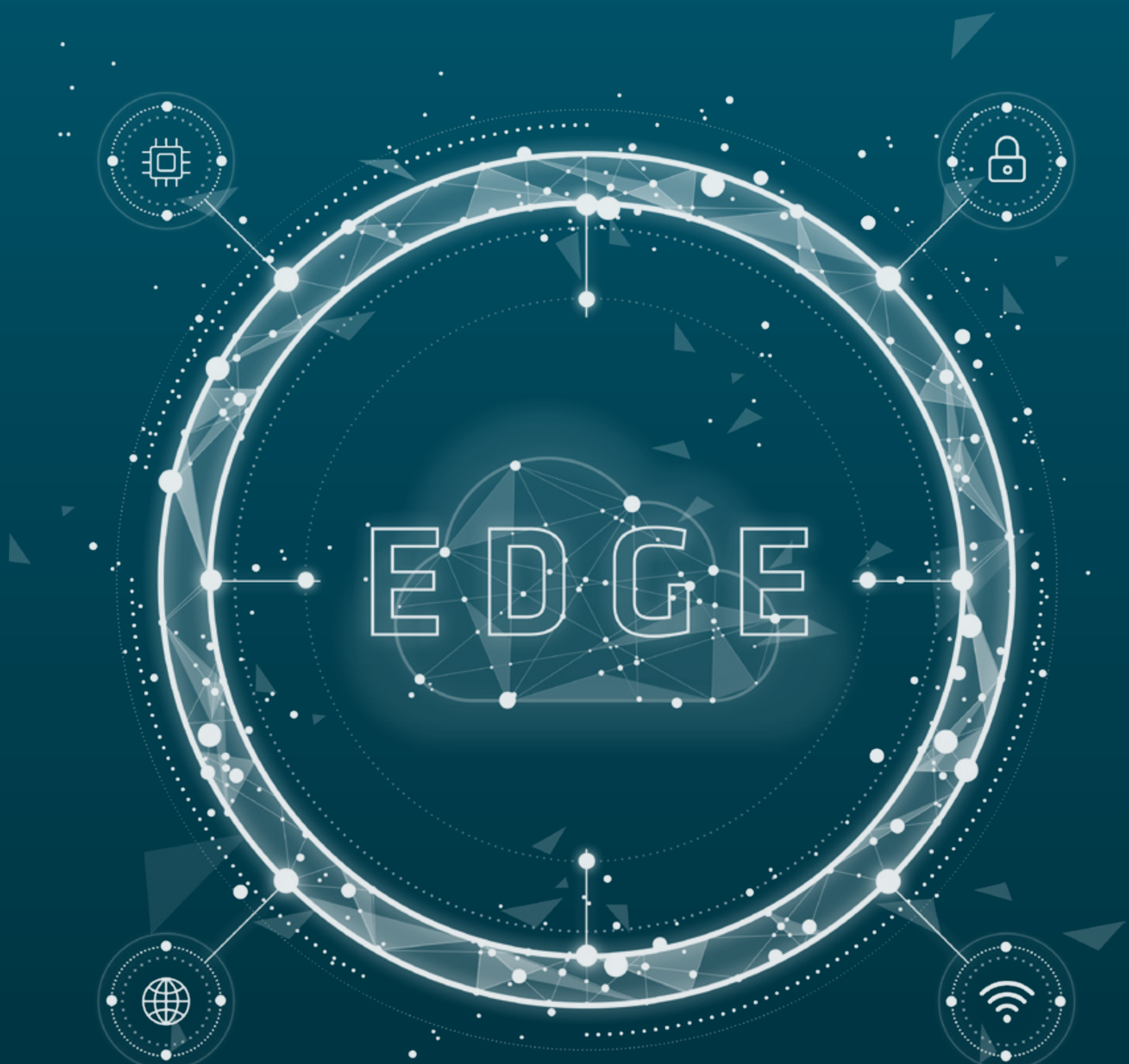


5G-ACIA White Paper

Industrial 5G Edge Computing – Use Cases, Architecture and Deployment

5G Alliance for Connected Industries and Automation

Table of Contents

1 Executive Summary

This document addresses manufacturing use cases that may require or benefit from edge computing, while proposing non-public deployment options for a number of cases and making recommendations on how appropriate network architectures could be built.

We examine how the edge computing features of 5G networks can improve industrial use cases, concluding that they deliver many other benefits besides latency. Use cases can also profit from locality and therefore greater data privacy as well as the ability to scalably run compute-intensive applications such as media processing while consuming less network bandwidth. Solutions are also endowed with enhanced data aggregation capabilities as well as greater resiliency and reliability.

We present examples of use cases that we have examined and show that all of them could be deployed using standard 5G network functions and configuration capabilities. We also discuss the details of various options for configuring the components of an edge computing solution (such as edge runtime infrastructure, network functions, and application functions) to illustrate how the flexibility provided by the standards makes it possible to meet the requirements of use cases in different setups.

About 5G-ACIA

The **5G Alliance for Connected Industries and Automation** (5G-ACIA) was established to serve as the main global forum for addressing, discussing, and evaluating relevant technical, regulatory, and business aspects of 5G for the industrial domain. It embraces the entire ecosystem and all relevant stakeholders, which include but aren't limited to the operational technology industry (industrial automation companies, engineering companies, production system manufacturers, end users, etc.), the information and communication technology industry (chip manufacturers, network infrastructure vendors, mobile network operators, etc.), universities, government agencies, research facilities, and industry associations. 5G-ACIA's overarching goal is to promote the best possible use of industrial 5G while maximizing the usefulness of 5G technology and 5G networks in the industrial domain. This includes ensuring that ongoing 5G standardization and regulatory activities adequately consider relevant interests and requirements and that new developments in 5G are effectively communicated to and understood by manufacturers.

2 Introduction

Edge computing is a distributed information technology architecture in which client data is processed physically close to where it originates, namely at or near a network's periphery. Together with novel network functions, this approach enables new deployment options to help 5G and future networks achieve both very low latency and very high reliability.

Several 5G-ACIA publications have already discussed edge computing in industrial networks. The second edition of the white paper "5G for Connected Industries and Automation" [1], for example, talks about the generic concept of industrial edge computing and introduces the idea of an edge data center (a cloud located close to mobile devices where data is generated). It also describes generic use of network slicing in industrial contexts.

The primary objective of this white paper, which is explained in greater detail in chapter 3, is to analyze factory floor use cases that already rely on or could benefit from edge computing in various non-public network configurations (see [2] for details). There we also discuss the concept of "edge" by describing different edge locations and mapping use cases to them. In chapter 4, we then go further and discuss possible architectures, uses for existing standard functionalities, and possible gaps. Chapter 5 provides recommendations on deployment and configuration options for solution components and analyzes the drawbacks and benefits of each one.

Scope of This White Paper
This white paper addresses use cases described in various 5G-ACIA white papers and 3GPP's Technical Report 22.804 [3], Technical Specification 22.104 [4], and Technical Report 22.832 [5] (for example, on gap analysis and its applicability to use cases and related edge computing requirements) as well as newer edge computing use cases that may even be applicable to 5G-ACIA itself (like for testbeds, proofs of concept, ecosystem trials, etc.). It describes the requirements for various edge locations in different factory domains, identifies the existing use cases that would benefit the most from edge computing for meeting basic requirements, and discusses the advantages of edge computing for the identified use cases. It analyzes implementation, feasibility, and deployment options for the addressed use cases and points out possible gaps in existing specifications and especially those of 3GPP. Finally, it proposes best deployment practices and makes recommendations.

3 Use Case Requirements and Edge Computing Benefits

3.1 Use Cases: Introduction

In this section, we present industrial edge computing use cases that could be implemented with cellular wireless technology (especially 5G) and examine the potential benefits of doing so.

The primary advantages of edge computing for 5G networks are improved latency, greater reliability, local handling of data and therefore greater data privacy, reduced bandwidth requirements as a result of local processing, and facilitated data collection from devices. The scalable processing capacity of local edge clouds makes them ideal for use cases that require very low latency. As an edge site's distance from the communicating devices increases, the area served (e.g. for data aggregation purposes) expands and latency increases accordingly. The reliability of a network service diminishes with increasing distance from the operational technology (OT) premises. It is often required by law for data to be kept at a geographical location that is close enough to keep these problems in check. Other inherent advantages of edge computing are discussed in detail in connection with individual use cases.

In this paper, we examine use cases in which wireless edge devices are used inside or near a factory, executing selected operations locally while sending other data to the nearest edge cloud for speedy processing. Exactly which types of operations can be processed in the cloud depends on the selected deployment option.

For identifying the benefits of edge computing and its applicability to the considered industrial use cases, we focus on the following:

- Computing and aggregating functions that either receive input from various data sources to save bandwidth or carry out analyses locally
- Offloading of heavy computational and coordinating tasks from individual end devices to distributed infrastructure
- Low-latency responses and closed-loop feedback
- Security and data privacy considerations

Some of the presented use cases are based on 3GPP's Technical Report 22.804 [3], Technical Specification 22.104 [4], and Technical Report 22.832 [5] and the 5G-ACIA white paper "Key 5G Use Cases and Requirements" [6] while others are new industrial use cases. They are evaluated to determine their suitability taking advantage of edge computing. The annex of this white paper contains a lengthier example explaining in detail how to identify potential edge computing benefits for use cases in 5G wireless networks. The following examples are therefore kept concise.

3.1.1 Selected Edge Use Cases

Process automation with closed PI/PID control loops
This use case is based on a 3GPP closed-loop control use case for process automation [4]. Several sensors installed in a plant carry out continuous measurements. The resulting data is sent to a controller that decides whether or not to set actuators. A plant can have one or more such controllers. Here the use of a proportional integral (PI) or proportional integral derivative (PI/PID) controller is considered. This use case has very strict requirements in terms of latency and service availability. For the required service performance values for closed-loop control in process automation, see [4], Table A.2.3.1-1. Local data processing, aggregation, and storage are all crucial for the success of this automation use case.

Benefits of edge computing: reduced latency, improved reliability, efficient use of local computing capabilities and available bandwidth, additional aggregated data analytics in the local cloud

Remote access, monitoring, and maintenance based on aggregated sensor data and data aggregation functions
This is an existing 3GPP use case (TS 22.104) [4]. The logic for triggering remote access can be moved from the device to the edge. Data is collected from multiple sensors installed on devices and aggregated to the edge infrastructure where it is processed for instructing a remote control center or diagnostic system to perform the required maintenance. Periodic access to the resulting huge volume of sensor data and rapid processing of the aggregated data are the prerequisites for taking advantage of edge computing capabilities.

Benefits of edge computing: greater efficiency as a result of locally aggregating sensor data, less bandwidth required for device maintenance, remote software upgrades, offloading of heavy local computations, more reliable data due to reduced distance.

3.1.2 New and Evolved Use Cases

Many new and evolved use cases involve factory automation and human-machine interfaces. They are based on existing use cases discussed in the cited sources but have evolved further as a result of our analyses.

Factory automation: mobile robots and automated guided vehicles (AGVs), robot tooling (see [3], section 4.5)
AGVs can be deployed to move products, product components, tools, or raw materials around a factory between storage areas and production lines in accordance with logistical requirements. To execute these complex tasks, AGVs have to be mobile robots with the ability to process information flows on inventory and other things, handle materials, monitor and control activities, recognize images and more. This scenario involves both indoor and outdoor communication.

There are two options in this use case: a) AGVs are operated by a centralized automatic controller based on input received by video streaming and or other methods such LIDAR sensor data, or b) they are controlled by a human operator. Modules for video processing and detection and/or tracking as well as simultaneous localization and mapping (SLAM) functions process incoming video streams and provide information on the robots' environment.

The main prerequisites for AGV control use cases are a device-to-network endpoint latency of about 5 ms, high service reliability with an uplink speed of 3 to 8 Mbit/s for 1080 pixel images, and a remote control bit rate of 100 kbps.

The following additional requirements apply to the robot tooling use case:

- A cyclic data communication service with a minimum cycle time of one millisecond for precise cooperative robotic motion control and between one and 10 milliseconds for machine control
- Jitter < 50% of cycle time

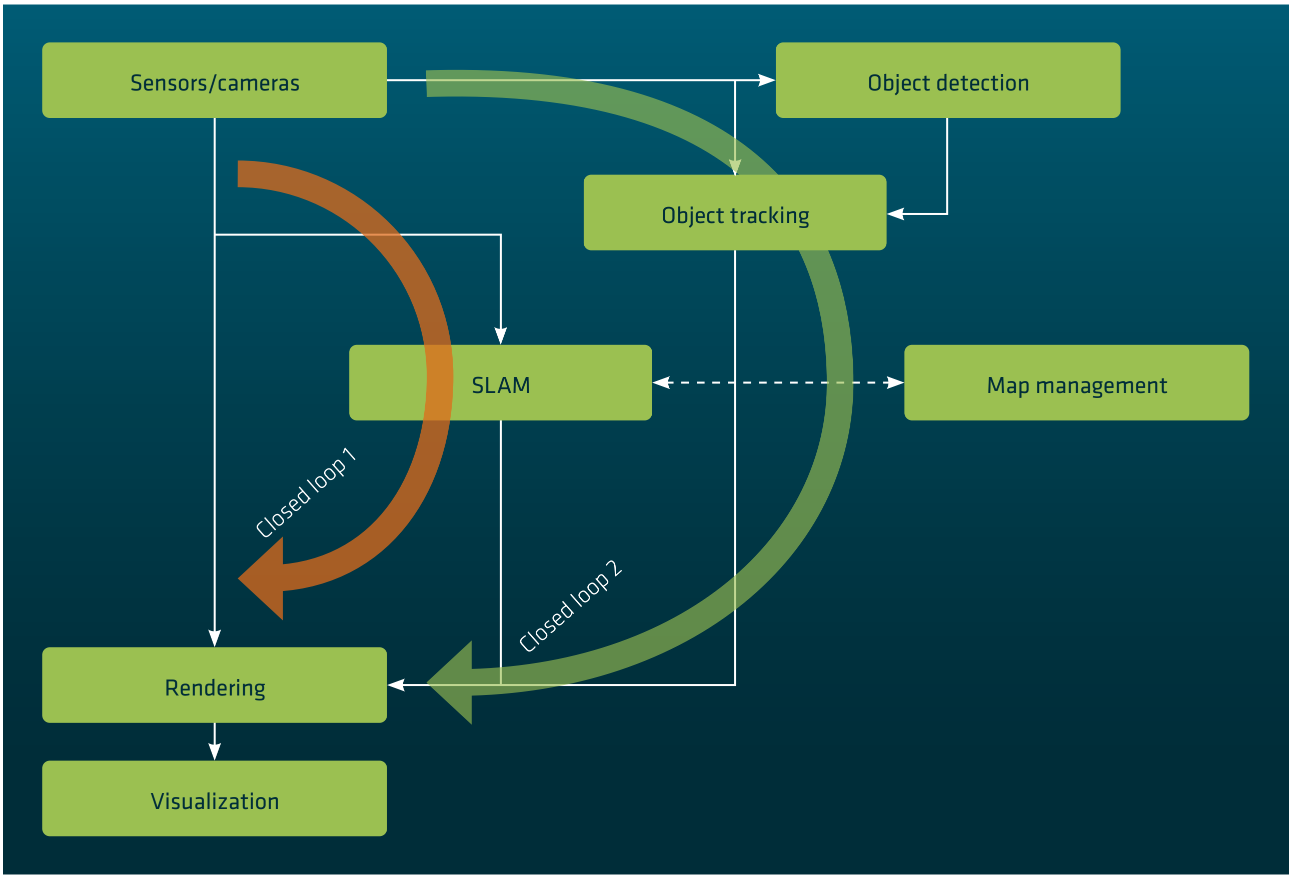
As the number of mobile robots increases, hardware acceleration in the video processing function becomes essential for using LIDAR sensor data and video streaming to control them.

Benefits of edge computing: very low and ultra-low latency for robot control, high reliability (99.999%), real-time video processing for SLAM, on-premises data for meeting privacy and security requirements, and bandwidth savings from locally processing the video data

Factory automation: routing of a fleet of robots and collaborative robot management
This use case is an enhanced version of one described in section 5.11 of [5]. In a smart factory, multiple mobile robots or AGVs carry large and/or heavy workpieces from one place to another. They need to collaborate for safely and smoothly moving large items. This is achieved with an application that controls the drives and motions of the individual mobile robots/AGVs. Usually one of them assumes the role of leader and controls the others.

The edge computing application is like a director who has the task of coordinating the motions of all of the robots. It involves a) capturing the momentary pose, position, and end-of-arm tool angle and position of each robot, b) comparing their status with interpolations to prevent collisions, c) calculating new setpoints, and d) sending these to each robot. Low-latency (<10 ms) communication is required among the director and mobile robots. The robots can communicate with one other via a sidelink mechanism and choose a leader to communicate with the 5G network on their behalf.

Figure 1: Simplified functional architecture for AR



Source: 5G-ACIA / ZVEI e. V.

Table 1: Benefits of edge computing: offloading options

Offloading option	End-to-end latency budget (communication + computation)	Communication latency (uplink and downlink)	Uplink data rate	Downlink data rate
Low offloading	50 ms	1–10 ms	5 Mbit/s	Moderate (augmentation coordinates only)
Moderate/high offloading	20/30 ms	1–10 ms	5–60 Mbit/s	Moderate (augmentation coordinates only)

Source: 5G-ACIA / ZVEI e. V.

Benefits of edge computing: real-time communication between robots and a centralized control function running on the premises

Human-machine interfaces
AR offloading options for industrial use cases

Different AR-enabled use cases can have different requirements and characteristics depending how the AR device user and object move, which of them "owns" the augmentation, the type of AR device, etc. Figure 1 shows a simplified AR functional architecture with the main types of AR functionality, the relationships among them, and the main control loops. Table 1 shows the typical offloading options and corresponding network requirements.

AR-based product quality and augmented field procedure with digital twins
This use case for optimizing product quality involves developing and merging complex digital models to create an idealized digital twin corresponding to the design intention. Streaming data from actual physical systems is then analyzed to ascertain how closely it corresponds. Machine learning is harnessed to detect and compensate for negative factory-specific anomalies in the production environment.

AR-based field procedures of this type can increase both safety and operating efficiency by augmenting field procedures and making them mobile. Such a system is able to:

- guide and assist the user step-by-step through the workflow and record all details of its execution,
- overlay field data as augmented context in an image of the physical area,
- generate up-to-date, accessible information on processes and other important aspects,
- guide manual inputs and trigger actions,
- take pictures and store comments, and
- support both manual and automatic data capture.

The general prerequisites for this use case are low latency (< 20 ms), feedback loop control, and high availability (> 99.999%).

Greater safety for manufacturing workers with AI and XR
Collaboration between humans and machines is essential for achieving flexible production. Specifically, it involves combining humans' ability to learn quickly with that of machines to precisely perform repetitive tasks. In the context of their interactions, it's essential to take steps to ensure the safety of human workers since there is a risk of serious injuries in the event of failures. The environment should therefore be constantly monitored by sensors to analyze and assess safety risks posed by objects that are close to humans.

This use case comprises various approaches for ensuring the safety of humans who interact with machines in the context of an industrial process:

1) Closed-loop safety control
2) Ultralow-latency communication
3) Connection of AGVs, XR devices, and machines via 5G wireless
4) Real-time assessment and mitigation of risks

Benefits of edge computing: real-time control and monitoring in wireless scenarios, ultralow-latency communication, and greater operating efficiency, reliability, and availability

Logistics and warehouse automation: mobile positioning and asset tracking
Products, tools, and assets vary greatly and can be difficult to locate in large, complex deployments. Mobile positioning can help improve this situation, both indoors and outdoors. This use case involves monitoring a large number of items, is calculation-intensive (especially for moving targets), and is typically prone to measurement errors.

5G radio access technology delivers various enhanced parameters for improving positioning accuracy, especially when using time- and angle-based methods. Beamforming may increase measurement accuracy, although it depends on the power received. In these use cases, the edge computing requirements are local processing capabilities, data storage, and application infrastructure.

Benefits of edge computing: low latency for wireless real-time control, greater network reliability, and improved privacy due to local control of data.

Monitoring and maintenance: video surveillance service
In this use case, multiple video surveillance cameras are deployed to monitor some of the most critical process steps in an industrial plant. This adds value by letting users watch the actual production process if it's disrupted. The cameras are in turn monitored by operators and optionally also recorded, and the data they generate is locally stored in the edge.

As a rough estimate, each camera will require an uplink data rate of more than five megabits per second, but the required bandwidth will depend to a large extent on the encoding method used as well as the desired quality.

The primary requirements for this use case are local storage and local processing capabilities.

Benefits of edge computing: proximity of the edge computing application environment to the use case deployment site results in greater reliability, reduces the required network bandwidth, opens up additional low-latency automation options, and can also provide data privacy.

3.2 Mapping Use Cases to Edge Deployment Location Options

Telecoms operators have different options for supporting an edge cloud within their own facilities, on corporate premises, or elsewhere. From a telecommunications network perspective, it's possible to distinguish operational technology (OT) premises, Near Edge (between about one to 100 km from the data source), Far Edge (up to 500 or 1000 km from the data source), and centralized options (usually from the operator's perspective, but "edge" can also be interpreted from a global viewpoint). The distance between the OT premises and the Far or Near Edge isn't fixed, since it depends on the vertical applications, customer needs, quality of service, and service agreements in each case. What's more, an industrial 5G network can be implemented in different ways: it can be a completely isolated private network or integrated in a public network and deployed as either an on-premises private edge or a public edge.

5G-ACIA has also defined an industrial automation communication structure consisting of an OT production domain, an IT enterprise domain, and a public network domain as deployment options for supporting industrial use cases (see section 2 in [8]).

In later sections, we will examine feasible deployment options in greater detail. We will also discuss the capabilities of each deployment option and the kinds of edge computing benefits that they can provide for use cases, as well as the factors that determine the relative simplicity or complexity of a given model.

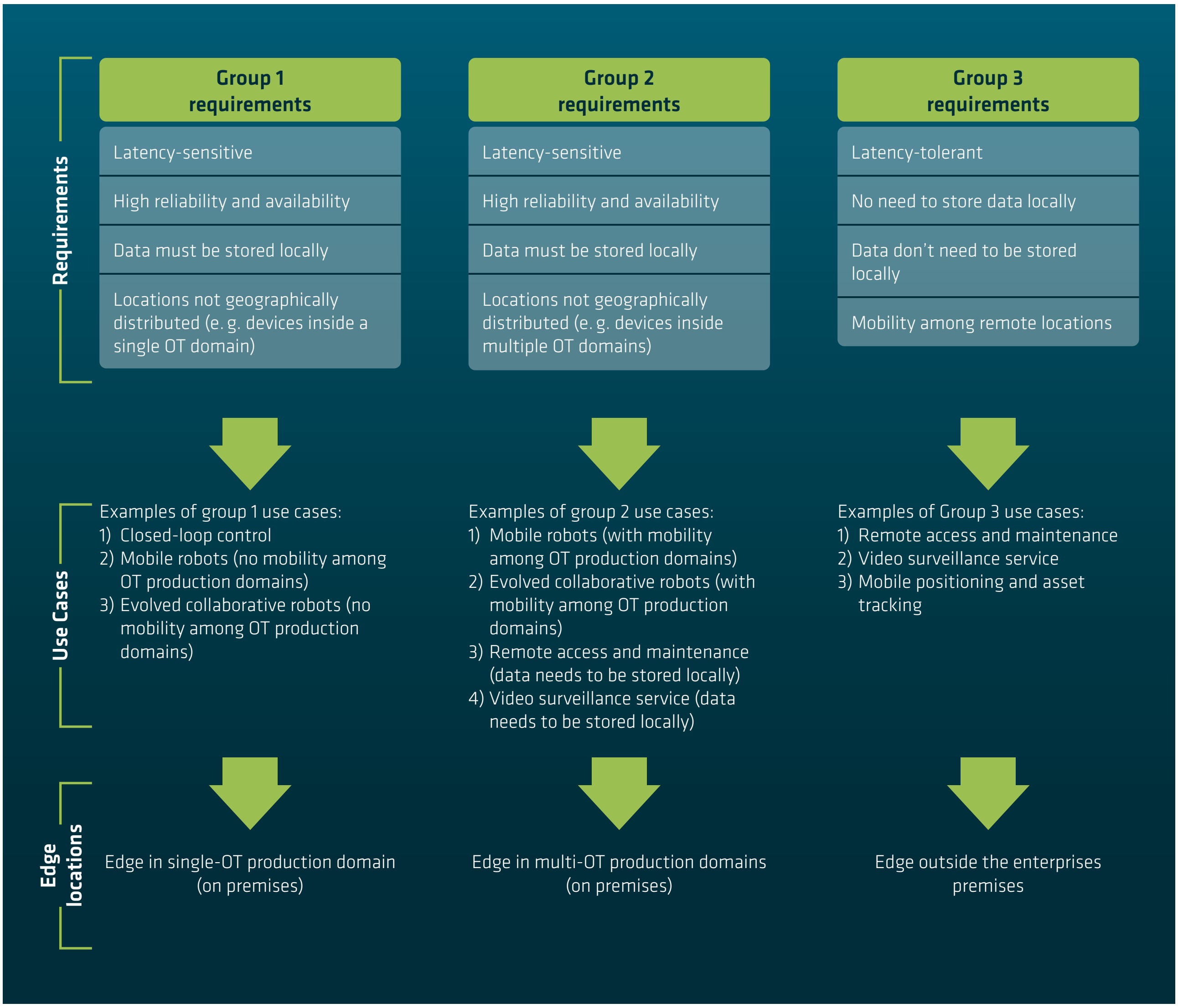
In this section, we map use case requirements to edge node location options (presented in figure 2). This is done while keeping in mind that the edge nodes can be located in any of the three industrial automation domains and that, from a telecommunications operator perspective, the "edge" can be on the OT premises or at a Far Edge, Near Edge, or central location. This results in clearly defined edge deployment options, which we will refer to from here on. Table 2 shows the mappings.

Different use case requirements (figure 2) are divided into three groups.

The first group involves use cases that have very stringent requirements in terms of latency, service reliability, and availability. They are also characterized by the need to store sensitive private information locally within an enterprise. All of their devices and entities are distributed locally within the OT domain. Either there is no need for mobility at all, or else the mobility of devices and entities is limited to that one domain.

The second group contains use cases that are basically similar to those of the first group but involve a shop floor infrastructure with two or more interlinked OT production domains.

Figure 2: Approach for mapping use case requirements to edge location options



Source: 5G-ACIA / ZVEI e. V.

Table 2: Mapping of edge location options to proposed edge node locations from a telecommunications operator's perspective, based on analyses of use case requirements and applying the terminology of the 5G-ACIA white paper "Integration of Industrial Ethernet Networks with 5G Networks" [4]

Deployment options for edge computing

Communication service provider's bird's-eye view	On OT premises			Far edge (approx. 1 to 100 km	Near edge (regional or within city)	Central (from CSP's perspective)
Definition from 5G-ACIA white paper	OT production domain		IT enterprise domain	Public network domain		
Edge node locations proposed in this paper (based on authors' analyses of requirements)	Single OT production domain	Multiple OT production domains		Outside enterprise premise (but nearby)	Central location	

Source: 5G-ACIA / ZVEI e. V.

In other words, devices and entities are distributed locally across multiple domains within the enterprise. This makes it important to support mobility between domains (like, say, a mobile robot traveling from one OT domain to another).

The third group comprises latency-tolerant use cases with devices and entities at geographically distinct locations. It's essential to support mobility among these remote sites to the required extent, but there are no data privacy obligations and no necessity to store information locally within the enterprise.

By way of example, the closed-loop control use case described in [4] falls into the first group. When robots are stationary or their mobility is constrained to a single OT production domain, they also belong to the first group. If mobile robots need to move between OT production domains, however, the use case belongs to the second group. Scenarios involving more evolved, collaborative robots can belong to either the first or the second group depending on the degree of mobility. Remote access and maintenance, video surveillance services, mobile positioning, and asset tracking are all delay-tolerant use cases and therefore normally belong to the third group. If they involve private, OT-sensitive information that is stored locally within the enterprise, however, they fall into the second group instead.

In the first group defined above, the edge is on the premises and contained within a single OT production domain. In the second group the edge is also on the premises, but it serves multi-OT production domains and can be situated (for example) within a larger enterprise-wide IT domain. In the third group, the edge is outside the company premises in the public network domain.

The proposed edge location options correlate to user plane function (UPF) deployment options and the corresponding advantages and drawbacks, which are described in section 5.2.

4 Example Deployments for Use Cases

In this chapter, we analyze various use cases and propose deployment options for different edge node locations (as described in chapter 3) with edge computing architecture components specified by 3GPP, ETSI MEC, and GSMA OPG.

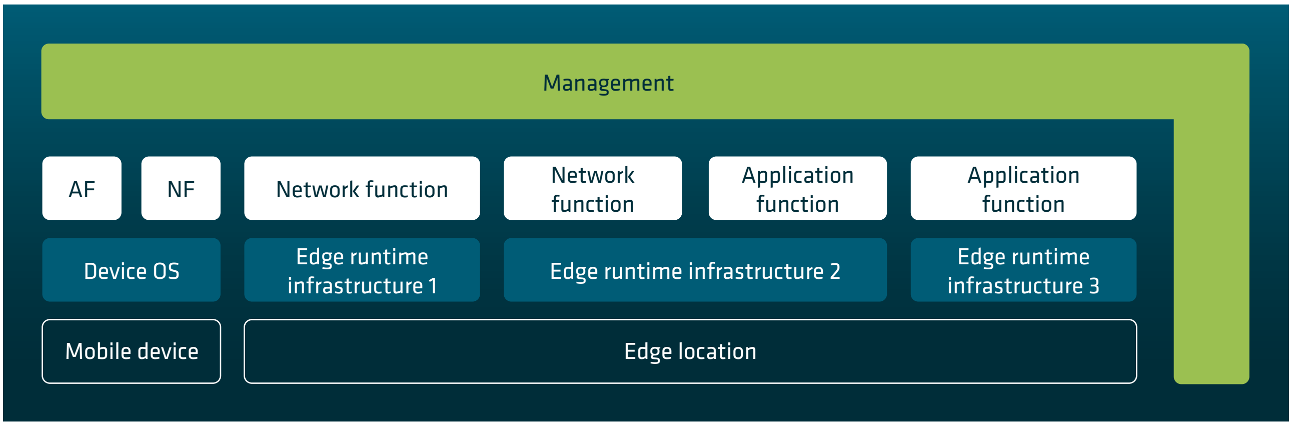
For the sake of simplicity, here we use various high-level terms to describe the components of an edge computing system. Figure 3 shows an example: a full high-level stack solution comprising an edge runtime infrastructure, network functions, application functions, and management of all these elements. The point of this exercise is to highlight the functionality and components that are most relevant to actually deploying a use case.

- The edge location (dark turquoise) can be an "edge in a single OT production domain", an "edge in multiple OT production domains," or an "edge outside the enterprise premises." Despite this, we treat the "central site" and "mobile device" as locations for inserting functions in use cases that require them.
- The edge runtime infrastructure (light turquoise) can be dedicated hardware, the cloud, or any other infrastructure that is capable of running the network functions and application functions. The mobile devices have a special device operating system as the runtime infrastructure for their NF and AF components. This infrastructure may also include required firewalls and network routing functions.
- The network functions (white) are based on a similar concept, namely the virtual network function defined by ETSI NFV. They comprise all functions of the 3GPP network that are required in order for it to work as intended. They don't include the underlying infrastructure or the functions needed to manage and operate it.
- The application functions (white) are typically independent of the communication network infrastructure but required for the use case.
- The management layer (black) isn't discussed in detail in this document. It may comprise multiple components that are presumably able to manage and operate all or parts of an edge computing system.

4.1 Factory Automation: Mobile Robots

Factory automation involves automated control, monitoring, and optimization processes and workflows for a factory's production systems. It typically comprises discrete applications with specific requirements in terms of low, bounded latency, high availability and reliability, cybersecurity, and functional safety. In the following, we examine a specific use subcase, namely that of mobile robots.

Figure 3: A high-level diagram showing the components of an edge computing system



Source: 5G-ACIA / ZVEI e. V.

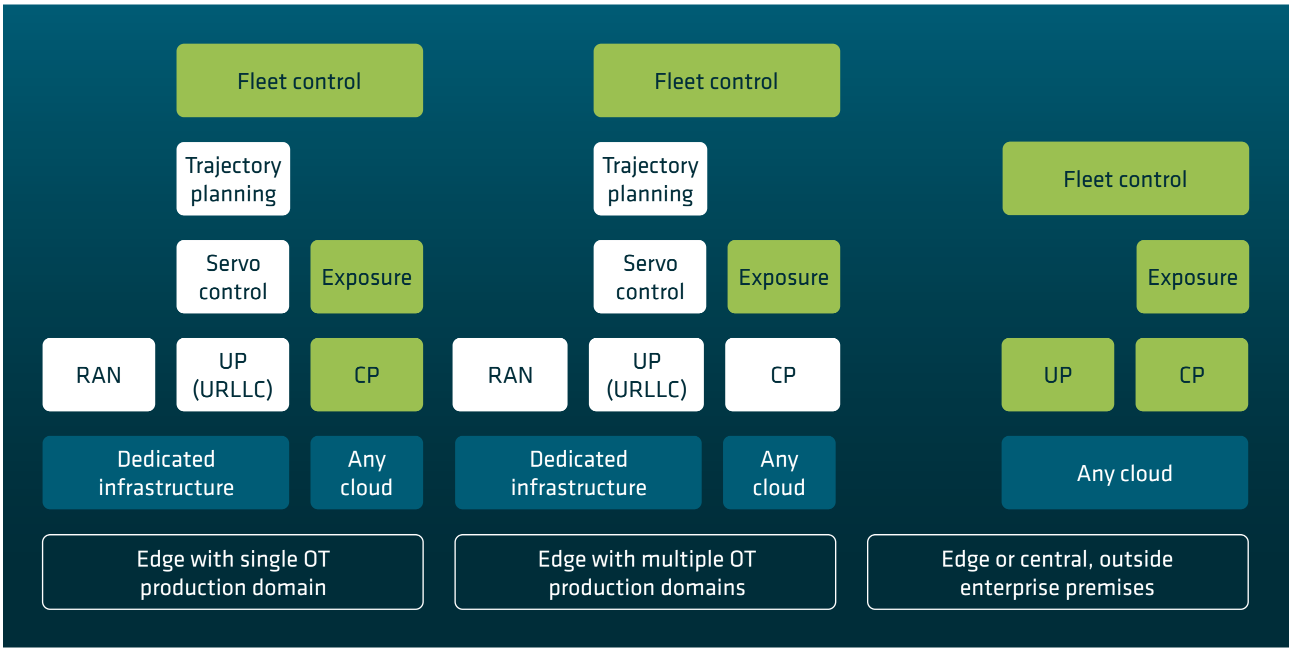
Mobile robots deliver a whole new level of flexibility in manufacturing by handling production assets in ways that include but aren't limited to transportation, storage, commissioning, placement, and picking. The related use cases (see [4] for more examples) are typical of factory automation scenarios that can benefit from the edge. In edge-enabled robotics, control intelligence is uncoupled from the robot platform, with complex computing tasks and functions being virtualized and offloaded to the edge ecosystem. This approach significantly reduces the computational load for operating mobile robots. Centralized control also makes it more efficient to operate both individual robots and entire fleets.

Depending on which control function is offloaded, cloud-enabled mobile robot control can involve multiple data streams with differing characteristics and requirements. Servocontrol calls for reliably constant, deterministic communication with a time lag of between only one and 10 milliseconds, LiDAR traffic is sporadic with a periodicity of 25 to 100 ms, and camera streams need a large bandwidth. Servocontrol requires ultralow-latency communication, which is enabled by the 5G URLLC feature. If a 5G domain is integrated in a legacy time-sensitive networking (TSN) domain in a factory, the 5G-TSN support feature should be applied. The user plane (UP) of the 5G Non-Public Network must be deployed on the factory premises, although the control plane (CP) can be shared with the public network domain.

Figure 4 shows different function placement options for implementing use cases. Instead of dealing with infrastructural and management aspects, it focuses on critical latency levels (aspects with less-than-critical latency are surrounded by green borders). To minimize latency, servocontrol and trajectory planning components should be deployed in one or more OT edge domains. Since fleet control poses more relaxed latency requirements, it can be deployed off the enterprise premises in the edge.

Capabilities can be exposed and the 5G network configured by invoking procedures such as network resource manage-

Figure 4: Different deployment options for placing telecommunications and application functions for a specific factory automation use case involving mobile robots. Telecommunication network functions and application components are also shown. Functions in green boxes may tolerate higher latency.



Source: 5G-ACIA / ZVEI e. V.

ment, e2e QoS management, TSN stream configuration, etc. To minimize latency, the servocontrol and trajectory planning components should be deployed in the edge within one or more OT domains. Because fleet control has more relaxed latency requirements, it can be deployed in the edge off the enterprise's premises. The network control plane and associated exposure capabilities can be located almost anywhere owing to their more relaxed latency requirements. A variety of redundant transmission solutions exist for ensuring robust communication between mobile robots and the control application in the edge (see [9]).

4.2 Process Automation: Closed-Loop Control

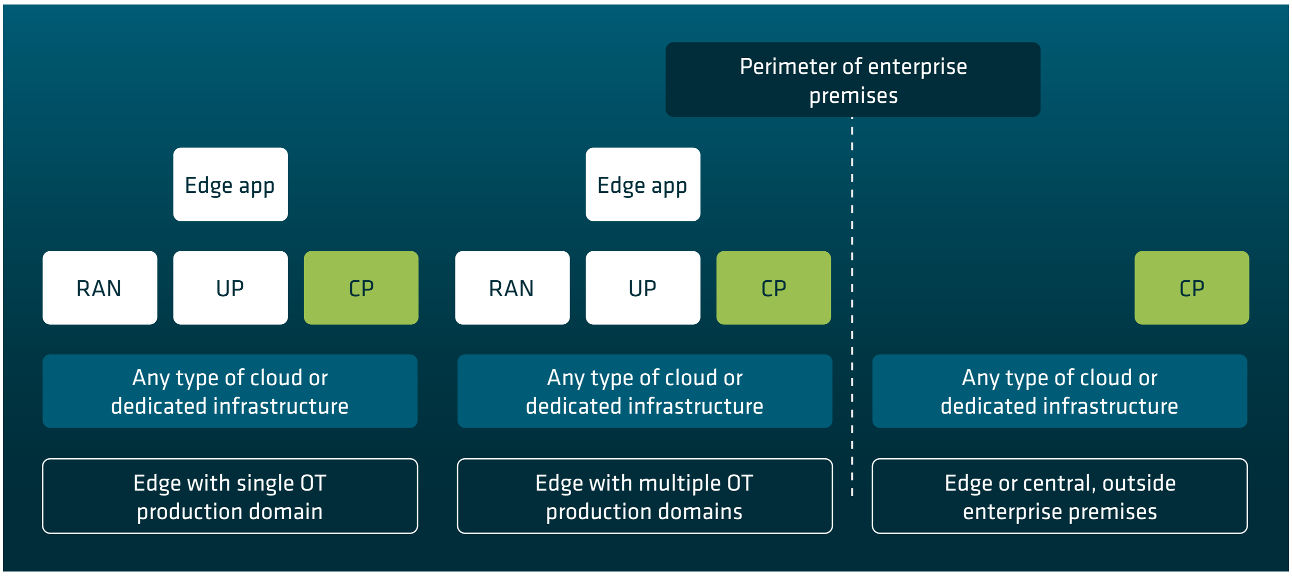
Industrial process automation applications typically pose very specific requirements in terms of determinism, reliability, redundancy, cybersecurity, and functional safety. They apply to controlling the production and handling of substances such as chemicals, foods and beverages, pulp, etc. in order to boost efficiency, reduce energy consumption, and enhance the safety of facilities. Closed-loop control use cases (see [4]) are a typical example of process automation for which the edge can be taken advantage of.

Here we consider a use case involving multiple sensors and actuators installed in a plant, with each sensor performing continuous measurements. The values obtained can be sent to a closed-loop controller application in the edge that performs calculations with them for sending control signals to actuators.

This use case has quite stringent requirements in terms of latency and service availability, but doesn't require interaction with the public network (which would necessitate, for example, continuous service and roaming capabilities) [4]. Because user data is only handled locally, the user plane also needs to be within the non-public network and OT production domains as shown in figure 5.

Measurement data received from sensors and processed data sent to actuators can be stored locally in the edge and

Figure 5: Function deployment options for closed-loop control. The green bordered-items are more relaxed in terms of requirements like data privacy, latency, and service availability. Other aspects, such as the edge runtime infrastructure type and management, aren't shown here.



Source: 5G-ACIA / ZVEI e. V.

analyzed either later or in real time, for instance for statistical analysis, optimization and so on. Dedicated data analysis applications for efficiently determining how to optimally set actuators can be moved to an OT production domain in the edge computing platform. The on-premises edge used on this case makes it possible to control production more efficiently while keeping the data within the OT production domain in order to maximize security, privacy, bandwidth, availability, and reliability.

4.3 Logistics and Warehouse Automation

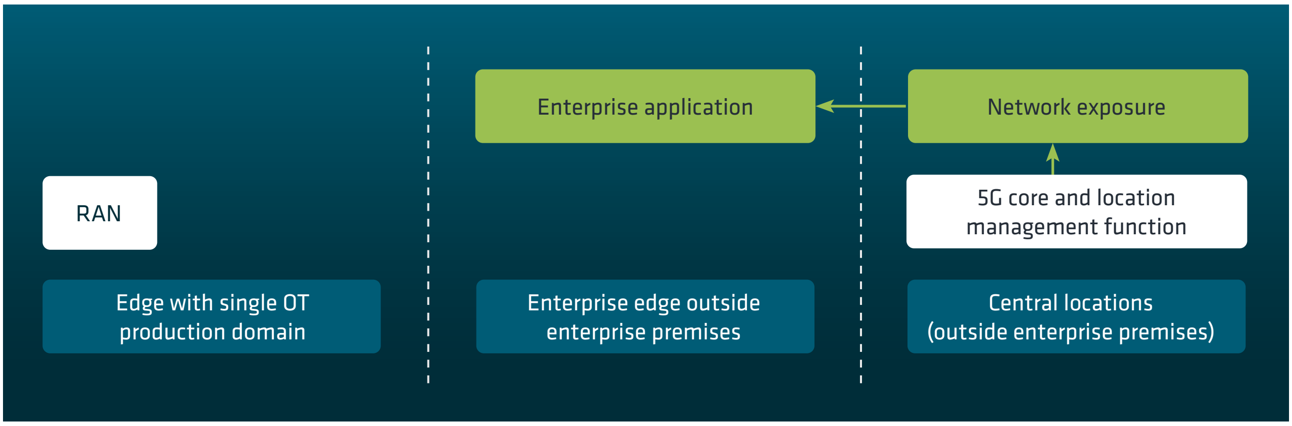
A particular localization use case may require a high level of data confidentiality and locality, as well as close to real-time operation in some cases. This may call for a local edge computing setup right on the premises, in which case it's essential to locally deploy the 3GPP-defined location management function (LMF) as well as all other functions that are needed for operation, such as radio functions, databases, and the access and mobility management function (AMF). Network exposure capabilities are essential for enabling interactions with the application (to provide it with location information or reply to location queries).

The currently provided support for precise localization and positioning doesn't depend on the user plane function (UPF), only on core network control functions and the exposure service. Real-time positioning (figure 7) also requires the related 5G core control functions to be located at the network edge, and it's recommended that they be close together to prevent network components from introducing any delays. As shown in the non-real-time case in figure 6, it's basically possible to deploy each function anywhere, even at a central location. However, for data security reasons the solutions in light green boxes may only be used if data is allowed to leave the premises, for example for remote assistance and monitoring purposes.

4.4 Human-Machine Interfaces

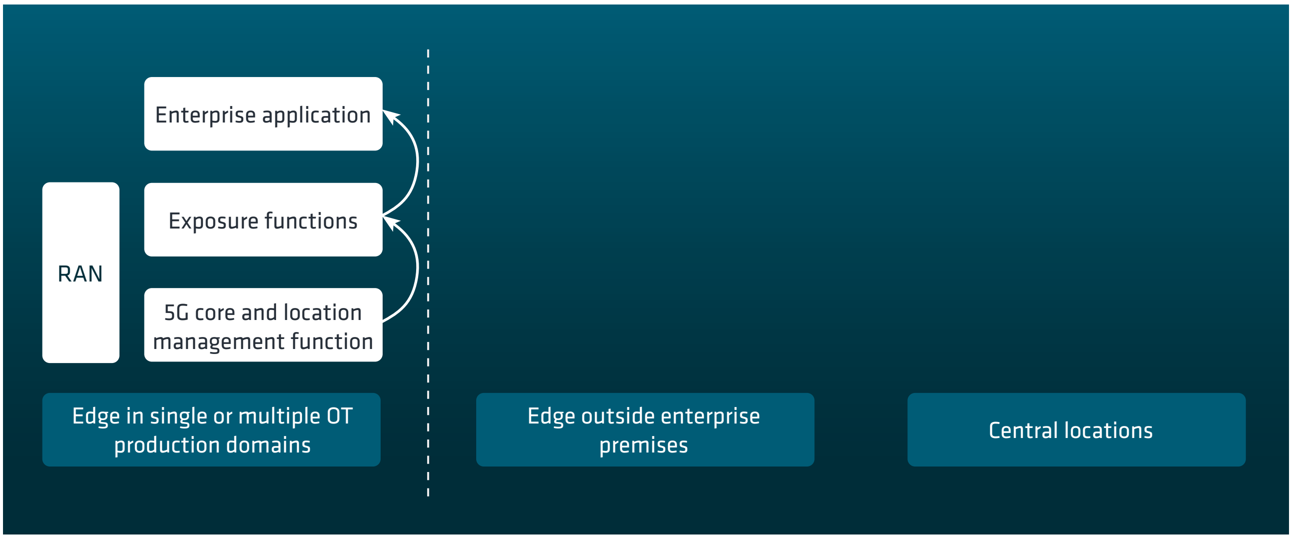
Close interaction between humans and machines is an essential prerequisite for achieving flexible production. In a given manufacturing environment, each class of machinery should be associated with a certain way of interacting with humans. To achieve this, a closed-loop safety management function monitors the data generated by all equipment (such as mobile robots and AGVs) in order to assess and mitigate risks. This function relies on an edge computing setup, since highly complex machine learning models and large data volumes are

Figure 6: An example of precise non-real-time localization showing various components of the system besides RAN deployed at different edge and central sites.



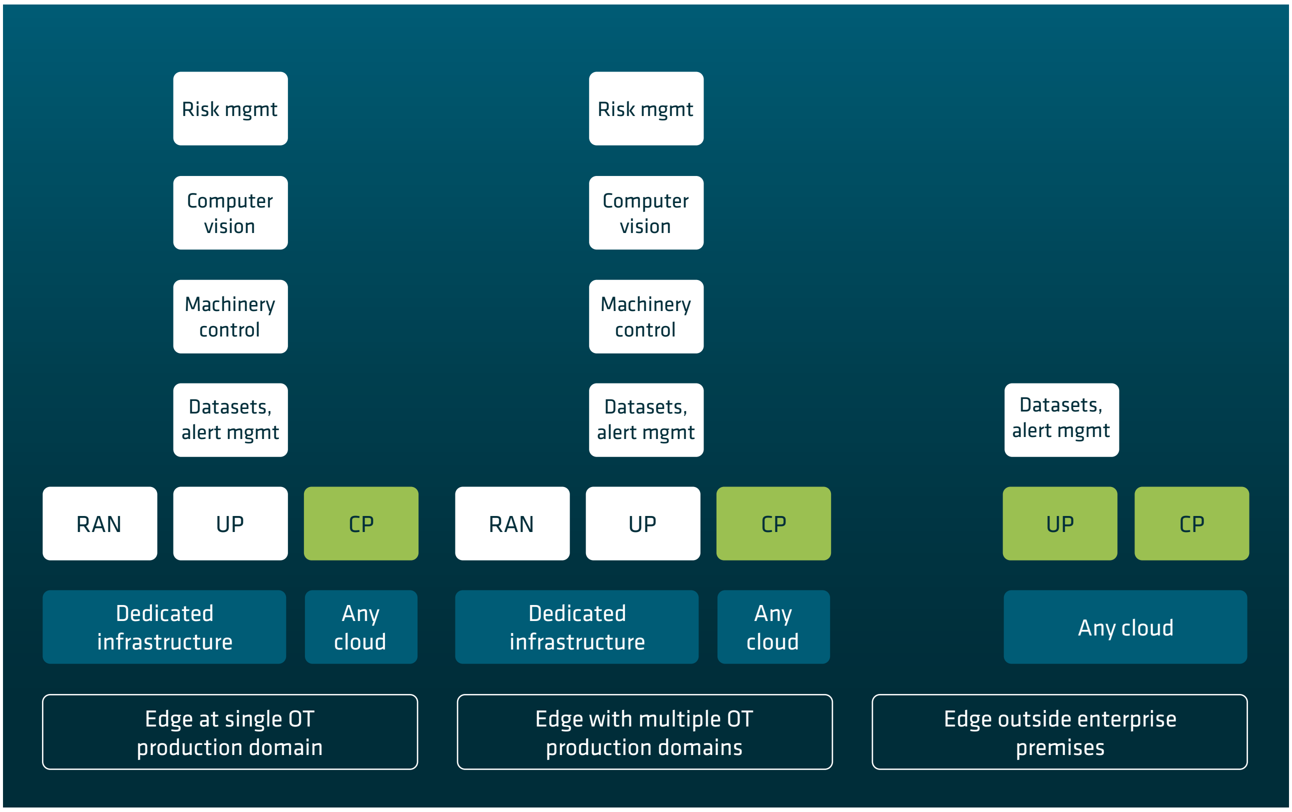
Source: 5G-ACIA / ZVEI e. V.

Figure 7: An example of precise real-time localization showing the need for core network control, location, and exposure functions to be located close to the device.



Source: 5G-ACIA / ZVEI e. V.

Figure 8: Deployment options for a human-machine interoperation use case. Multiple instances represent different options for deploying use case elements, showing dedicated infrastructure for components that are probably time-critical.



Source: 5G-ACIA / ZVEI e. V.

involved. This is particularly important when the machines only have limited processing resources at their disposal. Another benefit of shifting processing to the edge is that this makes it possible to access data from anywhere in the factory for predicting safety risks and optimizing production. Edge processing is also needed to ensure that equipment required for safe human-machine interactions responds promptly.

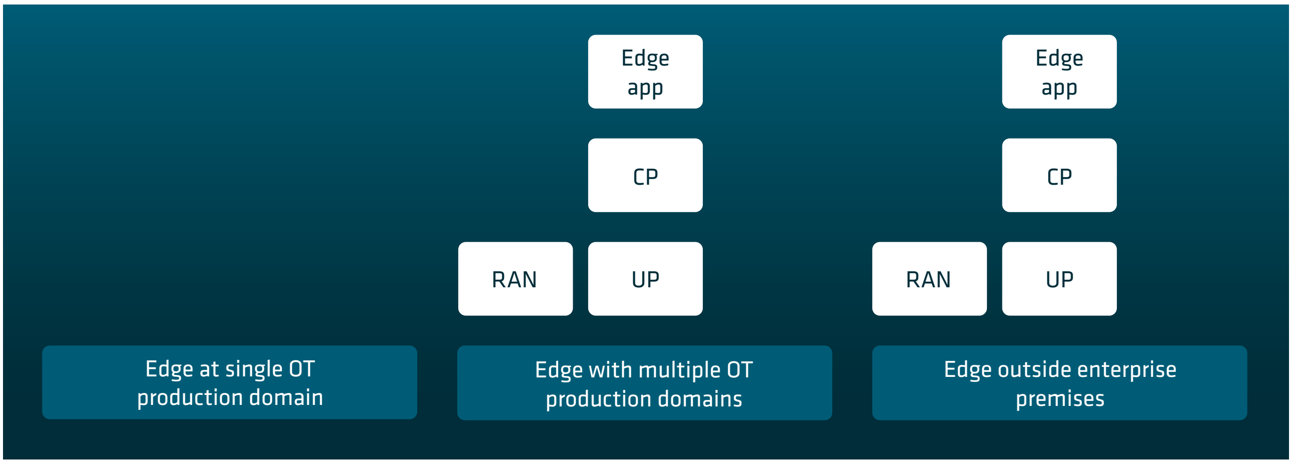
The applications of this closed loop that rely on edge computing are those that tightly interact to ensure safe human-machine interactions. They include risk management (identifying the safest movements and trajectories for machines and equipment), computer vision (detecting and distinguishing different objects in the area), machine control (actual operation of equipment), and alert management (which sends special safety warnings to workers). Figure 8 shows deployment options while making it clear that the user plane and application functions can be situated in one or multiple OT production domains. Data asset management and related data channels, including the core network control plane functions themselves, can be relegated to edge sites with more relaxed latency requirements.

4.5 Monitoring and Maintenance

Monitoring and maintenance are about tracking certain processes and/or assets in the context of industrial production without directly influencing the processes themselves. (This is in contrast to the closed-loop control systems typically used in factory automation.) They mainly involve applications such as condition monitoring and predictive sensor data-based maintenance, as well as big data analytics for optimizing future parameter sets for a particular process. The remote access and maintenance use case (see [4]) is a typical example of process automation that can benefit from the edge. In this use case, maintenance-related data from numerous sensors installed on devices is collected and aggregated, then sent to the edge where it is processed to determine whether to trigger maintenance alerts at remote control centers or in diagnostic systems. The logic for activating remote access is also shifted from the device to the edge.

The devices and entities for this use case can be installed at geographically distributed locations, and data acquisition and maintenance aren't latency-critical processes in any case. Consequently, the edge for this use case could be situated in the public network domain. If sensitive maintenance data is involved, it can be kept both on the premises and in the edge for serving multiple OT production domains. These options are shown in figure 9 below.

Figure 9: Function deployment options for remote access and maintenance.



Source: 5G-ACIA / ZVEI e. V.

5 Deployment Considerations for Standard Components of the Edge Computing Solution

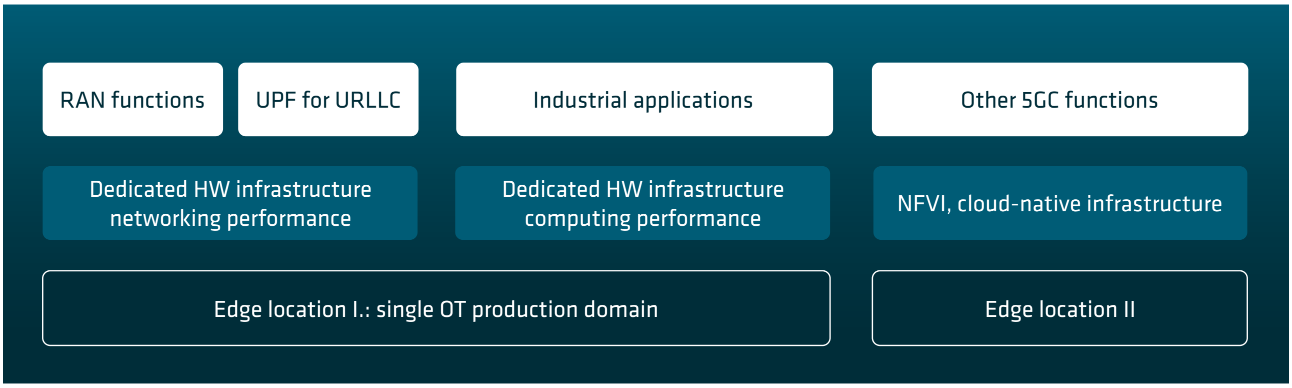
This section describes options for implementing and deploying the edge computing solution components using standard toolset and functional components.

5.1 Edge Runtime Infrastructure

Some kind of runtime infrastructure is required to deploy network functions and applications. Different types may be needed depending on the use case and nontechnical considerations, for example when implementing brownfield deployments or leveraging cloud solutions to increase flexibility or reduce the total cost of ownership (TCO). Virtual machine-based solutions are still commonly employed for 3GPP functions in networks, but cloud-native design is becoming the new standard. To achieve both very low latency and high reliability, however, it may be necessary to implement some network functions in bare-metal systems or proprietary HW infrastructure (this is discussed in [4]).

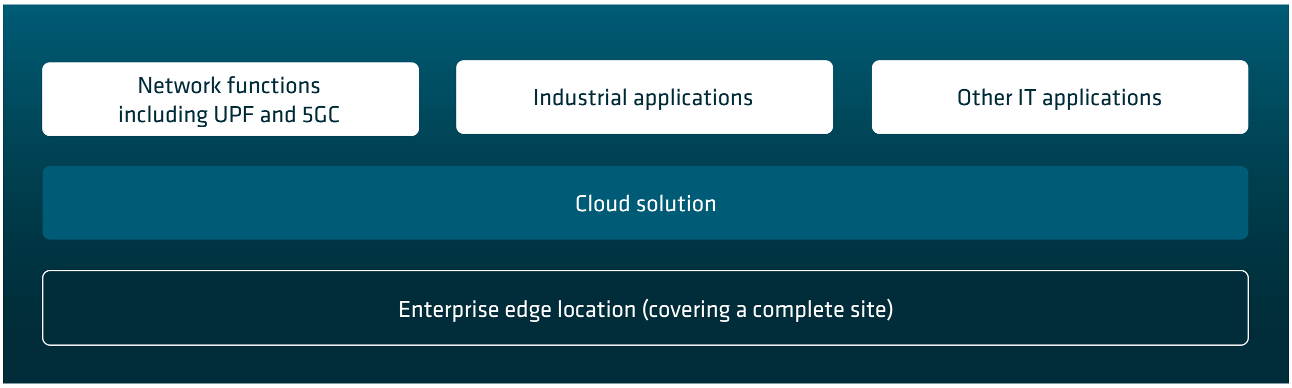
Appropriate runtime infrastructure for edge application functions can involve proprietary hardware, bare-metal installations, local IT infrastructure, infrastructure extensions for network functions, and/or third-party cloud runtime platforms. In many of these cases, special hardware acceleration

Figure 10a: Example deployment of use cases requiring high reliability and ultralow latency



Source: 5G-ACIA / ZVEI e. V.

Figure 10b: Example deployment of use cases with low (but not ultralow) latency



Source: 5G-ACIA / ZVEI e. V.

may be required to achieve efficient processing, for example using GPUs, smart network interface cards, P4 switches, persistent memory, or real-time operating systems.

The benefits and drawbacks of using different infrastructure types in various deployment setups greatly depends on the individual scenario. Flexibility is therefore very important. Figure 10a shows a typical deployment architecture for a high-reliability, low-latency solution in which proprietary infrastructure is needed in order to highly efficiently and resiliently handle a networking and computational workload in an industrial or other application. However, it is also possible for a runtime infrastructure environment to be shared by telecommunications and use case application components, as shown in figure 10b. This is an option for applications that have less stringent requirements in terms of latency and reliability.

5.2 Edge Deployment Locations and Data Plane Connectivity Path Considerations

UPFs play a key role in a 5G network for accessing edge computing systems and applications (see [9]). In particular, a local UPF steers user plane traffic toward a data network containing targeted edge applications. This UPF may be part of the edge implementation, as mentioned in [5]. In this case the UPF deployment correlates to the edge location. For integrating industrial infrastructure with 5G networks, an earlier 5G-ACIA white paper [5] presented three possible options for deploying UPF physical locations: in an OT production domain, in an IT enterprise domain, and in the public network domain. There may also be others. We based the three edge location options presented in section 4.2 on these UPF deployment models, especially for cases in which the edge is within a single-OT or multi-OT production domain on or off the enterprises premises.

In the following analysis, we examine possible locations for the UPF. Figure 11 shows three different possible locations for the UPF. In this example, the edge application and its runtime environment are both situated at the same location with the UPF. Although this configuration isn't mandatory, it is a very likely choice for many practical deployments. The figure also shows the data path running from the devices to the application servers. Here we won't discuss either the type of underlying edge runtime infrastructure or whether it is somehow shared between the telecommunications and enterprise applications; instead we will consider the advantages and disadvantages of different data path options and locations based on the previously identified requirements.

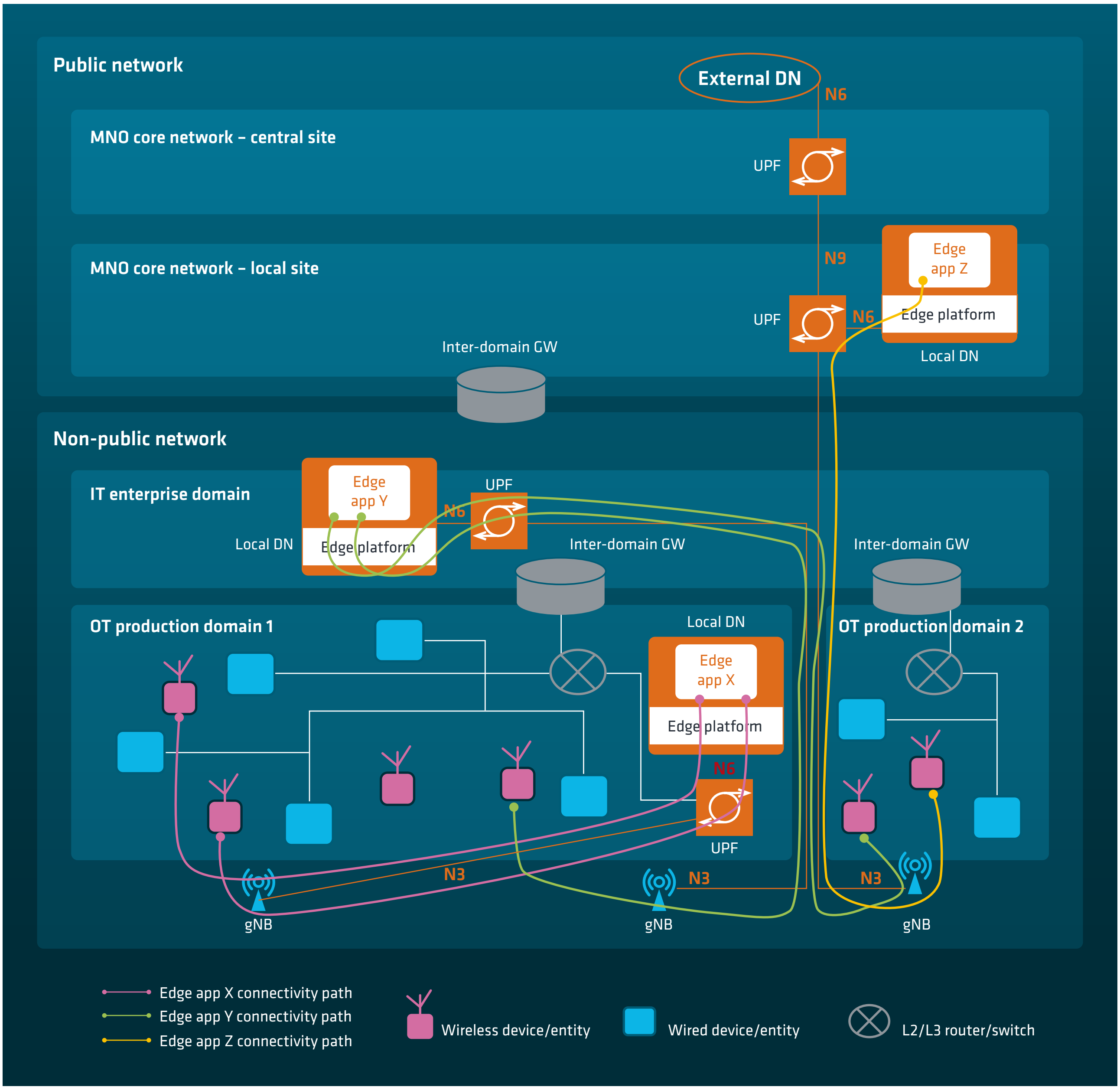
UPF deployed in a single OT production domain
In this deployment option, the UPF is located on the premises and integrated in the enterprise's shop floor infrastructure. This is a very likely option for a standalone non-public network (SNPN) (scenario 1 in [2]). It can also support Public Network Integrated Non-Public Network (PNI-NPN) scenarios 2 and 3 of [2], with either only RAN or both RAN and control plane shared with the public network.

One advantage of this approach is that, owing to the fact that the wireless and wired connections with the edge platform pass via the local UPF deployed in the OT production domain, it provides the shortest data path to the edge application without traversing interdomain gateways with FW/NAT functionality and therefore supports low latency requirements. Another is that, because the edge infrastructure and edge application are both situated within a closed enterprise environment (in a private network), privacy-sensitive and security-relevant data can be locally processed and stored. Deploying them in the same isolated OT production domain also yields benefits in terms of reliability and availability by preventing the external environment from causing any failures.

The drawback of this option is that it requires the edge computing solution components to be deployed in every existing L2/L3 OT infrastructure, resulting in greater complexity and costlier implementation. If an enterprise has multiple segregated OT production domains, each of which has its own edge site, seamless mobility of a wireless device between domains may cause issues whenever its IP anchor point (i.e., the UPF) changes. Owing to these advantages and drawbacks, this edge location option is appropriate for delay-critical, highly reliable, secure use cases that don't require seamless mobility outside OT production domains. A typical use case that works well here is closed-loop control.

UPF deployed in a multi-OT production domain
In this deployment option, the edge runtime infrastructure and associated UPF are both located on the premises. Although it resembles the previous option, its edge runtime infrastructure is shared by multiple OT production domains. The edge site can be deployed either within the enterprise shop-floor infrastructure or within the IT enterprise domain. This option corresponds to the same deployment scenarios as option 2 in [2]: SNPN with shared RAN or PNI-NPN with shared RAN and control plane.

Figure 11: Three different edge locations with a possible data path for each one. In all three cases, the UPF corresponds to the edge site.



Source: 5G-ACIA / ZVEI e. V.

This solution provides data privacy and low latency. The edge site is on the premises, its service area contains multiple OT production domains, and the wireless and wired data connectivity paths are slightly longer than in the first option. However, they are still short enough to support many use cases that require low latency.

Data that is sensitive in terms of privacy and security can be safely stored locally within the enterprise. Since all user plane traffic takes place on the premises and nothing in the external environment can cause any failures, this option is also good in terms of reliability and availability. Another advantage compared to the single OT domain option is that it enables seamless mobility for wireless devices between OT production domains, since all of the latter share a common IP anchor point (the UPF).

The drawback of this option is that it requires 5G network elements to be integrated into the existing IT or OT infrastructure, which increases the complexity and cost of implementation. If there is a FW/NAT gateway between different OT production and IT domains, and if it goes via the wireless domain data plane, latency may be slightly greater and reliability/availability somewhat lower than in the single-OT production domain deployment option. It might not be applicable to use cases that require extremely low latency or very high reliability/availability.

A typical use case that fits well here is mobile robots that need to move between OT production domains.

UPF deployed outside the enterprise premises
In this deployment option, the edge and associated UPF are located outside the enterprise premises in third-party cloud or mobile network operator data centers, which can be either local or centralized.

This deployment option is most likely in the PNI-NPN deployment scenario, namely scenario 4 in [2], in which an NPN is hosted by the public network. In this scenario, all network traffic is routed via an external network to or from the NPN subscribers.

The advantage of this option is that it doesn't require the edge and associated UPF to be integrated in an enterprise IT/OT infrastructure. It therefore incurs low costs and can be quickly deployed. Since this option addresses the ordinary MNO network, it can support typical public network features such as mobility within and outside the enterprise. Because the edge data center may be in close geographical proximity, it delivers low-latency communication and computing capabilities for data aggregation and benefits from significant savings in terms of the bandwidth of the transport network.

The downside of this option is that any traffic flowing to or from devices within the OT production domain has to pass through two FW/NAT gateways to reach the edge, one between an OT production and an IT enterprise domain and the other between the IT enterprise and the public domain. This results in additional delays and more complex gateway configurations. What's more, if the edge and the associated UPF deployment are far from the enterprise premises, in other words deeply inside the core network hosted by the MNO, this may result in latency and availability levels that only support use cases that require neither real-time communication nor high reliability. In addition, because the edge infrastructure is located in the public realm, due to privacy and security concerns it may not be the best choice for storing sensitive private information.

Weighing these advantages and disadvantages, this edge location option is appropriate for less latency-critical use cases and especially for use cases spanning geographically distributed locations for which it isn't necessary to keep sensitive data safely stored inside the enterprise. A typical use case here is asset tracking.

5.3 Data Plane Connectivity Path Considerations

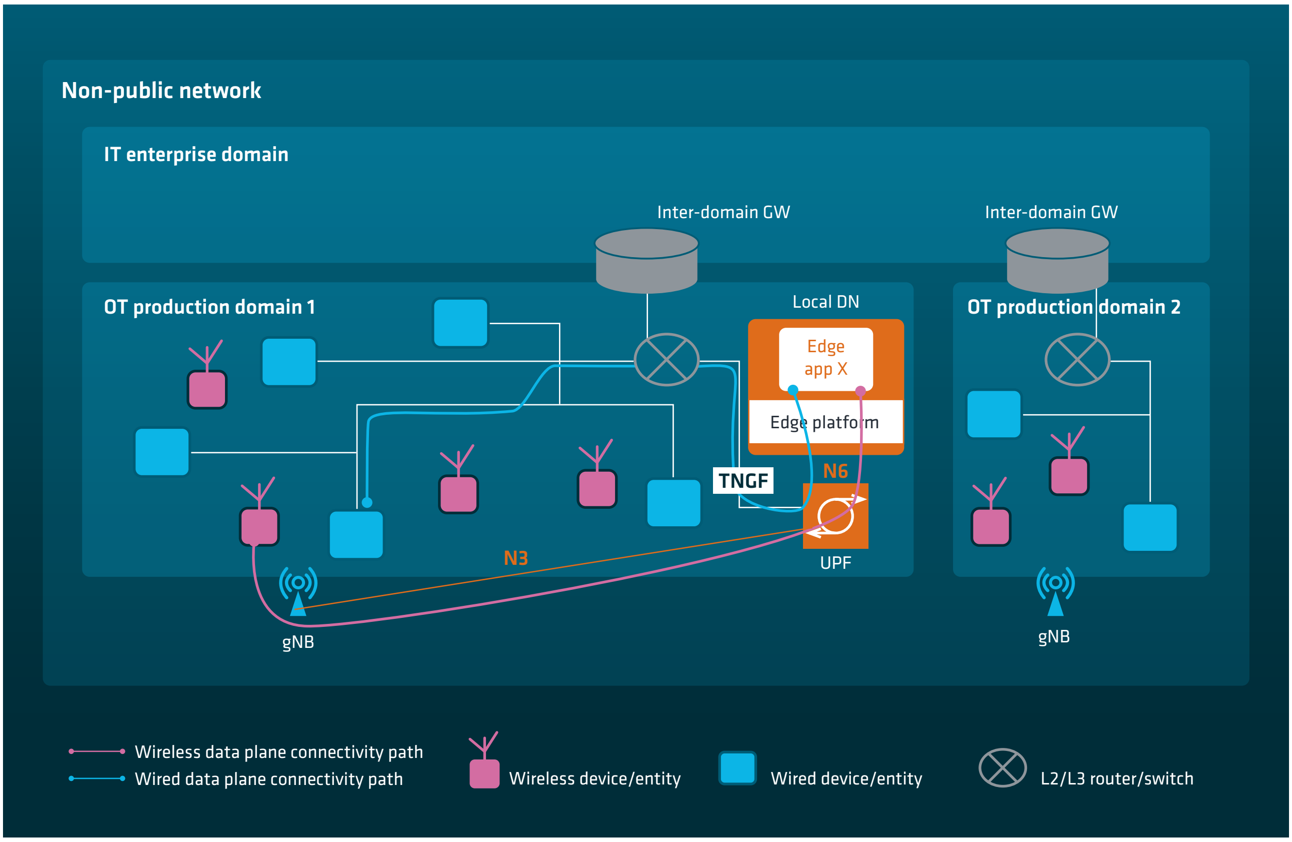
Legacy installations can pose challenges when introducing 5G wireless networks and edge computing capabilities for shop floor use cases. In many cases, wired devices will continue to be used, for several reasons. Some wired devices won't necessarily benefit from wireless connectivity, or else it would be too costly to build wireless versions. Some wired devices, like industrial computers, also host edge application servers. In this section, we present different approaches for the single-OT production domain edge location option, some of which can be used instead of other edge location options. For background information on the scenarios, we recommend the 5G-ACIA white paper "Integration of 5G with Time-Sensitive Networking for Industrial Communications" [4].

Data paths that traverse the 5G system

One approach for connecting to the edge is for the data paths from both wired and wireless devices to pass through the 5G system (5GS). This option makes it possible to take advantage of UPF with layer 2 and 3 automation control systems as described in Annex A of [5]. Since 3GPP treats wired OT devices as trusted but non-5G-capable (N5GC) devices, according to TS 23.501 and 23.316 a Trusted Non-3GPP Gateway Function (TNGF) should be involved in the connectivity paths.

Figure 12 shows an example in which both wireless and wired data plane connectivity paths traverse the UPF to the edge, which is deployed within in a single-OT production domain (OT Production Domain 1). The data path for the wireless devices runs from the gNB to the UPF there, while the data path for wired devices reaches the UPF via the TNGF. The UPF offloads data traffic from wired and wireless devices to the local data network via N6 toward the edge platform and associated edge application. If the edge platform is shared by multi-OT production domains (in other words, deployed in the IT enterprise domain), the UPF plays the role of common IP anchor and traffic aggregation point, in addition to supporting seamless mobility for devices moving between OT production domains. If the edge and associated UPF are located within the mobile network operator (MNO) network, in other words in the public domain, a UPF near the enterprise is chosen for

Figure 12: Wireless and wired data plane connectivity paths via the UPF to the edge deployed in an OT production domain



Source: 5G-ACIA / ZVEI e. V.

steering data traffic to the local data network that contains the edge runtime infrastructure and applications.

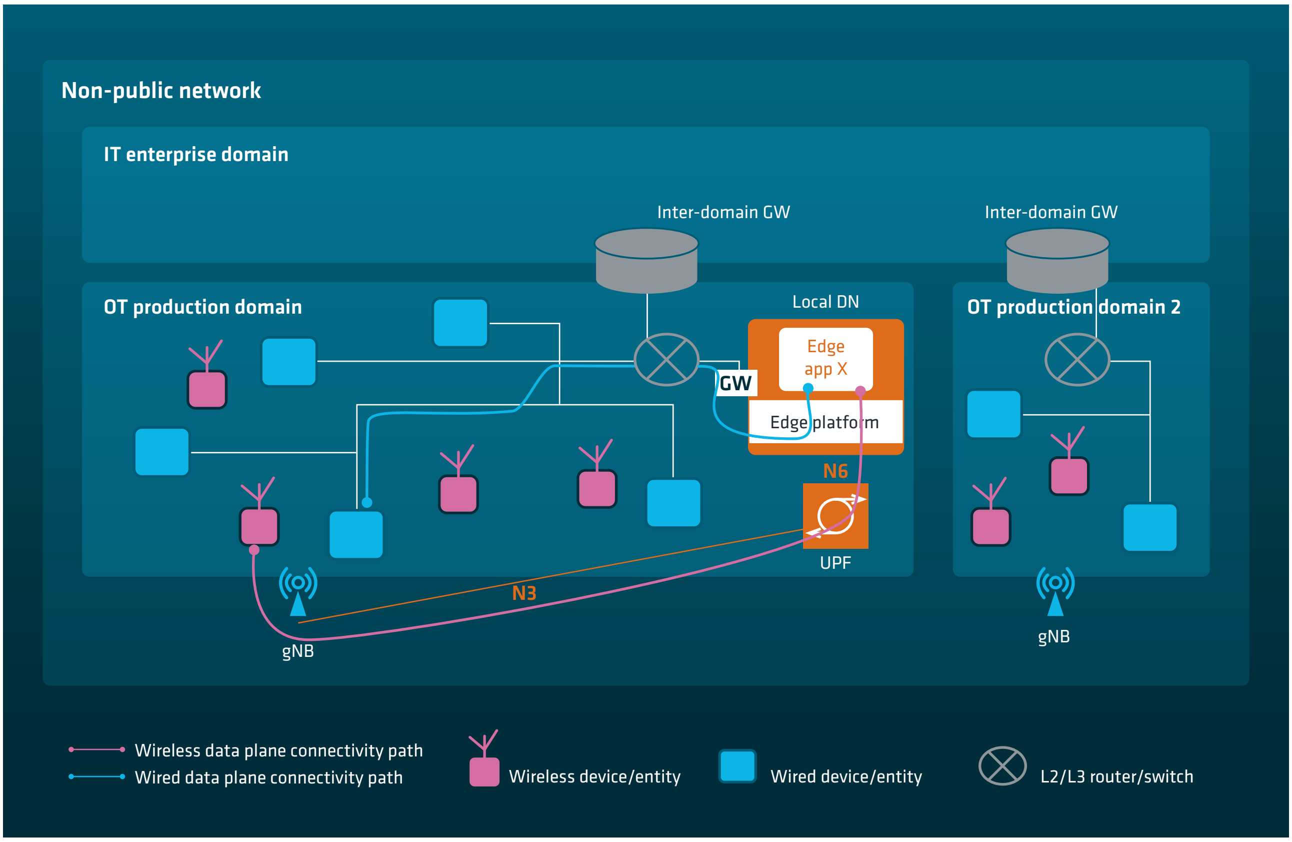
The advantage of this connectivity approach is that the edge application and platform communicate with both wired and wireless devices via the 5GS. The edge application and platform can therefore be access-agnostic, which makes them easier to design and implement. The drawback is that it is necessary to integrate wired access into the 5GS, thus increasing the deployment's overall complexity.

Data paths that bypass the 5G system

Another connectivity approach is to have wireless devices pass through the UPF as in the previous option, but with the wired access to the edge designed to bypass the 5GS. Then the data path from a wired device passes through the L2/L3 infrastructure and is linked to the edge platform via a gateway. Figure 13 shows an example of a single-OT production domain. This connectivity scenario is also straightforward to use when the edge is deployed in an IT enterprise or the public domain.

The advantage of this scenario is that it is easy to implement. The downside is that the edge application and platform need to be orchestrated to support communication both with wired devices via LAN and in parallel with wireless devices via a 3GPP network. In addition, if wired and wireless devices (such as controllers) need to exchange information, they should do so via the edge application and platform.

Figure 13: Wireless data plane connectivity paths via the UPF and wired connectivity paths that bypass the UPF to the edge, which is deployed in an OT production domain



Source: 5G-ACIA / ZVEI e. V.

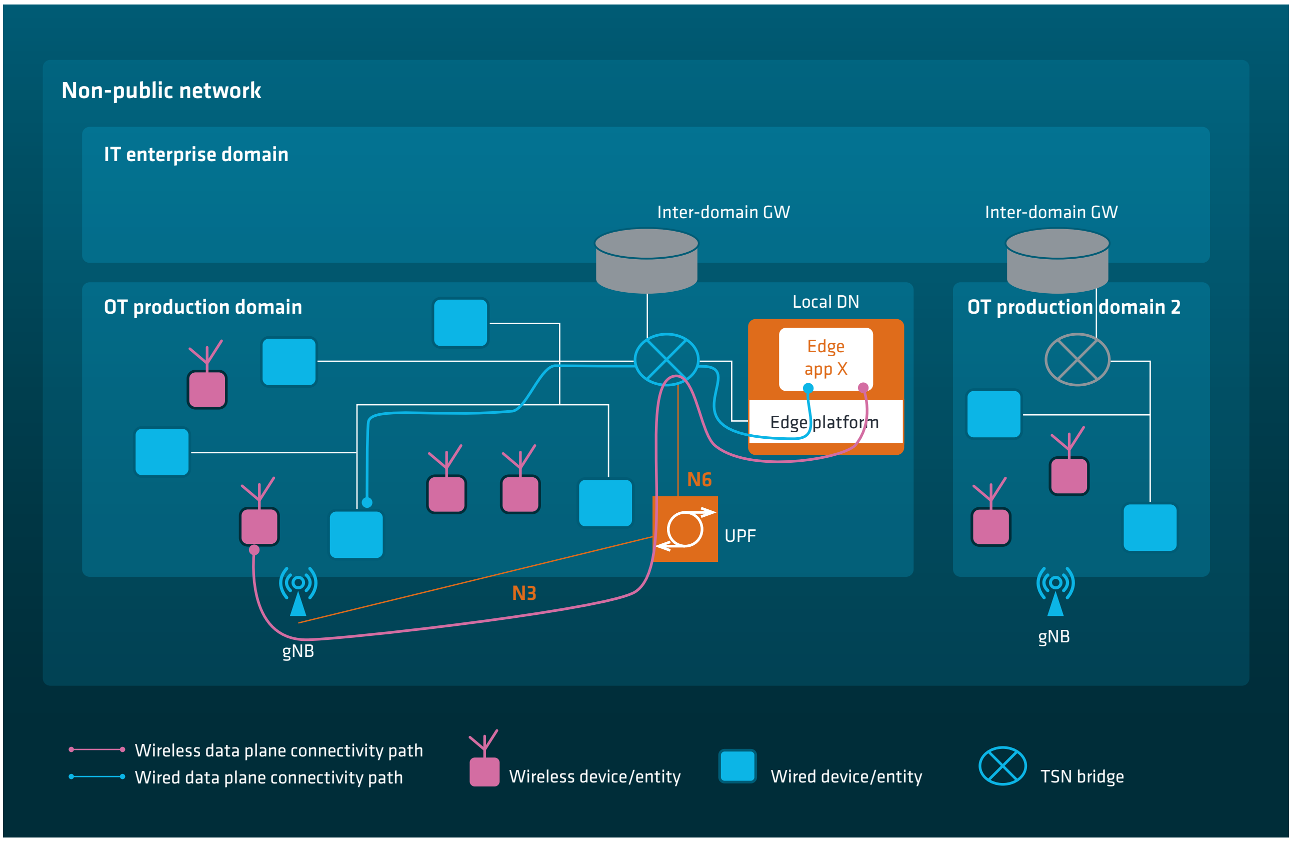
Data paths for time-sensitive communications

In Release 16, 3GPP expanded the 5G system architecture to support IEEE 802.1 Time-Sensitive Networking (TSN) as specified by TS23.501. This 3GPP-enabled 5G-TSN integration can be taken advantage of when TSN technology is used to connect devices to the industrial application deployed in the edge. In this case, the 5G network domain is treated by TSN control plane entities as one or more virtual 5G TSN bridges on a User Plane Function base, which ensures seamless integration of the 5GS and wired TSN domains. Figure 14 shows an on-premises deployment case for this connectivity scenario, in which the UPF and edge platform are deployed within the OT production domain.

To reach industrial applications deployed in the edge, wireless devices are connected to a 5G virtual bridge while the N6 interfaces of the UPFs are linked to a legacy wired TSN bridge. Figure 14 shows a scenario in which the UPF is deployed in the OT production domain, but the solution also works if the edge is in the IT enterprise domain with multiple UPFs deployed in the OT production and enterprise domains. This connectivity setup can support scenarios with multiple on-premises edge locations.

An advantage of this solution is that wired devices can use the legacy TSN network. This eliminates the need for the 5GS to provide wireline access support for wired devices. Beyond this, no other specific features are required because 3GPP has built TSN support into the 5GS. TSN technology can therefore be used as a converged communication solution. One constraint does exist, however: in practical terms TSN can only be deployed on the premises, so this connectivity option can't be used if the edge is somewhere else.

Figure 14: Seamless integration of wireless device in wired TSN segment 3GPP 5G-TSN



Source: 5G-ACIA / ZVEI e. V.

5.4 Choosing the Data Plane and Edge Application Server

UPF selection and deployment have been discussed in detail in the preceding sections. The presented options for placing the UPF and steering data plane connectivity paths show clearly that where a UPF is deployed affects not only latency and availability, but also how easy or hard it is to implement and the associated costs. The network configuration between the UPFs and edge application servers (EASs) also affects latency, redundancy, privacy, and complexity, especially considering that in most cases these correspond to different entities. The method used to select the data plane and thus also the serving EAS (to which the device will actually connect) depends on the applicable standards and the types of applications involved, among other things.

The 3GPP specifications describe features for

- enabling the 5G Core network to select, based on the UE's location, a UPF that is able to steer traffic to the data network's most appropriate local access point, which can be an OT production domain, an enterprise IT domain, or an edge platform off the premises, and
- supporting application clients (AC) in the UE for identifying the address of the appropriate (e.g., topologically closest) edge application server (EAS) in the data network.

Some deployment options support coordinated UPF selection and EAS discovery. The 3GPP specifications describe features for the following:

- **IP-type traffic**: When the application layer relies on IP communication, the EAS is selected by the DNS server in response to a query by the UE. The DNS query must be handled by an authoritative DNS server within the data network. A DNS resolver within the 5G Core network can be configured to receive the DNS query, possibly add more information (like the ECS option defined by 3GPP technical specification # 23.548 [12]), and forward it to the corresponding authoritative DNS server in the data network.

  When a DNS query is sent, the UPF forwarding the DNS can be:
  - a local UPF (distributed anchor, see 3GPP technical specification # 23.548 [12]) that was previously selected when the UE initiated the PDU session for a particular data network and/or network slice or
  - a local UPF that is dynamically selected based on information derived from the DNS query. This can be done by:
    - ◊ establishing a new PDU session and selecting a UPF for a specific data network and/or particular network slice based on the destination address of the DNS query or
    - ◊ reconfiguring an existing PDU session by selecting a local UPF based on a session breakout or distributed anchor model. The local UPF can be chosen by the 5G Core network (see EASDF in 3GPP technical specification # 23.548 [12]) based on information included in the DNS query or the subsequent DNS response (for example, an EAS IP address).
- **Ethernet-type traffic**: When the application uses Ethernet communication, traffic is tunneled between the UE and a UPF chosen by the 5G Core network when establishing the PDU session. UPF selection can be performed for a specific data network and/or network slice requested by the UE. If IEEE TSN features are used for Ethernet-type traffic, the 5G Core network exposes the API that the TSN Central Network Controller (CNC) invokes to configure the 5G system like a TSN bridge. In this case, the CNC steers traffic to the appropriate local access point of the data network (the 5G system's "bridge port").
- **Application-layer methods**: An Edge Enabler Client (EEC) in the UE (possibly implemented by the device OS) supports the application clients for discovering the address (such as the URI or IP address) of the appropriate EAS. If the EAS address is a URI, an additional DNS query is required to discover the EAS, reverting to what was described above for the IP traffic case.

5.5 Examples of Private and Public Network Integration Scenarios

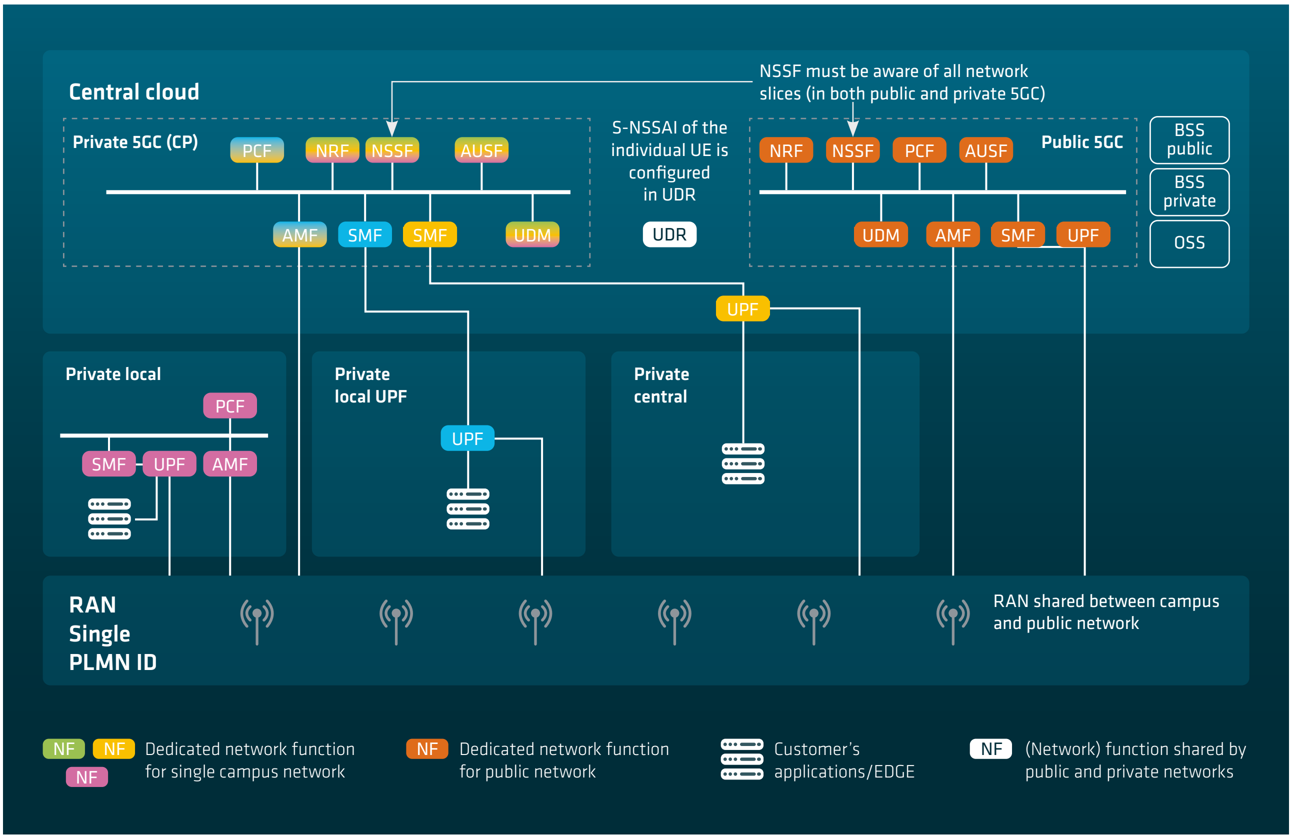
PNI-NPN deployments offer a variety of choices for deploying different network functions to meet specific customer requirements, such as low complexity, local handling of data, URLLC, a positioning service, etc. 5G network elements can be also deployed in an edge-cloud environment to achieve greater flexibility, scalability, and efficacy.

When a simple solution is wished, a central private network is an option (see figure 15). In this scenario, all network functions are deployed at a single central location. The customer's applications and edge can be deployed either centrally or locally. The resulting greater simplicity and absence of local network functions can save resources while also supporting use cases that don't require very great latency.

It is possible to deploy user plane functionality (UPF) on a campus (see figure 15 – "Private local UPF"). In this case, customers' applications and associated data traffic can benefit from not having to traverse the transport network to the cloud. This approach achieves ultra-low latency for traffic flowing between an enterprise's devices and the corresponding applications.

Sometimes there is a greater need for latency in connection with additional services (such as positioning). This can be the case if more network functions have to be on the factory premises (such as control plane functionality like AMF, SMF, and LMF - see figure 15 Private local). To meet the requirements for use cases in scenarios of this kind, control plane entities could also be located in the local edge. They could also handle control plane data on the premises, with only subscription data remaining in the main cloud.

Figure 15: Private network (PNI-NPN) deployment scenarios according to use-case-specific needs



Source: 5G-ACIA / ZVEI e. V.

Figure 15 shows all of the scenarios in which a separate 5G cloud (devoted exclusively to managing private networks) could be used to provide dedicated services to private networks (such as positioning, control plane functionality e. g. AMF, SMF, and LMF, URLLC, etc., that would not be available to public subscribers), thus distributing the control plane. Public networks may be required by law to meet additional requirements (such as support for emergency calls and lawful interception capabilities) that don't apply to private networks and therefore don't need to be enabled in a private 5G cloud.

A PNI-NPN can have more than one slice. For example, a possible two-slice solution might have one for URLLC and another for all other communications within the private network. One of them could include a radio access network (RAN) scheduler with functionality for ensuring the availability of radio resources.

Private networks often have access to a dedicated spectrum (like the hotly debated so-called "industry spectrum" in Germany), which they may only use for a specific purpose and not for public mass market services. Built-in functionality in the RAN associates users of a network slice with the corresponding spectrum. This can be accomplished with dedicated cells within the dedicated spectrum that are only available to UEs that belong to the corresponding campus network (identified by a network slice ID, the so-called single-network slice selection assistance information or S-NSSAI). These dedicated cells can also be protected using the closed access group (CAG) feature. Alternatively, there can be a single cell encompassing both mass market and dedicated spectrum. Within such a cell, only UEs with radio resources within the spectrum corresponding to UE's S-NSSAI are made available.

6 Conclusions

We have reviewed a number of industrial automation use cases and assessed whether they can benefit from or even require edge computing capabilities. The analyzed capabilities as well as the use cases are based on specifications defined by industrial and standardization forums. It has been observed that the capabilities, and therefore also the benefits, of the use cases include latency and locality as well as the capabilities of scalable computing resources, which can improve data aggregation and save bandwidth with local processing.

After identifying the relevant use cases, their principal communication and computation requirements, and the benefits of using the edge, we have proposed alternative locations for edge implementation and deployment options. 3GPP addresses the general requirements for implementing and deploying use cases in terms of networking and communication. We have investigated several deployment options in greater detail, especially for the data plane (the User Plane Function (UPF) of the 3GPP 5G Core network) and analyzed the drawbacks and benefits of each option.

For cloud and runtime environments, we have identified several possible implementation options involving multiple edge site instances that host different bare-metal and cloud platforms, without going into detail on the management and orchestration aspects. We have also outlined the interactions between runtime application instances and the network and how they affect the deployment options for 3GPP exposure interfaces, and concluded that local exposure of capabilities is required for some use cases.

In general, the standard feature set and flexible deployability of 5G enable a wide variety of edge computing configurations for implementing all currently anticipated use cases. We have also observed that the deployment options excel in terms of flexible use in a variety of runtime environments.

7 Key Terms and Acronyms

3GPP
The 3rd Generation Partnership Project (3GPP) is an umbrella term for a consortium embracing a number of standards organizations worldwide that are collaborating to develop globally accepted specifications for mobile telecommunications. As its name implies, it was originally created to establish specifications for the third generation (3G) of mobile communication systems. It has continued working on subsequent generations, including the fifth generation (5G), which is considered in this white paper.

5G-ACIA
The 5G Alliance for Connected Industries and Automation is the globally leading organization for shaping and promoting Industrial 5G.

5G-SMART
5G-SMART is an EU-funded research project devoted to demonstrating, validating, and evaluating the potential of 5G in actual manufacturing environments.

AGV
Automated guided vehicle.

AMF
Access and Mobility Management Function.

EAS
Edge application server.

Edge computing
Edge computing takes storage and computation closer to where data is sourced. Traditionally, cloud computing has taken place in remote data centers. Edge computing moves part of this activity right onto or very close to the premises to achieve data privacy, an actionable feedback loop, bandwidth savings, local management etc.

ETSI MEC
ETSI multi-access edge computing.

ETSI NFV
ETSI Network Functions Virtualization.

Far Edge
Far Edge is the edge computing infrastructure that is deployed farthest from the cloud data center(s) and closest to the users. It can be deployed at enterprises and factories. MEC infrastructure typically gets deployed as a Far Edge. The applications that run there are characterized by very low latency, high scalability, and high throughput; typical examples are AR/VR and gaming apps.

GCS
Guidance control system.

GSMA OPG
GSMA Operator Platform Group. The GSM Association represents the interests of mobile network operators worldwide.

IP
Internet protocol.

LMF
Location management function.

MNO
Mobile network operator.

NAT
Network address translation.

Near Edge
Near Edge is an edge computing infrastructure deployed between the Far Edge and the cloud data centers. While Far Edge computing infrastructure hosts applications specific to the location where it is deployed, a Near Edge hosts generic services. A mobile network operator service aggregation point can constitute a Near Edge.

NEF
Network Exposure Function.

Network function
Based on the Virtual Network Function defined by ETSI NFV, this term refers to all functions that are needed for the 3GPP network to operate. It does not include either the underlying

infrastructure or the functions needed for management and operation.

NIC
Network interface card.

NPN
Non-public network.

OT
Operational technology.

P4
Programming Protocol-independent Packet Processors, an open-source, domain-specific programming language for network devices.

PDU
Protocol data unit.

PID
Proportional integral derivative (PI/PID).

PNI-NPN
Public Network Integrated Non-Public Network, a 5G network for private enterprise use.

Private edge
This is when the storage and computing infrastructure is located on the premises and devoted exclusively to private use. It may be run and managed by a factory operator or service provider or jointly with the owner of the premises or factory. It provides data localization and low latency and supports applications that require time-sensitive data. The usual latency requirement is < 15 ms.

Public Edge or MNO Edge
This describes a shared infrastructure with storage and computation capabilities that can accommodate multiple tenants and is located close to where data is generated. A Public Edge is managed by a service provider, with users receiving access to virtual space and a computing platform for launching applications that communicate with a local data source. Usually the latency is greater than 15 ms. The bandwidth capacity and data security (device to edge) vary depending on the service provider's offering.

Public cloud
This comprises data centers that are typically operated by cloud service providers such as AWS, Google, Microsoft, etc. A public cloud can be located in the Internet or distributed. Far Edge and Near Edge are examples of public clouds in which a cloud infrastructure is shared by multiple tenants via secure connections.

Non-public network (NPN)
3GPP has broadened the scope of private networks by introducing non-public networks. Previously, a private network meant an isolated network with certain device access. Non-public networks have two deployment options: SNPN and PNI-NPN. An SNPN is a standalone non-public network, while a PNI-NPN is a public network integrated non-public network connected to the operator's network. 5G-ACIA has defined four different non-public network configurations in the 5G ACIA white paper "5G Non-Public Networks for Industrial Scenarios" [2].

OPCF
The OPC Foundation, an industry consortium that creates and maintains standards for open connectivity of industrial automation devices and systems. It has established interoperability standards for securely and reliably exchanging data in the industrial automation space and other industries. The standards are platform-independent and ensure a seamless flow of information among devices of multiple vendors.

Private network
Usually a standalone deployment of a non-public network. The term is also commonly used to designate a network that is under a property owner's control, although it may be managed by a third party.

RAN
Radio access network.

SLAM
Simultaneous localization and mapping.

SMF
Session management function.

SNPN
Standalone non-public network.

TNGF
Trusted Non-3GPP Gateway Function.

TSN
Time-Sensitive Networking, a set of standards under development by the Time-Sensitive Networking task group of the IEEE 802.1 working group.

Time synchronicity
Clock synchronicity, or time synchronization precision, is defined between a sync master and a sync device. The requirement on the synchronicity budget for the 5G system is the time error contribution between ingress and egress of the 5G system on the path taken by clock synchronization messages.

UE
User equipment (wireless 5G transceiver).

UPF
User plane function.

URLLC
5G ultra-reliable low-latency communications.

8 Annex: General Method for Evaluating the Relevance of Edge Computation to Different Use Cases

A wide variety of use cases exists, each of which has **different requirements** in terms of, for example, end-to-end latency, data rate, service continuity, security and privacy, service availability, and reliability. Latency can be important for one use case, while another may require privacy or reliability or both. Edge computing offers excellent **capabilities and features** that can help meet the requirements of specific use cases. These can include, for example, local data collection, data processing, data storage (including localized content), data analytics, data aggregation, and firmware and software updates.

For evaluating the "edge computing relevance" of a given use case, we apply the principle that a use case's eligibility for edge computing increases with the number of requirements that an edge system's capabilities can help meet.

8.1 Overview of Method

The method comprises the four main steps shown in figure A1.

Step 1 involves briefly describing the use case and assumptions about it. The description mentions any related functions or operations that could potentially be accomplished with edge computing. The assumptions capture and compare possible types of computations (e.g., edge versus remote cloud computing or edge versus device computing) and the type(s) of data communicated (e.g., control commands, management data, measurement data, video transmissions etc.).

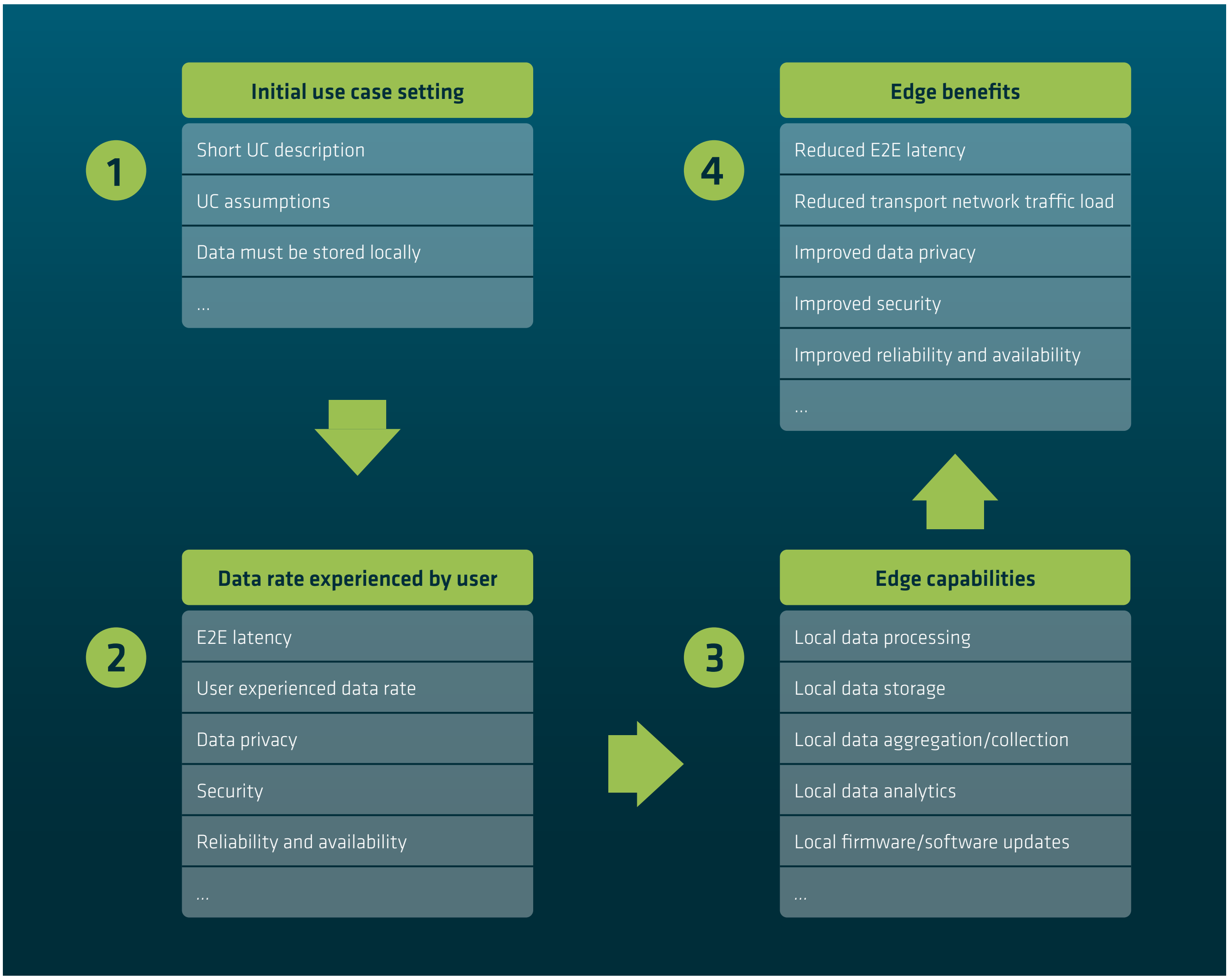
Step 2 consists of describing requirements that are critical for or important to a use case. These can be service performance requirements like end-to-end latency or service availability, or other requirements unrelated to performance such as offloading of heavy computational loads to another device, centralized coordination, etc.

Step 3 is identifying any "edge computing" capabilities and features that may help support the specific use case requirements defined in the second step.

Examples of these **edge computing capabilities and features** are:

- Local data capture
- Local data processing
- Local data storage (incl. localized content)
- Local data analytics
- Local data aggregation from a large number of sources
- Local firmware/software updates (from edge devices to end devices)
- Bandwidth savings (localization)
- Offloading of heavy computations (from end device(s) to edge computing)
- Coordination of end devices and entities
- Security
- Privacy

Figure 16: The main steps of the method used to evaluate the relevance of edge use cases



Source: 5G-ACIA / ZVEI e. V.

Step 4 involves mapping edge computing capabilities to relevant use case requirements to identify achievable benefits: for example, that local data processing in the edge cloud can help meet a use case's low latency requirement.

The next sections explain how the four steps of this method work in detail, taking the case of a mobile robot as an example.

8.2 Examples of Edge Relevance Analysis Based on the Defined Method

8.2.1 Use Case Description and Assumptions (Step 1)

The mobile robot use case is described in detail in 3GPP TS 22.104 and TR 22.804. A mobile robot is a programmable machine that's able to perform a wide variety of tasks while following programmed paths. Automated guided vehicles (AGVs) are a subgroup of mobile robots.

Mobile robots and AGVs are monitored and controlled by a guidance control system (GCS), which assigns jobs to them, sends them up-to-date process information, and manages traffic while preventing collisions [TS 22.104].

Edge computing infrastructure (or a remote cloud) can be used to host remote control system functionality of this kind (for example, GCS). In particular, edge computing capacities can store and process video and images received from cameras of the AGVs. After processing the data, control and management commands are sent from the edge to the AGVs. For instance, if the edge application recognizes an obstacle in an AGV's path or detects a malfunction, it sends a command instructing the AGV to perform an emergency stop.

The generic assumptions for this use case are as follows:

- Remote control system (guidance system) functionality is hosted in the edge infrastructure.
- Only communications between mobile robots and the edge computing infrastructure are considered (individual mobile robots don't communicate with one another directly). In other words, the edge controls a number of mobile robots and communicates bidirectionally with each one.
- Uplink communications (from a mobile robot to the edge) include (1) process-related measurement data and (2) video or image data.
- Downlink communications (from the edge to a mobile robot) include (1) process data for controlling and managing mobile robots and (2) emergency stop commands (such as the alarm command).

The following analysis sheds light on the relevance of edge computing by considering a case in which GCS functionality can reside in the remote cloud.

8.2.2 Requirements for the Mobile Robot Use Case and Edge Capabilities for Meeting Them (Steps 2 and 3)

The service performance requirements for this use case are presented in TS 22.104 [4], A 2.2.3. The most important KPI values are an end-to-end latency of between one and 10 ms for periodic machine control communication, communication service availability greater than 99.9999%, and service reliability equivalent to a mean time between failures of about 10 years with periodic machine control communications. These values should be achievable for a large number of mobile robots (up to 100 per square kilometer). The video streaming uplink requires a data rate above 10 Mbit/s. High bandwidth is essential for this use case, since the video streams from the cameras mounted on all the robots can generate heavy network traffic.

Two additional requirements for this use case are data privacy and security, which are covered in the 5G-ACIA White Paper "Key 5G Use Cases and Requirements" [6] and 3GPP TS 22.261.

Summing up, the requirements for this UC (the second step of the method) include *end-to-end latency, reliability, availability, data privacy, security,* and *high bandwidth*.

From the list of edge computing capabilities and features (see section 8.1 above), we have selected the following three capabilities (the third step of the method) that may potentially help meet the requirements of this use case: *local data processing, local data storage,* and *local data aggregation/ collection.*

Local data processing: The edge application locally processes video data or images received from cameras on mobile robots and control data received from sensors on mobile robots.

Local data storage: Video data or images received from the cameras on mobile robots and control data received from sensors on mobile robots can be stored locally in the edge. The stored data can be used for statistical purposes, caching, analyses, etc.

Local data collection and aggregation: When there are many mobile robots, the images, video data, and measurement data are locally collected and aggregated for processing in the edge.

8.2.3 Benefits of Edge Computing for the Mobile Robot Use Case (Step 4)

To illustrate the achievable benefits, the fourth step of the method maps edge computing capabilities onto relevant use case requirements.

1) Local data processing helps meet the following requirements:
   - Reduced E2E latency (benefit #1):
     There is no need to send images, video data, or control information deeply into the core network for processing. This results in a shorter end-to-end roundtrip latency for communications between the edge and mobile robots. Note that low latency is very important for control commands in general and especially for alert commands such as emergency stop.
   - Improved reliability and availability (benefit #2):
     The edge is hosted within one or more reliable enterprise environments. This precludes failures induced by external events. The communication data path (TS 22.104) to the edge infrastructure is shorter than that to the remote central cloud, which reduces the risk of failures.
   - Improved data privacy (benefit #3):
     The edge infrastructure is located inside a closed enterprise environment (a private network). This ensures the safety of sensitive private information flowing between the edge infrastructure and mobile robots.
   - Improved security (benefit #4):
     The communication link between the edge cloud and mobile robots runs within an enterprise environment (a private network controlled by OT) and is therefore secure. In contrast to this, a communication link between mobile robots and a remote central cloud includes an external (third-party) network that may not always have guaranteed security.

2) Local data storage helps improve privacy similarly to the local data processing feature (benefit #4):
   Edge computing data is stored within a closed enterprise environment (a private network) for later use for statistical purposes, caching, analysis, etc. This ensures that sensitive private information is kept locally where external third parties can't access it.

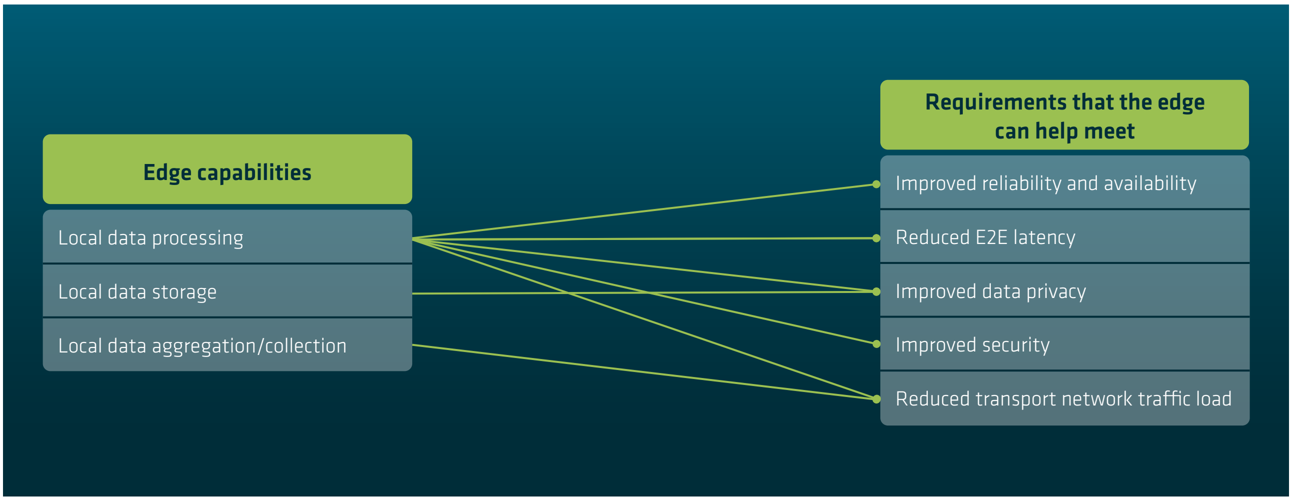
3) Local data collection and aggregation help meet the need to save bandwidth (benefit #5):
   The cameras of mobile robots produce a considerable volume of video streaming data traffic (a mobile robot like that in use case 4 in A 2.2.3 of TS 22.104 can generate more than 10 Mb/s). Collecting and aggregating these data streams from multiple mobile robots and localizing them to the edge infrastructure instead of remotely routing them to the central cloud uses less bandwidth and significantly reduces the network load.

8.2.4 Summary: the Relevance of Edge Computing to Mobile Robot Use Cases

Edge computing capabilities can help meet **five requirements** that have been defined for mobile robot use cases by 3GPP TS 22.104, TS 22.261, and the 5G-ACIA white paper "Key 5G Use Cases and Requirements" [6]: end-to-end latency, reliability and availability, reduced network traffic load (bandwidth savings), security, and data privacy.

Figure 17 shows the benefits that taking advantage of different edge capabilities can yield in connection with mobile robots/AGVs. This use case can therefore be considered a prime candidate for deriving benefits from edge computing.

Figure 17: Relationship of edge capabilities and benefits for mobile robots use case



Source: 5G-ACIA / ZVEI e. V.

9 References

[1] 5G-ACIA White Paper, *“5G for Connected Industries and Automation,” 2nd edition,* published in March 2019.

[2] 5G-ACIA White Paper, *“5G Non-Public Networks for Industrial Scenarios,”* published in July 2019.

[3] 3GPP Technical Report 22.804, *“Study on Communication for Automation in Vertical Domains,”* Rel. 16.

[4] 3GPP Technical Specification 22.104, *“Service requirements for cyber-physical control applications in vertical domains”*, Rel. 18.

[5] 3GPP Technical Report 22.832, *“Study on enhancements for cyber-physical control applications in vertical domains,”* Rel. 17.

[6] 5G-ACIA White Paper, *“Key Use Cases and Requirements,”* published in May 2020.

[7] 5G-ACIA White Paper, *“Integration of 5G with Time-Sensitive Networking for Industrial Communications”*, published in June 2021.

[8] 5G-ACIA White Paper, *“Integration of Industrial Ethernet Networks with 5G Networks,”* published in November 2019.

[9] 3GPP Technical Specification 23.501, *“System architecture for the 5G System (5GS),”* Rel. 17.

[10] 3GPP Technical Specification 23.401, *“General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access,”* Rel. 15.

[11] ETSI White Paper no. 28, *“MEC in 5G networks,”* published in June 2018, ISBN no. 979-10-92620-22-1.

5G-ACIA White Paper
Industrial 5G Edge Computing –
Use Cases, Architecture and Deployment

Contact
5G Alliance for Connected Industries and Automation
(5G-ACIA), a Working Party of ZVEI e. V.
Lyoner Strasse 9
60528 Frankfurt am Main
Germany
Phone: +49 69 6302-209
Fax: +49 69 6302-319
Email: info@5g-acia.org
www.5g-acia.org

Published by
ZVEI - German Electro and
Digital Industry Association
5G Alliance for Connected Industries and
Automation (5G-ACIA), a Working Party of ZVEI
www.zvei.org

Published in February 2023
Design: COBRAND Berlin

10 5G-ACIA Members

As of January 2023

KETI
KEYSIGHT TECHNOLOGIES
Lenovo
LS telcom
MAVENIR
MC TECHNOLOGIES
MITSUBISHI ELECTRIC
MOXA
MUGLER TELCO NETWORKS.
Nitto
NOKIA
NXP
ONELAYER
orange
Panasonic
PEPPERL+FUCHS
PHOENIX CONTACT
Qualcomm
RELIABLE REAL-TIME RADIO
Radisys
ROHDE&SCHWARZ
s&t ISKRATEL
SAL SILICON AUSTRIA LABS
salzburgresearch
Schneider Electric
SICK
SIEMENS
SINTEF
SoftBank
SONY
ST life.augmented
Telefónica
T · ·
telent
TRUMPF
ublox
UNIBERG
UPLINK
verizon
VIAVI Solutions
vodafone
VOLKSWAGEN AKTIENGESELLSCHAFT
WAGO
Weidmüller
XITASO
YOKOGAWA
ZTE

5G-ACIA.org