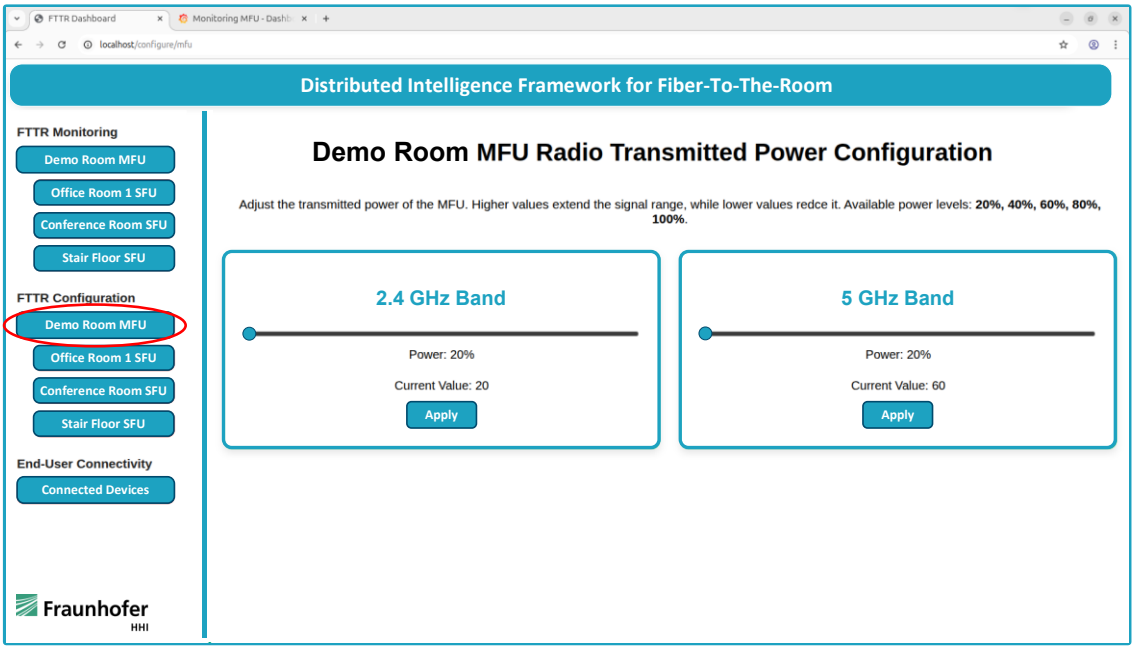


Figure 69: The MFU menu for Wi-Fi operation mode configuration.

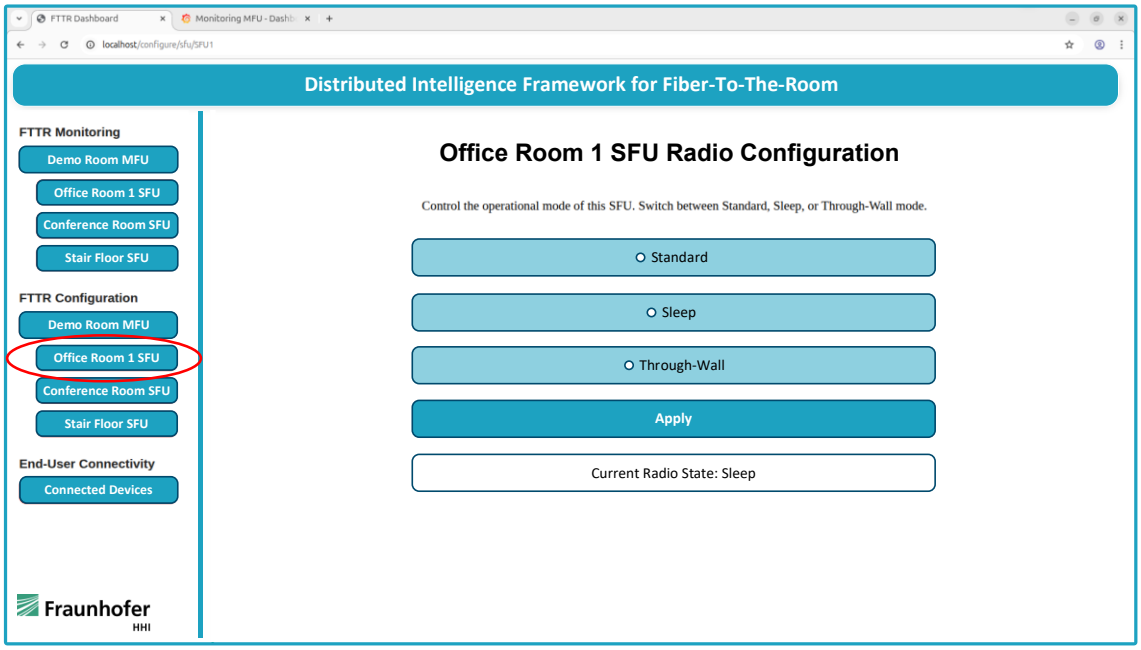


Figure 70: The SFU menu for Wi-Fi operation mode configuration.

In addition, the framework supports end-user device localization within the FTTR domain. It displays the MAC addresses of all devices currently served by the MFU or SFUs, along with a clear indication of which FTTR unit each device is connected to (Figure 71). The monitoring functionality further identifies the access technology used by each device, distinguishing whether the connection is established via Wi-Fi or Ethernet – a capability that is currently available only on the MFU.

© AIOTI. All rights reserved.

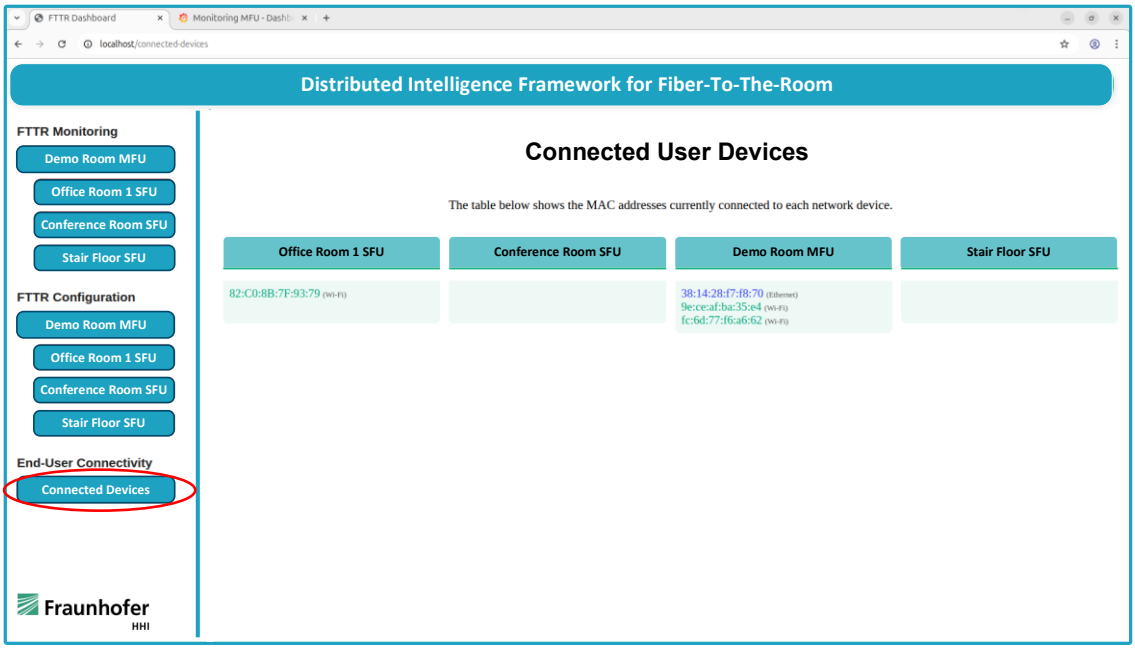
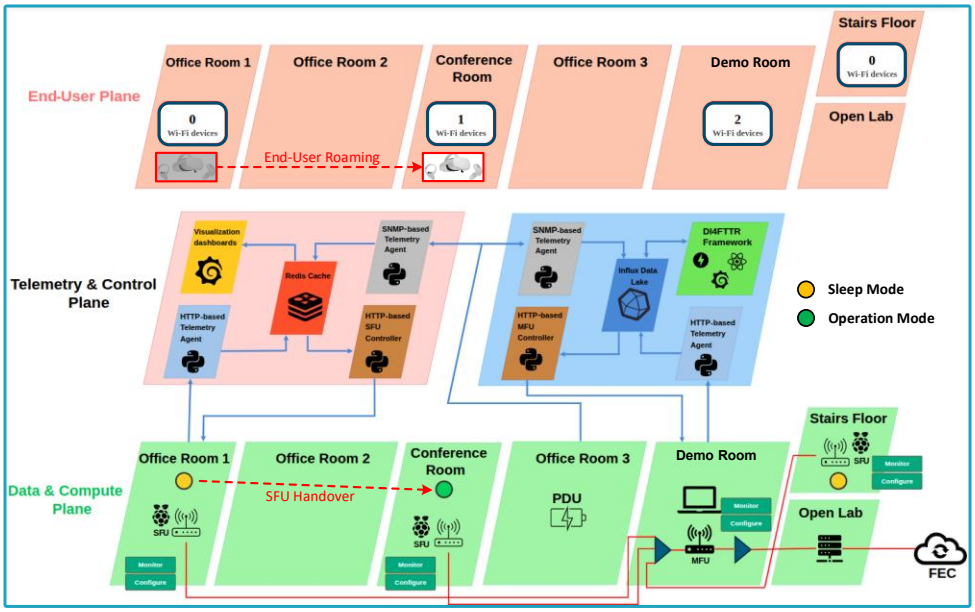


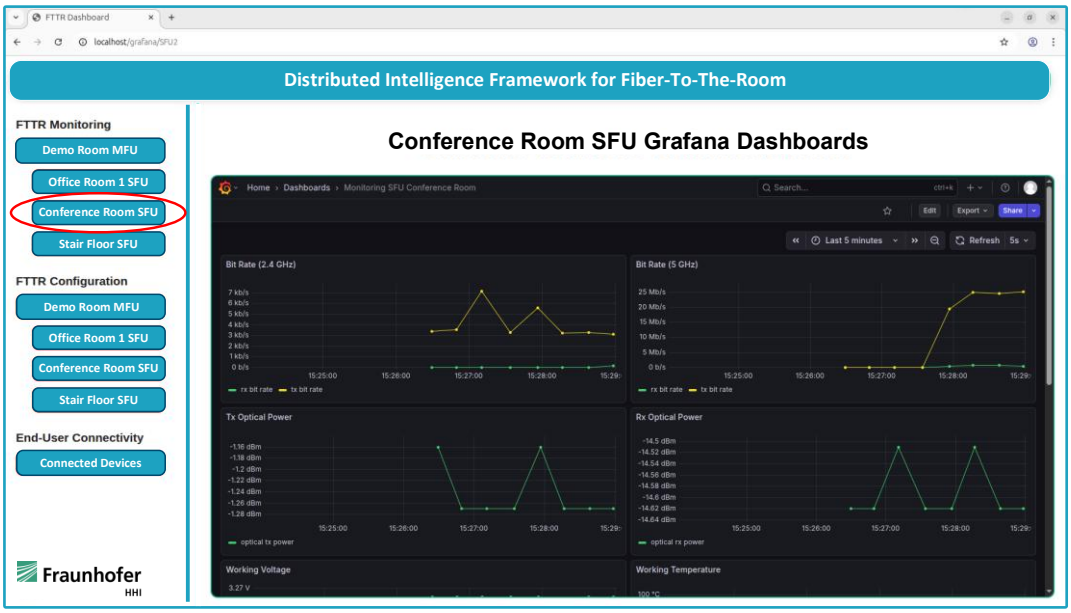
Figure 71: End-user device identification within the FTTR domain.

We demonstrate the operation of the DI Framework for real-time traffic monitoring at both the MFU and the SFUs, while two distinct services are sequentially delivered within the FTTR domain: 4K cinematic video streaming using two different applications (Bigscreen and Virtual Desktop Streamer), followed by immersive VR gaming. During the PoC demonstration, an end user equipped with VR head-mounted goggles moves across multiple rooms within the FTTR coverage area. This mobility scenario highlights seamless roaming and handover between the MFU and SFUs, while maintaining service continuity without any perceptible interruption. In parallel, the DI Framework dynamically activates an SFU when the user enters its coverage area and automatically transitions it into a low-power sleep mode once the user leaves its domain.

The demonstration starts in the Demo Room, which is covered by the MFU. From there, the user moves through several rooms served by different SFUs. Throughout this process, the DI Framework continuously visualizes traffic evolution at both the currently serving SFU and the MFU, effectively illustrating uninterrupted service delivery and smooth traffic migration during user mobility (Figure 72 and Figure 73).

© AIOTI. All rights reserved.





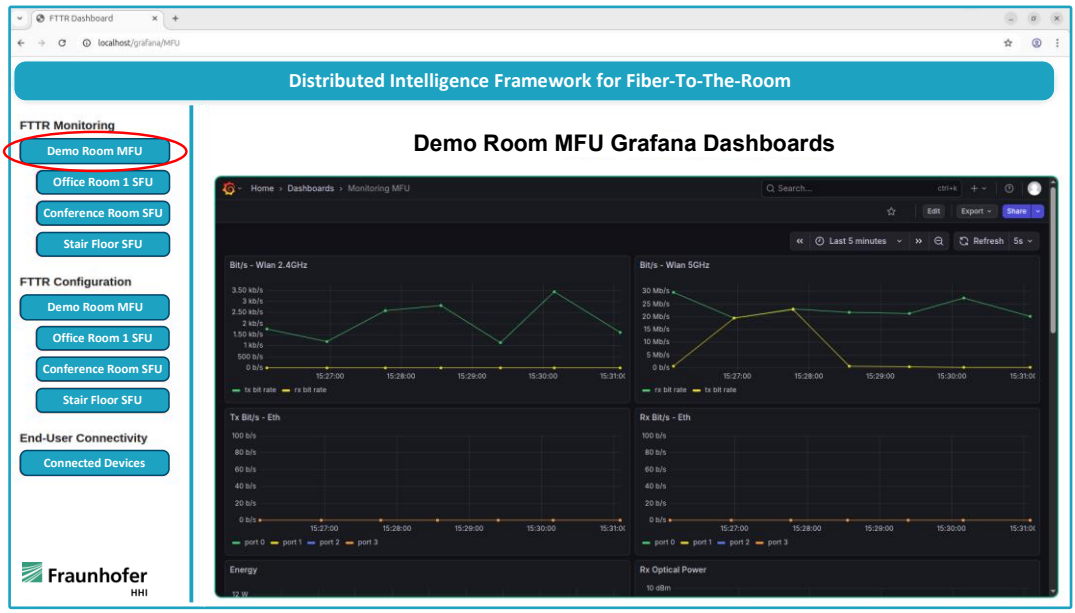
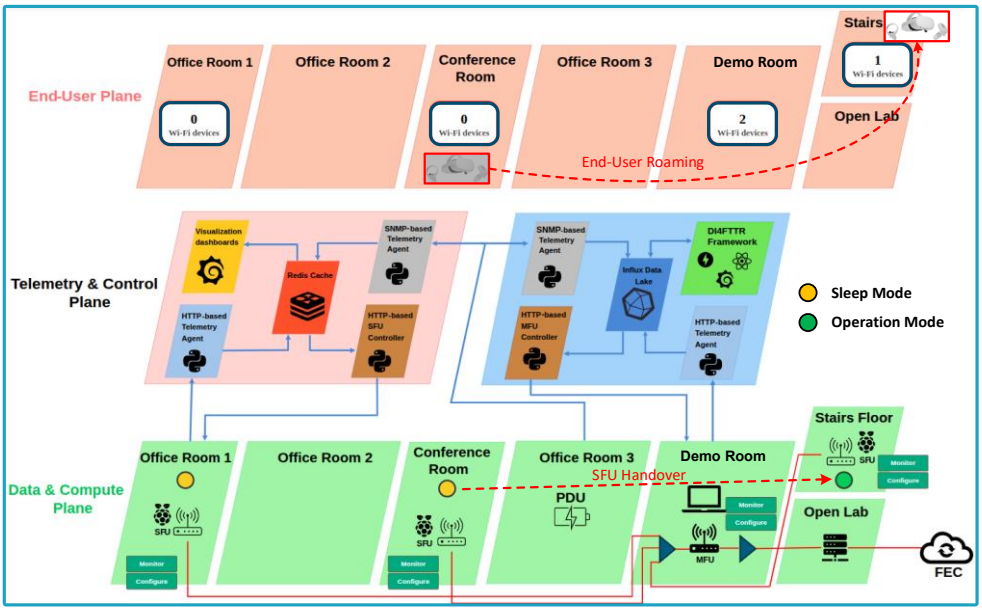
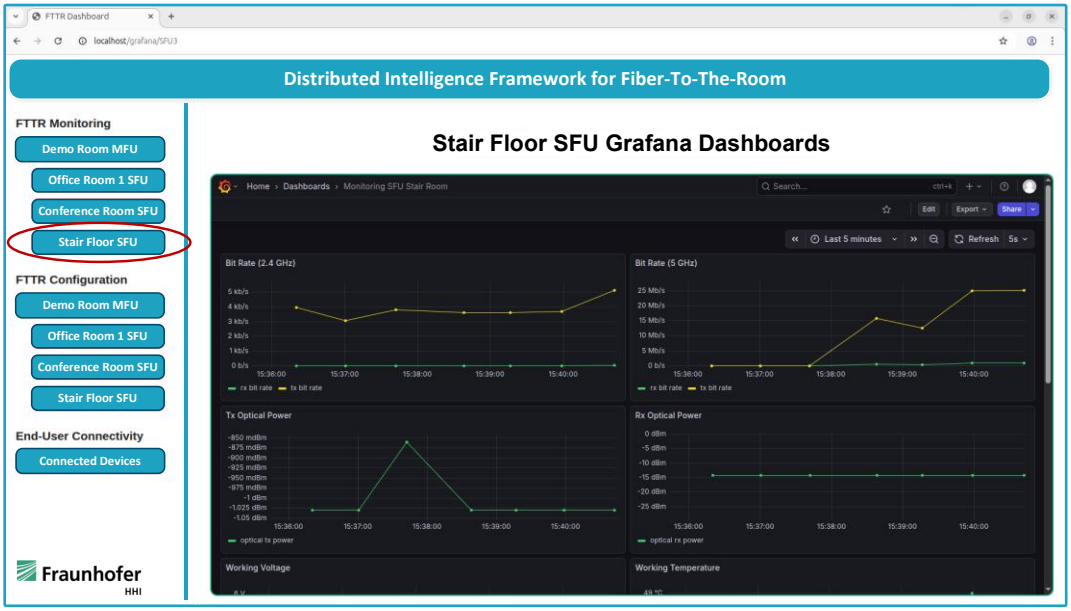


Figure 72: Roaming and handover of the end-user from Office Room 1 SFU to the Conference Room SFU during a Virtual Desktop 4K video streaming service, and the corresponding traffic evolutions on the SFU and MFU.

© AIOTI. All rights reserved.





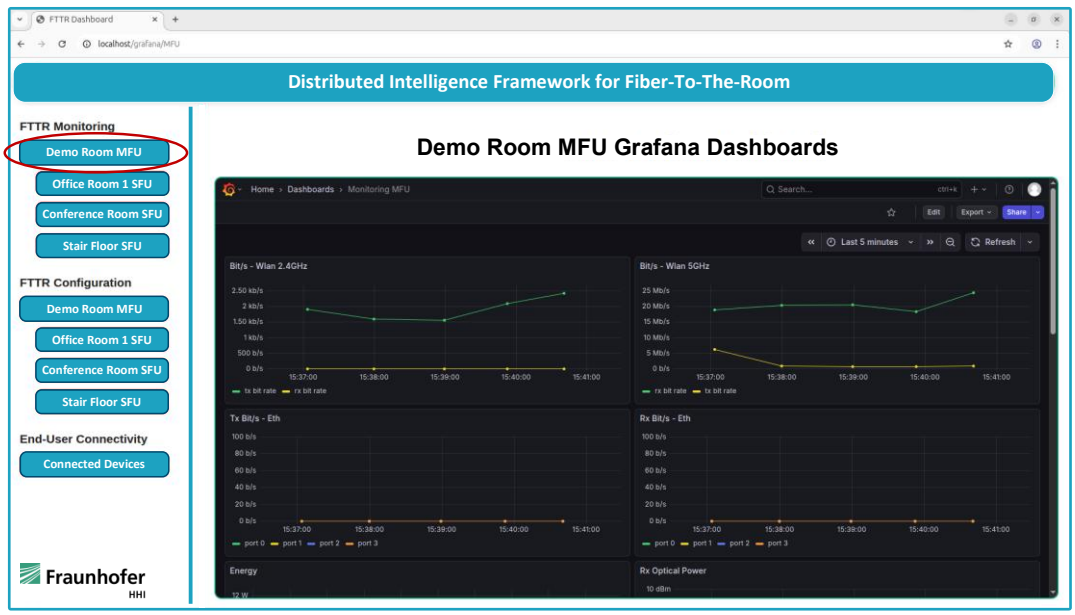


Figure 73: Roaming and handover of the end-user from Conference Room SFU to the Stair Floor SFU during a Virtual Desktop 4K video streaming service, and the corresponding traffic evolutions on the SFU and MFU units.

© AIOTI. All rights reserved.

To conclude, we present a benchmark analysis of FTTR service delivery performance in comparison with FTTH. The main KPIs considered include end-to-end latency and jitter of streamed traffic, perceived end-user experience in terms of service smoothness, and overall wireless signal quality and coverage provided by the access points (APs) namely the ONU in the FTTH setup and the MFU/SFUs in the FTTR setup, across all rooms of the office environment. To ensure a fair comparison, the FTTH ONU and the FTTR MFU are co-located in the Show Room, allowing the FTTH ONU to also cover the staircase floor. In addition, the FTTH ONU AP transmitted power is configured 100% to maximize coverage across the office area. Performance is evaluated for both representative applications previously analysed 4K video streaming and immersive VR gaming, across all four rooms of the office environment. To evaluate system performance, we first deployed a 4K video streaming service using the Virtual Desktop Streamer application on a laptop acting as the video server, generating an average traffic rate of approximately 30 Mb/s. The server was connected either to the FTTH ONU or to the FTTR MFU, while the video stream was rendered and consumed on VR goggles. In parallel, a continuous ICMP Echo Request/Reply session was established between the video server and the VR headset to measure end-to-end round-trip time (RTT). Based on the measurements, both testbeds exhibited comparable performance for the 4K video streaming use case. The observed latency ranged from 3 ms to 300 ms, while jitter remained within acceptable bounds, averaging 14.4 ms for the FTTH testbed (Figure 74) and 23.4 ms for the FTTR testbed (Figure 75). In both scenarios, latency spikes were observed only during user mobility events, specifically device relocation in the FTTH domain (Figure 74) and AP/SFU handovers in the FTTR domain (Figure 75).

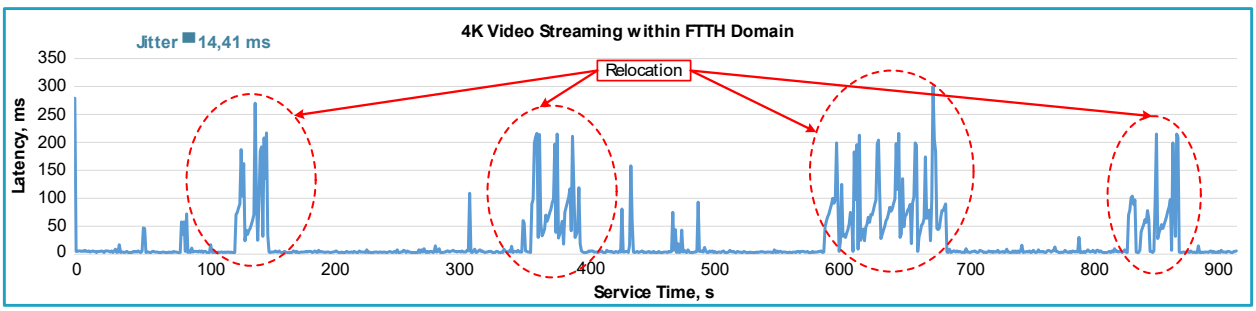


Figure 74: Latency variations and jitter for the 4K video streaming service within the FTTH domain.

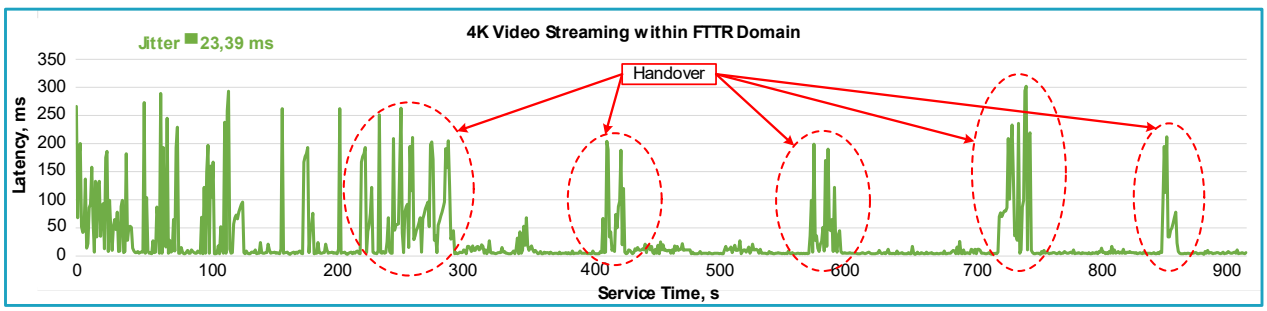


Figure 75: Latency variations and jitter for the 4K video streaming service within the FTTR domain.

From a quality of service (QoS) and user experience perspective, 4K video streaming remained smooth and uninterrupted in both networks, with no perceptible differences in service quality.

For the second use case, immersive VR gaming, a similar experimental setup was employed. A dedicated gaming laptop hosting the VR application acted as the server and was connected in turn to the FTTH ONU and the FTTR MFU. Unlike video streaming, VR gaming generated significantly higher traffic volumes, ranging between 180 Mb/s and 200 Mb/s, as illustrated in Figure 76.

© AIOTI. All rights reserved.

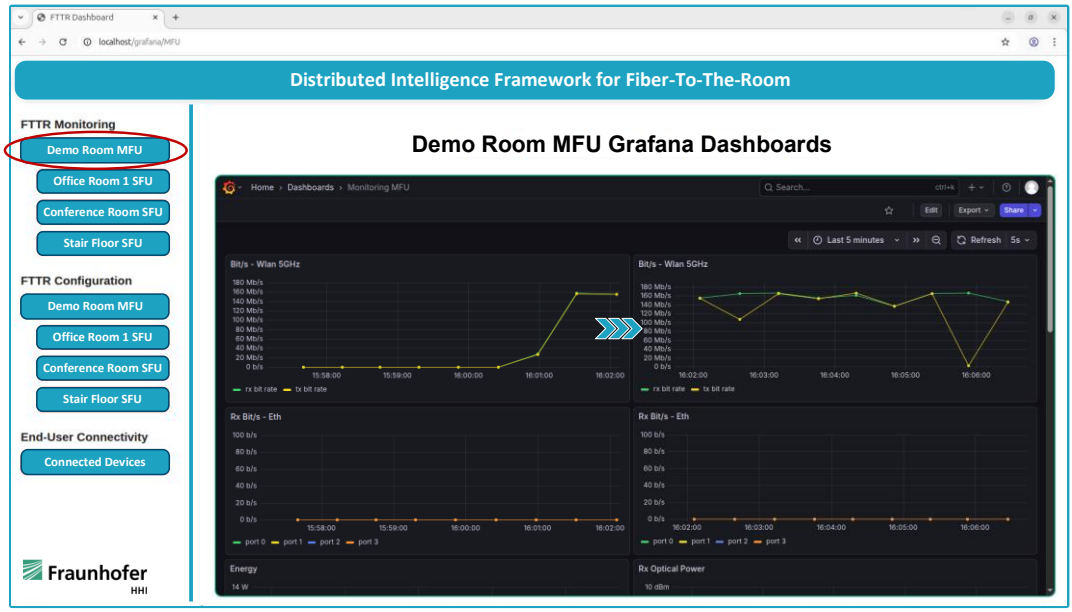


Figure 76: Monitoring of the VR/AR Gaming traffic on the MFU.

In the FTTH scenario, VR gaming performance was smooth only within a 5–7 m radius and required a clear line-of-sight (LOS) between the VR headset and the FTTH ONU access point. As the user moved to more distant locations such as Office Room 1 or the Stairs Floor where walls and doors attenuated wireless propagation, service quality degraded significantly. This degradation manifested as increased latency, frequent interruptions, and, in some cases, temporary session disruption. In contrast, in the FTTR testbed the VR gaming service remained smooth, seamless, and interruption-free across all four rooms, including Office Room 1 and the Stairs Floor. Continuous ICMP measurements further emphasized this difference: in the FTTH setup, more than 50% of Echo Requests timed out (Figure 77), whereas in the FTTR setup no packet loss was observed, achieving a 100% delivery rate (Figure 78). Jitter results further confirm these findings. The FTTH testbed exhibited an average jitter of 111.14 ms during VR gaming (Figure 77), while the FTTR testbed achieved only 11.56 ms (Figure 78), corresponding to an order-of-magnitude improvement. Minor variations observed in FTTR are mainly attributed to SFU handovers and session initialization effects.

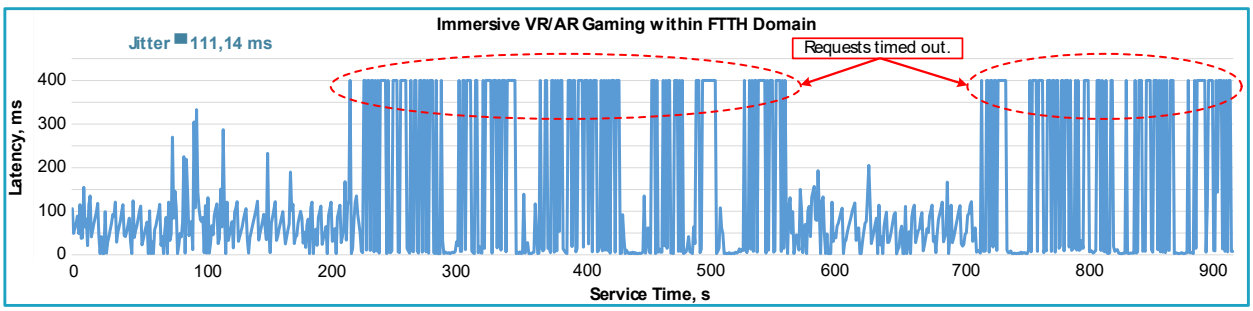


Figure 77: Latency variations and jitter for the VR gaming service within the FTTH domain.

© AIOTI. All rights reserved.

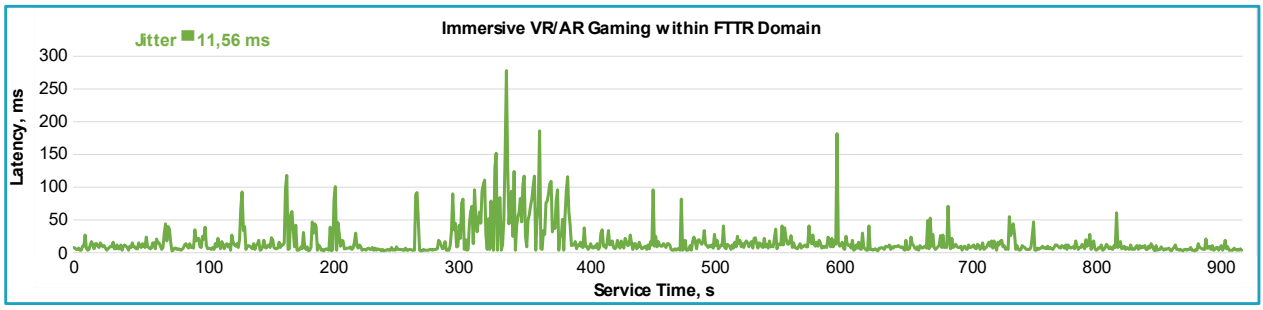


Figure 78: Latency variations and jitter for the VR gaming service in FTTR domain.

Conclusions

We have developed and demonstrated privacy preserving FTTR distributed intelligence framework at HHI premises in Berlin. The demonstration showed that it is possible to collect fine granularity telemetry from the MFU/SFUs to drive self-configuration in FTTR without the data ever leaving the customer's premises. Such results enable the development of more advanced distributed intelligence algorithms once the hardware requirements are met.

[SICAOFC26] M. Sica, M. Balanici, M. R. Raza, B. Shariati, J. K. Fischer, and R. Freund, "Distributed Intelligence Framework with Privacy-Preserving Features for FTTR Network Monitoring and Automation," in *Optical Fiber Communication Conference (OFC 2026)*, Los Angeles, CA, USA, paper M3Z.8, 2026. DOI: https://doi.org/10.1364/OFC.2026.M3Z.8.

© AIOTI. All rights reserved.

13. PoC report: Self-Supervised Incremental Learning for Insect Segmentation and Counting on Resource-Constrained Devices

Overview

This PoC proposes a self-supervised incremental learning (IL) framework for edge/IoT-based insect segmentation and counting, targeting pest control applications in agriculture (specifically monitoring of the invasive brown marmorated stink bug, Halyomorpha halys). The framework splits responsibilities across a three-tier architecture: (i) an offline cloud/server stage where a lightweight U-Net-inspired dual-output model (SemiY-Net) and a MobileNetV2-based triplet-loss classifier are initially trained on a small labeled dataset; (ii) low-power MCU-based edge nodes (OpenMV boards) that perform on-device image capture and inference, and extract and transmit only cropped, insect-candidate regions of the image (rather than full images) to reduce communication load, especially over low-data-rate networks such as LoRa; and (iii) a more capable edge server (Raspberry Pi 5) that reconstructs images from the transferred crops, pseudolabels them using a K-Means-based classifier (with an "uncertain" rejection mechanism, M-K-Means), generates synthetic training samples via replay, and incrementally retrains the deployed models — closing the loop by pushing updated models back to the edge nodes. This eliminates the need for continuous manual labeling and cloud connectivity. Experiments on a 476-image real-world orchard dataset (9 chronological subsets, HALY.ID project) show accuracy comparable to human annotation, ~10% segmentation improvement and threefold reduction in counting error versus a non-retrained baseline, low retraining time (363.3 s) and modest power draw (1.2 A at 5.1 V) on the edge server, and up to 90% reduction in transferred data volume via cropped-region transfer and JPEG compression.

Relevance for Computing Continuum

This PoC is a concrete, quantitatively evaluated demonstration of a three-tier compute continuum spanning cloud/server → edge server → constrained edge/MCU node, applied to a real agricultural IoT deployment. It directly addresses several core computing continuum concerns:

- Task/workload distribution across tiers based on device capability: inference and lightweight image pre-processing (cropping, blob detection) on MCU edge nodes; heavier pseudolabeling, image reconstruction, and model retraining on a more capable edge server (RPi 5); initial model training offline on a cloud/server.

- Data-minimizing communication strategy between tiers: only relevant image regions (~1–4 kB crops instead of ~212 kB full images) are transferred from edge nodes to the edge server, with quantified data-volume reductions (57–90%) and analysis of the accuracy/compression trade-off (JPEG quality 90 down to 1).

- On-the-edge continual/incremental learning without cloud roundtrips or human labeling, addressing data drift and the "cold start" problem of small initial labeled datasets.

- Resource-aware model design and retraining strategy (partial layer freezing, small batch/epoch counts, model compression/quantization for parameter transfer) suitable for battery- or energy-harvesting-powered continuum nodes.

- Empirical characterization of execution time (363.6 s average full cycle) and energy consumption (≈0.68 Wh per retraining cycle) on the edge-server tier, and separately on the MCU tier (2.6 s inference, ≈4.9 J), providing reusable benchmarks for continuum resource provisioning and duty-cycling strategies (e.g., estimated 45 retraining cycles, >1.5 years of operation, from a 37 Wh power bank).

- Reference architecture (Figure 79/Figure 80) [KGR26] diagrams a generic pattern — capture/infer at the extreme edge, reconstruct/label/retrain at the near edge, initial train in

© AIOTI. All rights reserved.

the cloud — that is transferable beyond the insect-monitoring use case to other resource-constrained, intermittently-connected computing-continuum scenarios (e.g., other smart-agriculture, environmental, or industrial visual-monitoring applications).

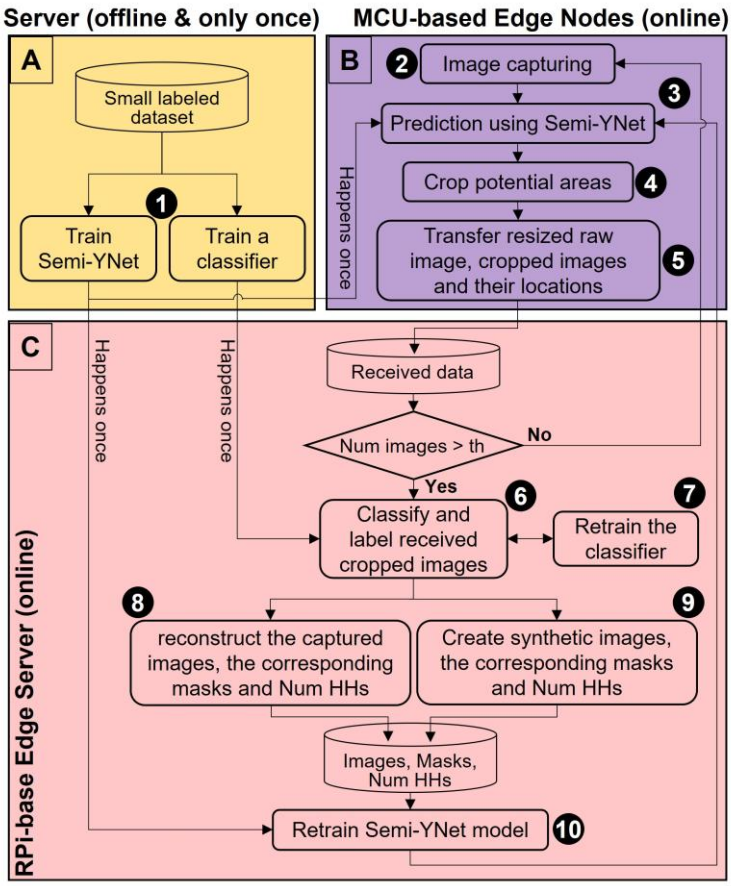


Figure 79: Proposed framework architecture and main components.

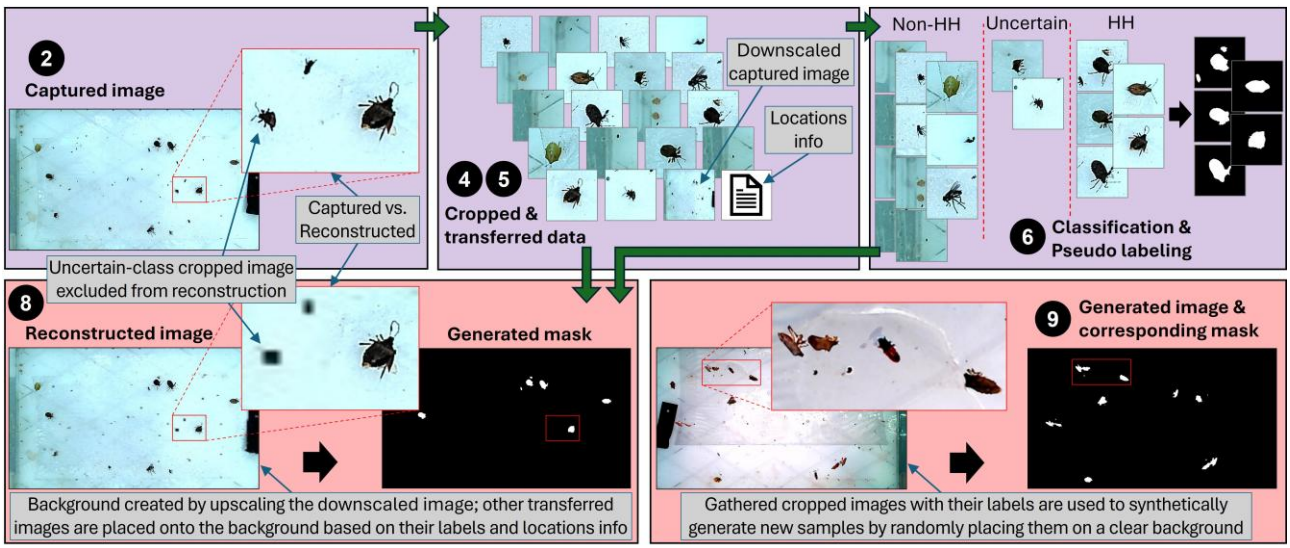


Figure 80: Main components visualization of the proposed framework. Steps 2, 4 and 5 are performed on edge nodes, while the remaining steps are executed on the edge server.

Relation to other AIOTI tasks: Agriculture, Research.

Time frame: The work is already published, and the underlying dataset is publicly available (Zenodo, DOI 10.5281/zenodo.16088064).

© AIOTI. All rights reserved.

References:

[KGR26] A. Kargar, D. Zorbas, M. Gaffney, B. O'Flynn, and S. Tedesco, "Self-Supervised Incremental Learning for Insect Segmentation and Counting on Resource-Constrained Devices," IEEE Internet of Things Journal, vol. 13, no. 2, pp. 2167–2178, 15 January 2026, doi: 10.1109/JIOT.2025.3630062

Dataset: https://doi.org/10.5281/zenodo.16088064

Project: HALY.ID – HALYomorpha Halys IDentification, https://www.haly-id.eu/

© AIOTI. All rights reserved.

14. 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 such factors as the type of service and the load that the ICT infrastructure needs to support over a certain 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.

© AIOTI. All rights reserved.

Energy savings using passive optical networks

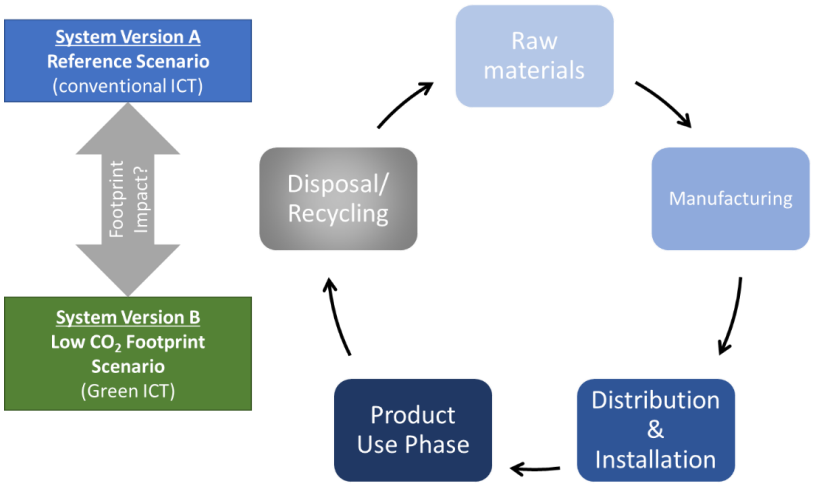


Figure 81: Modelling the carbon footprint of different system versions requires life cycle assessment including detailed technical data as well as measurements in testbeds for product use phase.

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 81), 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 82.

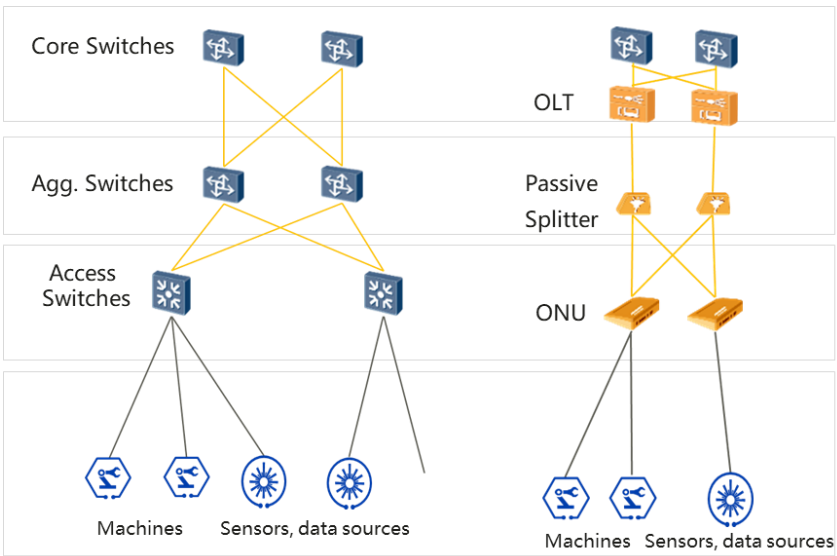


Figure 82: Switched network (left) vs. PON (right) architecture, including optical network units (ONU) and optical line terminal (OLT).

© AIOTI. All rights reserved.

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 83 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.

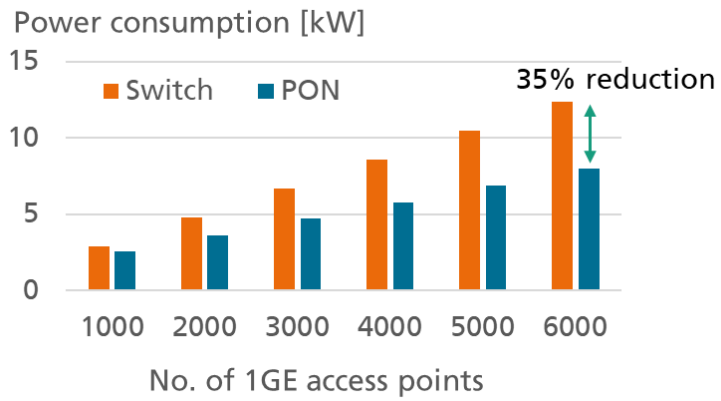


Figure 83: Reduction of power consumption enabled by PON in a scenario with a large number of access points.

© AIOTI. All rights reserved.

15. 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.

© AIOTI. All rights reserved.

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

© AIOTI. All rights reserved.

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 84.

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:

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

- $\boldsymbol{T\_EictBs}_{(l)(ts)}$: 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:

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

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:

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

- $\boldsymbol{T\_EictGr}_{(l)(ts)}$: 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:

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

  - 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.

© AIOTI. All rights reserved.

- Note that the superscripts **M**, **P**, **O**, **D**, shown in the equation terms introduced above and in Figure 84, denote that the carbon emissions calculations are related to the LC phases: Material, Product, Operation, Discard, respectively.

It can be derived that:

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

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

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

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

where:

- $\boldsymbol{EBs_{nict}{}^{M}_{(m)(l)(ts)}}$ 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;
- $\boldsymbol{EBs_{nict}{}^{P}_{(m)(l)(ts)}}$ 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;
- $\boldsymbol{EBs_{nict}{}^{O}_{(m)(l)(ts)}}$ represents carbon emission of each product/component (m) used in in the Baseline scenario, excluding the ICT infrastructure, obtained in the LC Operation phase;
- $\boldsymbol{EBs_{nict}{}^{D}_{(m)(l)(ts)}}$ 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: $\boldsymbol{T\_EGr\_nict_{(l)(ts)}}$; $\boldsymbol{T\_EictBs_{(l)(ts)}}$; $\boldsymbol{T\_EictGr_{(l)(ts)}}$.

<u>Equation for Total ICT Avoided Carbon Emissions</u>

Equation 6: $TAE\_ICT_{(l)(ts)} = T\_ EictBs_{(l)(ts)} - T\_ EictGr_{(l)(ts)}$,

where:

- $\boldsymbol{TAE\_ICT_{(l)(ts)}}$: 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.

© AIOTI. All rights reserved.

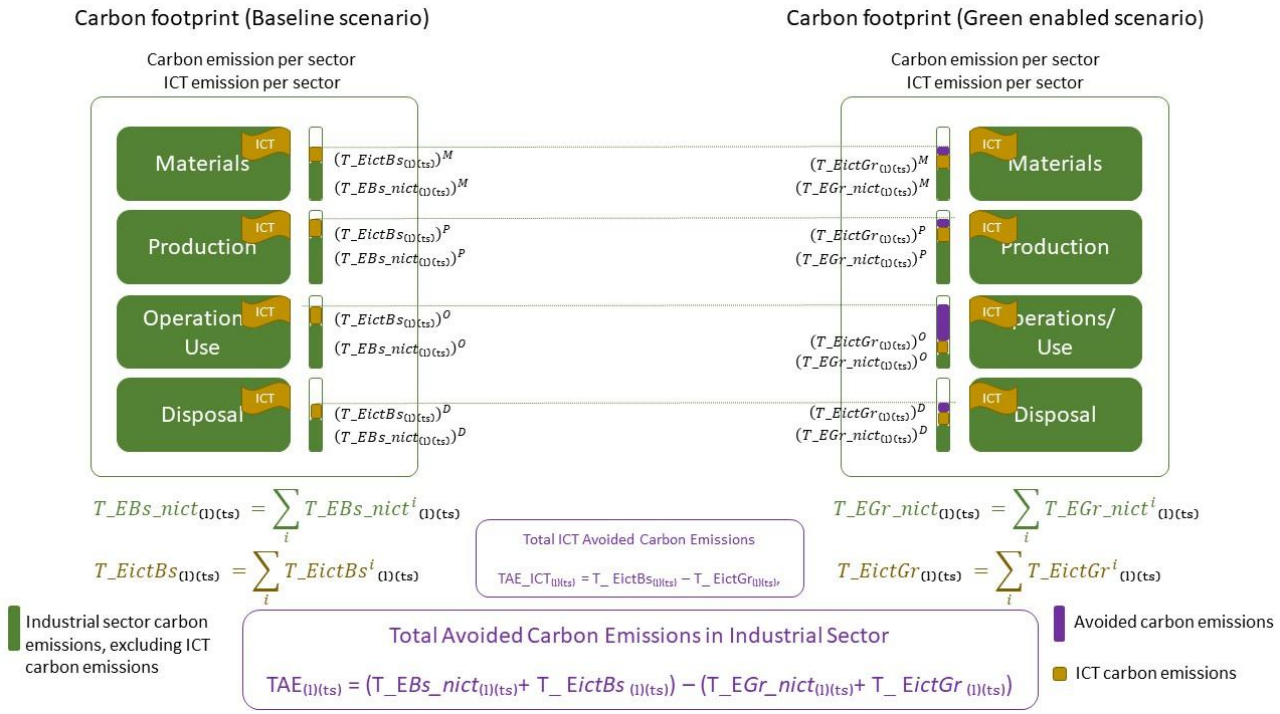


Figure 84: Visualisation of the total avoided carbon emissions, with no circularity support and when ICT is applied as an enabling technology, figure copied from "Alliance for IoT and Edge Computing Innovation 2023".

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.

© AIOTI. All rights reserved.

Contributors

Editor:

Ronald Freund, Fraunhofer HHI

Reviewer:

Valentina Peniche, AIOTI Secretary General

Contributors:

Anagnostis Paraskevopoulos	Fraunhofer HHI
Antonio Lalaguna Lisa	ACISA
Behnam Shariati	Fraunhofer HHI
David Hillerkuss	Huawei
Erwin Schoitsch	AIT Austrian Institute of Technology
George Suciu	BEIA Consult
Georgios Karagiannis	Huawei
Giacomo Tavola	Politecnico di Milano
Johannes Fischer	Fraunhofer HHI
Jun Zhou (James)	Huawei
Liang Zhang	Huawei
Marcus Brunner	Huawei
Muhammad Rehan Raza	Fraunhofer HHI
Mihail Balanici	Fraunhofer HHI
Hussein Zaid	Fraunhofer HHI
Nikos Giannakakos	UniSystems
Ricardo Vitorino	Ubiwhere
Ronald Freund	Fraunhofer HHI
Vasileios Karagiannis	AIT Austrian Institute of Technology
Zbigniew Kopertowski	Orange
Krzysztof Piotrowski	IHP Leibniz Institute for High Performance Microelectronics
Salvatore Tedesco	Tyndall National Institute, University College Cork

© AIOTI. All rights reserved.

Acknowledgements

All rights reserved, Alliance for AI, IoT and Edge Continuum Innovation (AIOTI). The content of this document is provided 'as-is' and for general information purposes only; it does not constitute strategic or any other professional advice. The content or parts thereof may not be complete, accurate or up to date. Notwithstanding anything contained in this document, AIOTI disclaims responsibility (including where AIOTI or any of its officers, members or contractors have been negligent) for any direct or indirect loss, damage, claim, or liability any person, company, organisation or other entity or body may incur as a result, this to the maximum extent permitted by law.

About AIOTI

AIOTI is the multi-stakeholder platform for stimulating AI, IoT and Edge Continuum Innovation in Europe, bringing together small and large companies, academia, researchers, policy makers, end-users and representatives of society in an end-to-end approach. We strive to leverage, share and promote best practices in the AI, IoT and Edge Continuum ecosystems, be a one-stop point of information to our members while proactively addressing key issues and roadblocks for economic growth, acceptance and adoption of the AI, IoT and Edge Continuum Innovation in society. AIOTI contributions go beyond technology and addresses horizontal elements across application domains, such as matchmaking and stimulating cooperation by creating joint research roadmaps, defining policies and driving convergence of standards and interoperability.