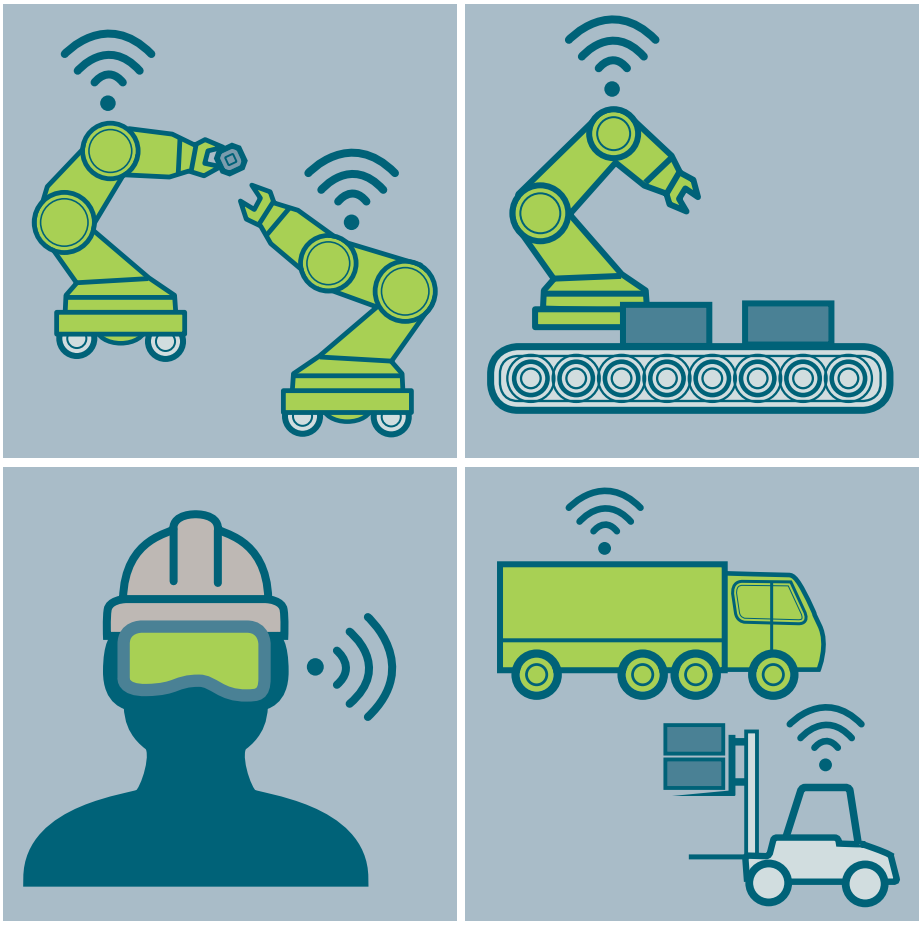
5G ACIA
5G Alliance for Connected Industries and Automation

White Paper

Integration of Industrial Ethernet Networks with 5G Networks



November 2019

5G Alliance for Connected Industries and Automation

Integration of Industrial Ethernet Networks with 5G Networks

Contact:
Email: info@5g-acia.org
www.5g-acia.org

Published by:
ZVEI – German Electrical and
Electronic Manufacturers' Association
5G Alliance for Connected Industries and Automation
(5G-ACIA), a Working Party of ZVEI
Lyoner Strasse 9
60528 Frankfurt am Main, Germany

www.zvei.org

November 2019

Graphics: ZVEI

Contents

1 Introduction

The objective of this white paper is to highlight the specific requirements associated with the integration of established industrial Ethernet networks with a 5G network and to identify possible solutions.

Industrial communication network technologies based on ISO/IEC/IEEE 8802-3 (as defined in [7] IEC 61784-2 and specified in [6]) are used in industrial manufacturing and control to enable real-time communication between controllers, machines and within machines at routing, control and sensor level. There are multiple protocols defined for industrial communication networks specifically based on Ethernet IEEE802.3, e.g. EtherCAT, Sercos III, Ethernet/IP, PROFINET, CC-Link IE Field or Modbus TCP. These protocols are therefore known as industrial Ethernet networks.

But compared to standard IT communication these technologies have specific needs e.g.

- Forwarding on Layer 2 (MAC, Ethertype, VLAN) – without IP
- Short communication cycle times and concurrent communication cycles with divergent times
- Typically, a large number of short Ethernet frames
- Highly precise synchronization
- Transmission of functional safety protocols (see IEC 61784-3)

Section 2 provides an overview of communication network technology use cases of the kind typically encountered in today's industrial scenarios. Building on this overview, section 3 describes the characteristics of existing industrial Ethernet solutions. Section 4 then presents a reference model for the integration of these wired technologies with 5G. The model provides insights into how data streams are transmitted and/or where they are terminated, how QoS is managed, how clock synchronization is achieved, etc. In addition, it provides potential users with guidance on how to implement 5G within an existing or a green-fields industrial communication network.

The 3rd Generation Partnership Project (3GPP) is a collaborative project that brings together standardization organizations from around the world to create globally acceptable specifications for mobile networks. As its name implies, it was first created to establish such specifications for the third generation (3G) of mobile systems. It has continued its work for subsequent generations, including the one considered here, the fifth generation (5G).

This paper refers to technical specifications (TSs) published by 3GPP, i.e. the 5G standards.

2 Industrial automation control applications

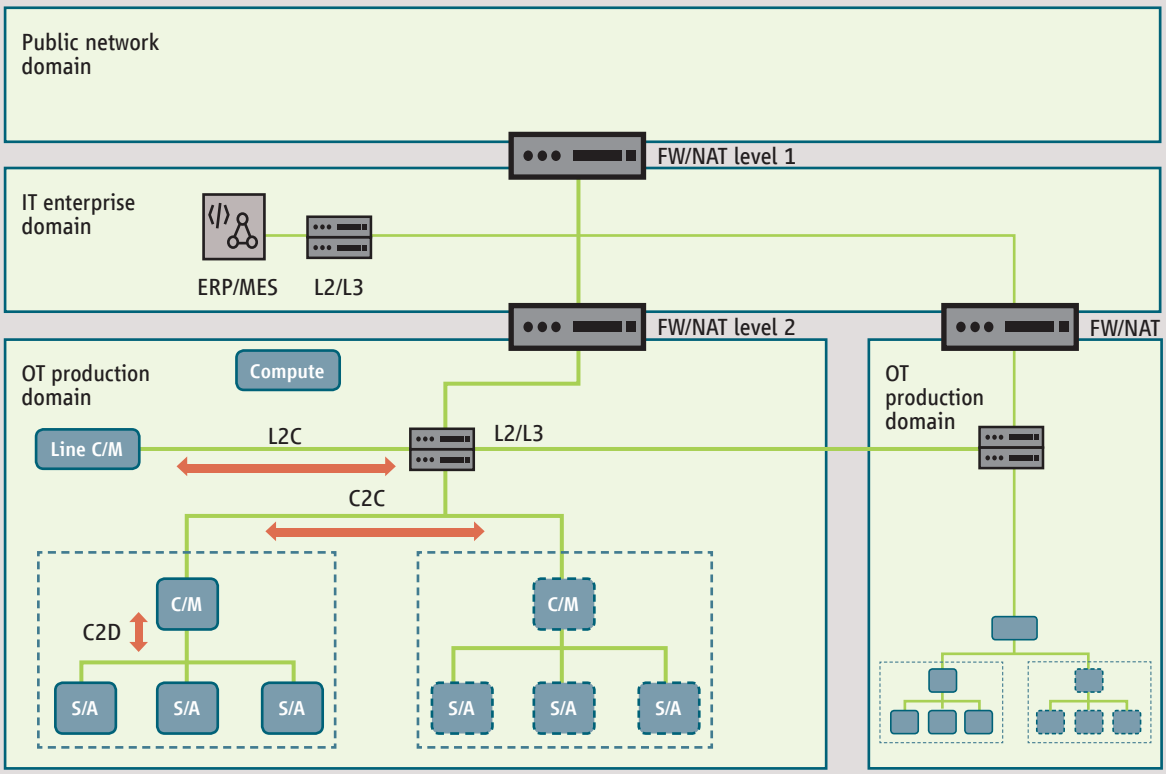
3GPP TS 22.104 [1] addresses a challenging category of vertical applications, namely cyber-physical control applications, which require very high communication service availability. Real-time Ethernet is one of the most established wired communication technologies for cyber-physical control applications, and TS 22.104 identifies requirements that 5G systems must meet to support it.

This white paper focuses on the use cases described in the 5G for Automation in Industry white paper [4], including their diverse communications network requirements. These use cases are as follows:

- Motion control
- Control-to-control communication
- Control-to-sensor/actuator communications
- Mobile robots
- Remote access and maintenance
- Closed-loop process control
- Process monitoring
- Plant asset management

In current industrial communications scenarios, the shop-floor IT infrastructure is separated from the enterprise IT by means of security gateways that perform firewall (FW) and network address translation (NAT) functions. Furthermore, the enterprise IT is separated from the Internet or WAN by a further layer of security gateways.
A topology of this type is given in Figure 1, which shows the public network domain, the

Fig. 1: Industrial automation communication structure



Source: 5G-ACIA / ZVEI

enterprise domain and multiple OT production domains, which may be physically segregated but belong to a single enterprise. Communication between two or more OT production domains may be via direct VPN connections traversing the FW/NAT gateway or indirectly via applications residing in the enterprise domain (e.g. ERP/MES).

Communication between L2/L3 infrastructure and and/or controllers/master (C/M), sensors and actuators (S/A) is predominantly wired, employing communications technology as described in the following section of this document.

Based on the main use cases described in [1], a number of use cases have been identified and detailed below. These represent the various communication paths and methods applicable to a generic 3-layer infrastructure.

2.1 Use case 1: line controller-to-controller (L2C) and controller-to-controller (C2C) communication

A production line comprises a series of machines that perform production tasks, plus a line control unit. Preconfigured control units for individual machines, with tested and approved internal communication, communicate with a supervisory control unit of the production cell or line.

Typically, the network connecting the machines has no fixed configuration, i.e. does not require certain controls to be present. The control nodes within the network often vary according to the current status of machines and the manufacturing plant.

Machines in a production cell may be identical but each performing a specific task (machine cloning). If for examples multiple machines, e.g. robots, are connected in a line, their outside interface to the line control must be configured. But all robots have the same type of control unit and internal network architecture.

The number of machines on a production line can vary from just a few to hundreds. The typical line size is between 10 and 30 machines.

For upstream/downstream data exchange between machine modules within a production domain (see Figure 1) the control units of the machine modules have to exchange data with other machine modules (controller to controller, C2C). Each machine may run its own unique data exchange cycle and the intervals may also differ. 5-10 machine modules can be assumed to be within a service area of 100x30x10m. Communication between controllers in a distributed control system (DCS), which is typical for closed-loop process control in the process industry, is a further example of controller-to-controller communication.

2.2 Use case 2: controller-to-device (C2D) communication

A machine has typically a control unit and several field devices (I/O boxes). Field devices can have inputs for sensors (S) and/or outputs for actuators (A). Typically, the machine controller (PLC or motion controller) has 1:n bidirectional communication relationships with (a set of) field device(s), e.g. sensors, actuators, drives.

Closed-loop-control via the network requires cyclic synchronous exchange of set-points from the control unit with feed-back values from the devices, e.g. a drive.

A typical machine configuration entails 10 to 20 field devices connected to several hundred sensors and actuators. Larger machines can have 100 and more field devices within a service area of 50mx50mx10m.

An automated production facility may have a large number (up to several tens of thousands in total) of sensors and actuators distributed over a large area for closed-loop process control and process monitoring. In a typical DCS configuration, each distributed controller may have control over hundreds of sensors and actuators via 10 to 20 I/O boxes.

In some machines there are functions that are either only optionally available (optional modules) or that may be mounted, for example, on a rotating part of the machine. The resulting sub-functions are integrated into the machine via sub-networks. From the control point of view, these sub-networks are treated as part of the overall machine network, with the requirements defined by the controller-2-device network.

2.3 Device-to-compute (D2Cmp) communication

Asset management functions for devices used in process automation, such as

- Inventories
- Firmware updates
- Condition monitoring and predictive maintenance
- Configuration backups
- Data analytics

are not usually managed by the network controller (PLC or DCS). These functions, typically implemented across the entire production facility or cloud-based, require secure access from outside the control network to devices inside the control network bypassing the controller.

As the machine controller is optimized for time-critical control algorithms and efficient communication with sensors and actuators, too much non-control-relevant traffic would place an unnecessary load on the controller. In wired installations, a gateway is used to perform these tasks – bypassing the controller.

The above-mentioned applications simply support the primary use case (“process control” or “safety monitoring”) performed by the controller but must not hinder it. This means robust device access with no impact on real-time traffic between devices or between device and controller.

The following network management services are required:

- Network management services
  - Routable layer 3
  - Segmentation
  - IP address management
  - Device discovery service (network scan)
- Security (authentication, authorization).

As a cloud-based application can access a device, the same device can access a cloud service, e.g. to check the availability of a new firmware version or to actively backup its configuration, i.e. on its digital twin.

Whether the connection is established by compute entity or by the device itself, it needs to be configurable in a secure way and with the ability to activate / deactivate the connection locally in the production facility or remotely by compute.

In addition, the engineering, management and diagnostic capabilities of the industrial Ethernet network must be suitable for integration with the 5G network.

3 Industrial Ethernet network principles

Industrial Ethernet networks, [6] and [7], are used in industrial manufacturing and process automation to enable real-time communication between controllers, machines and within machines at routing, control and sensor level. The use of Ethernet for real-time communication requires modifications in the use and/or implementation of Ethernet networks. This section describes the characteristics of existing industrial Ethernet solutions.

3.1 Principles of real-time communication

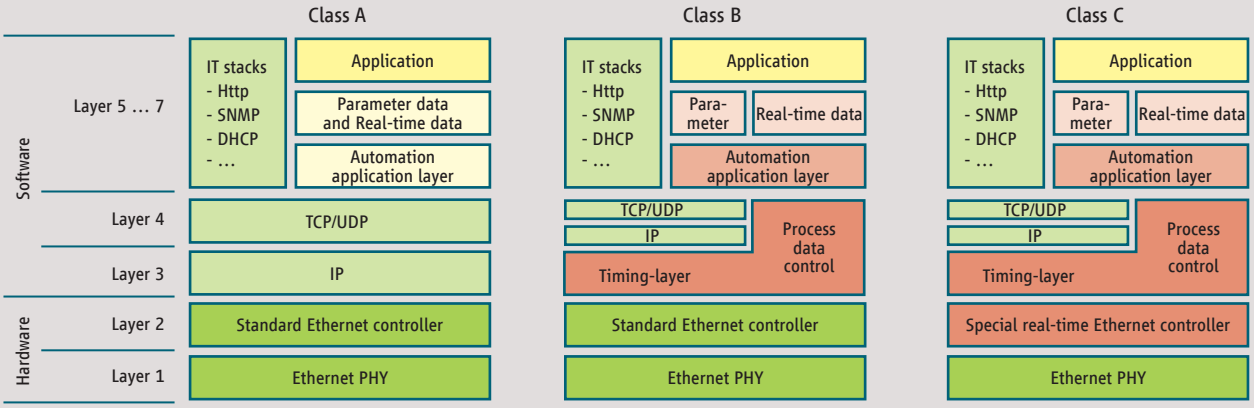
[5] describes a generic three-layer communication model:

- Application function: information model and software modules for the actual business function
- Middleware (shown as the automation application layer in Figure 2):
- this provides generic communication services, e.g. read/write, publish/subscribe, connect, browse, etc.
- Transport-oriented protocols: this provides a communication system that guarantees fulfillment of application requirements in terms of reliability, availability, real-time behavior, etc.

Depending on real-time and cost requirements, the industrial Ethernet network technologies defined in [6] employ differing implementation approaches to Ethernet and/or the IP layer in order to guarantee deterministic communication. Three performance classes of industrial Ethernet network technologies can be identified. Network performance typically improves from class A via class B to class C.

- **Class A** uses commercial, off-the-shelf Ethernet hardware and standard TCP/IP software stacks for process communication. Class A technologies are also referred to as best-effort-approaches.
  Examples: Ethernet/IP, Modbus TCP, Foundation Fieldbus
- **Class B** also employs standard Ethernet controllers, but does not use TCP/IP for process data communication. Instead, it deploys a dedicated process data protocol which is transmitted directly within the Ethernet frame.
  Examples: PROFINET RT, POWERLINK, EtherCAT Automation Protocol (EAP)
- **Class C** employs dedicated hardware down to Layer 2 (at least on the slave side)
  Examples: EtherCAT, PROFINET IRT, Sercos III, CC-Link IE,

Fig. 2: Basic industrial Ethernet network approaches



Source: 5G-ACIA / ZVEI

The industrial Ethernet network is not only optimized for real-time communication needs, but also for use on a machine: simple configuration and setup is required, with actionable diagnosis information to reduce machine set-up times and downtime in the event of faults. Typically, industrial machines operate 24/7, which requires 100% network availability.

3.2 Configuration principles

The logical communication relationships use communication paths on the end devices and infrastructure components that require configuration of the network.

For production facilities that deploy multiple machines of the same type in a cell or on a line, addressing the device may be more challenging. Within the machines, the same VLAN or IP addresses are used, and a different address can only be configured for line communication outside the machine (machine cloning).

Typically, the following mechanisms for address assignment are used in industrial Ethernet networks:

IP connections (layer 3)

Communication is established on or above the IP layer. Devices have an assigned IP address or retrieve the IP address during boot-up, e.g. via DHCP. With this information and additional services, such as ARP, the corresponding MAC addresses (L2) can be determined and layer 2 communication can then be employed for cyclical data exchange.

Some protocols use their own resolution protocol. PROFINET, for example, employs a name resolution service called DCP (Discovery and Configuration Protocol) that provides MAC address information for a configured name of the device.
Examples: EtherCAT Automation Protocol, Ethernet/IP

Bridged[1] networks (layer 2)

Communication is established directly on MAC address layer. A sender device can be configured to send the data as multi- or broadcast messages that all devices can receive. This is used to publish messages and allows direct device-to-device communication.

[1] Bridges according to IEEE802.1 are often called Switches

It can also be configured as a unicast message to have a dedicated connection between a sender and a device. This can be used e.g. as a poll request message. The receiver can accept multi- or broadcast destination addresses or unicast frames only.
Examples: EtherCAT Automation Protocol, POWERLINK, PROFINET

Non-bridged networks
Some industrial Ethernet networks do not use the address information given in the Ethernet frame, i.e. the MAC address. The specific address is part of the payload data within the Ethernet frame and processed by the devices connected to the network. Typically, the controller is connected directly to the network segment, i.e. without the usage of bridge devices.

If the controller is nevertheless connected to the industrial Ethernet network segment via a standard heterogeneous Ethernet network, the Ethernet frame must be routed through the standard network. The first device of the industrial Ethernet segment then requires L2 or L3 addressing capabilities. This is used, for example. to separate the controller from the network segment by means of a heterogeneous network using Time-sensitive networking (TSN) functionality defined by IEEE and IEC, [8] and [9].
Examples: EtherCAT, Sercos III

3.3 Periodic data exchange

Periodic process data exchange between the controller and the devices or between controllers usually consists of small frames: from the controller to each device to send the output data and from the devices to the controller or to other devices to read the input data. In other words, typically two periodic data streams are needed between the controller and each device. Short communication cycle times in the range of 10ms…1ms…0.5ms are common. Since differing application tasks may need differing information within the network, diverse communication cycle times are employed on network and within devices. An arbitration method is needed to avoid collisions or queues and to guarantee latency limits are not exceeded. The following mechanisms for media access control are used:

- Polling
  A client (the controller) sends a poll request frame, the server (network device) responds with a poll response frame. The client polls all servers sequentially.
  Examples: Modbus-TCP, POWERLINK, Ethernet/IP, EtherCAT Automation Protocol
- Cyclical, time-triggered transmission
  Each node sends its data cyclically, triggered by a local timer. The devices on the network run asynchronously, i.e. network load fluctuates.
  Examples: PROFINET RT, EtherCAT Automation Protocol
- Isochronous time-triggered transmission:
  Each device has a configured time slot for data transmission. Deterministic data is sent within a certain time window. Non-deterministic data can be sent in a second time window. Precise time synchronization across all devices is a prerequisite.
  Examples: PROFINET IRT
- Processing-on-the-fly with one frame for all devices
  The controller sends a single Ethernet frame (L2) to the network segment. The frame is processed on-the-fly within each device while being forwarded to the next device. Only the master actively transmits new frames, which avoids any collisions or delays. The device applications can run asynchronously or can be synchronized.
  Examples: EtherCAT, SERCOS III

3.4 Aperiodic traffic

Aperiodic traffic is used to transmit configuration and diagnosis information between network-connected devices. Transmission is typically trigged by an event, e.g. an exceeded temperature threshold (process event), a device malfunction (diagnosis event) or a lifetime counter (maintenance event).

Generally speaking, the mechanisms used in this context are aperiodic or have such long cycle times that they can be considered, from a networking perspective, sporadic.
[1] describes communication service performance requirements for aperiodic traffic.

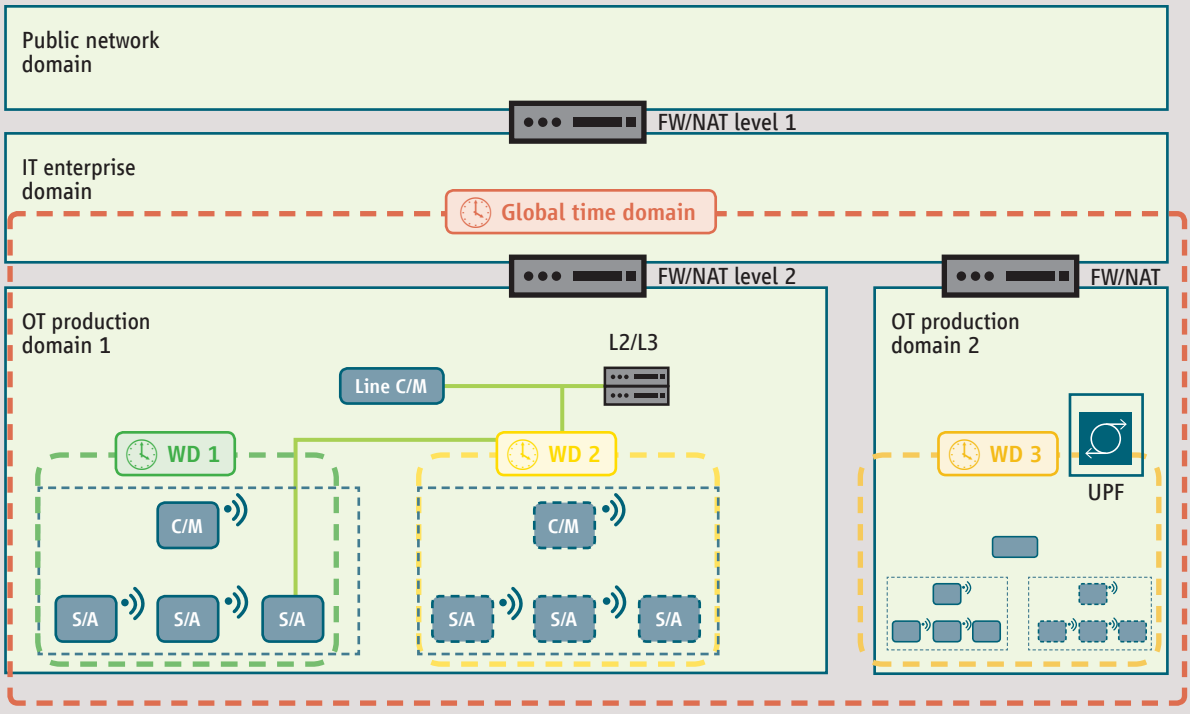
3.5 Synchronization methods

Synchronization of the I/O devices and the controller is required for various use cases.

- Time-triggered transmission of isochronous process data, see 3.3
- Time stamping, e.g. for sequence of events (SoE) or for distributed data capture

Synchronization requirements vary from very stringent, e.g. in motion control, to relatively lax, e.g. SoE tracking. To meet these diverse synchronization needs efficiently, the network can be divided into time domains.

Fig. 3: Time domains on the factory shop floor



Source: 5G-ACIA / ZVEI

The global time domain is used for overall synchronization within the production facility. Synchronization accuracy is typically ≤ 1µs. If very precise time synchronization can instead be provided by a working clock domain, an accuracy of ≤ 100 µs may be sufficient for the global time domain. Clock synchronization in the global time domain usually applies to all automation devices within the production facility. In other words, the global time domain usually applies to the entire factory shop floor or even multiple buildings, see Figure 3.

The working clock domain is used to synchronize specific applications or devices. Working clock domains are limited in size to ≤ 100m x 100m [1]. They often consist of a single machine or a group of neighboring machines that physically collaborate. The restricted size allows very precise time synchronization (≤1µs), with a tendency to increase accuracy to 100ns for some use cases. Robots, motion control applications, numeric control, and any kind of clocked / isochronous applications rely on the working clock domain to ensure actions are precisely coordinated..

A global time domain usually contains multiple working clock domains. Devices may participate in multiple time domains, leading to overlap between working clock domains.

It should be noted that the global time domain and the various working clock domains may have varying timescales, and synchronization accuracy and precision. Additionally, the clocks of these various domains are driven by multiple grand master clocks which cannot be expected to be synchronized.

The following mechanisms are employed for synchronization within industrial Ethernet networks:

Distributed clock synchronization

A time data from the master clock is distributed to its slave clocks via a protocol. Where every device has the same time, synchronized transmission and time stamping are possible. IEEE 1588 or derivatives or profiles, e.g. IEEE802.1AS-Rev, use IP multicast Frames or Layer 2 frames for time data distribution. Typically, the same delay is assumed for the sent sync message and the returned Delay_Req message. The variable residence time of the bridges is measured and added into a message correction field to improve accuracy.

Some industrial Ethernet networks apply their own principles to the distribution of time information from the master clock. EtherCAT, for example, uses specific commands in the payload data of the Ethernet frame to synchronize devices. The propagation delay between and within the devices must have low jitter to enable adjustment of time deviations between devices.

Synchronization frame from the controller

The controller transmits a synchronization frame at the beginning of each communication cycle. All devices adjust their local time upon receiving this frame. The synchronization frame must be transmitted high precisely with low jitter.

4 Reference model for integration of industrial Ethernet networks with 5G networks

4.1 5G system description

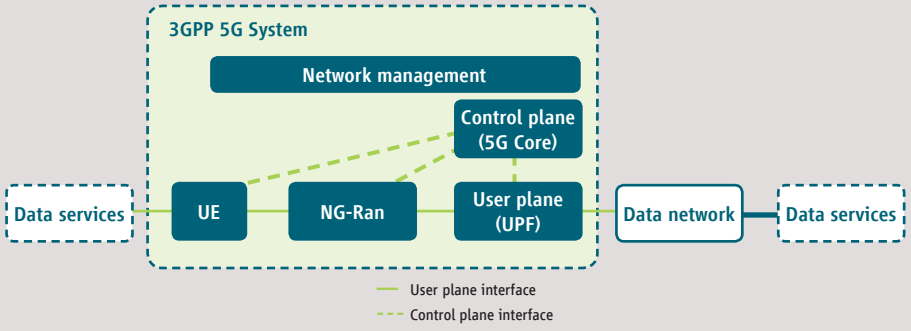
Figure 4 shows the high-level architecture of a 3GPP 5G system. The main components are

- the user equipment (UE),
- the radio access network (NG-RAN),
- the control plane (composed of multiple network entities not shown in the figure),
- the user plane, which connects with the data network, such as a factory's local area network,
- and the network management layer, which is responsible for provisioning and managing of the network and its services.

3GPP 5GS introduces the division into a control plane and a user plane function (UPF), allowing more scalable and flexible operation of the 5G network. The task of the user plane is to manage data exchange between the UE and the data network.

Furthermore, the radio access network is centrally controlled by the control plane in order to efficiently operate the 5G network, enabling seamless device mobility, access control, policy application, and other essential features. To a user of a 5G network, the internal structure is hidden, i.e., only layer 3 or layer 2 connections are visible while the actual control plane, user plane or radio access node configuration and settings are not relevant for network integration. Furthermore, it is important to understand that 3GPP is currently specifying functions relevant to wireless communication. In this regard, the UE contains a 3GPP-related protocol stack. Dedicated 3GPP working groups are currently defining interfaces between the UE and the application layer.

Fig. 4: High-level 5G system description



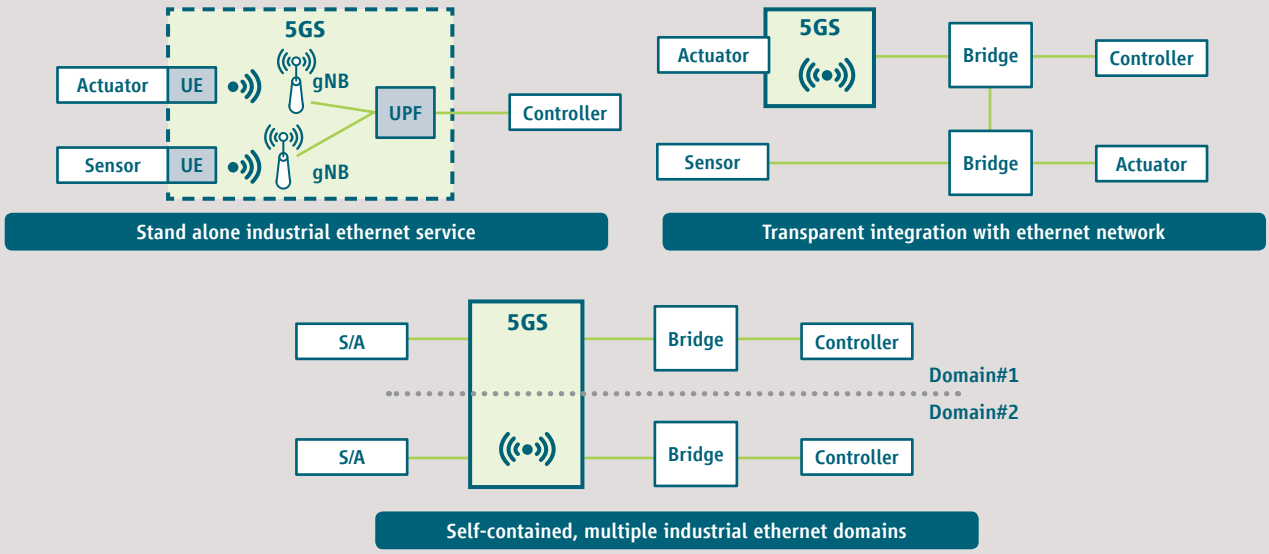
Source: 5G-ACIA / ZVEI

4.2 Integration of 5G network with the industrial Ethernet network

4.2.1 General integration scenarios (logical view)

This section describes the integration of a 5G network with an industrial Ethernet network. A distinction is made between bridged and non-bridged networks, as described in section 3.2.

Fig. 5: Integration scenarios



Source: 5G-ACIA / ZVEI

There are a variety of possible ways of for integrating a 5G network with an industrial Ethernet network; three very general options are illustrated in Figure 5. One is a stand-alone deployment where the 5G network directly supports and integrates all sensors, actuators, and controllers by connecting them through 5G user equipment (UE) or directly on the 5G network side (UPF). A system of this kind could be leveraged to integrate multiple base stations (gNBs) in order to increase wireless coverage. However, this approach is very limited in terms of interoperability, and would be limited to those cases where, for instance, a controller acts as proxy, e.g., between machine and production cell.

In order to enable seamless interoperability with existing wired networks, the 5G network can be integrated with an existing industrial Ethernet network transparently to the payload. The 5G network will then be perceived by the wired networks as Ethernet bridges and links. The challenge associated with this approach is to ensure interaction between the control and user plane layers to enable end-to-end industrial Ethernet network ervices, e.g.,

- end-to-end integration into the industrial network control and management process (see section 4.2.2 for an example using TSN),
- end-to-end integration into industrial automation network hierarchies (see section 4.3 on deployment models),
- or end-to-end time synchronization (see section 4.4).

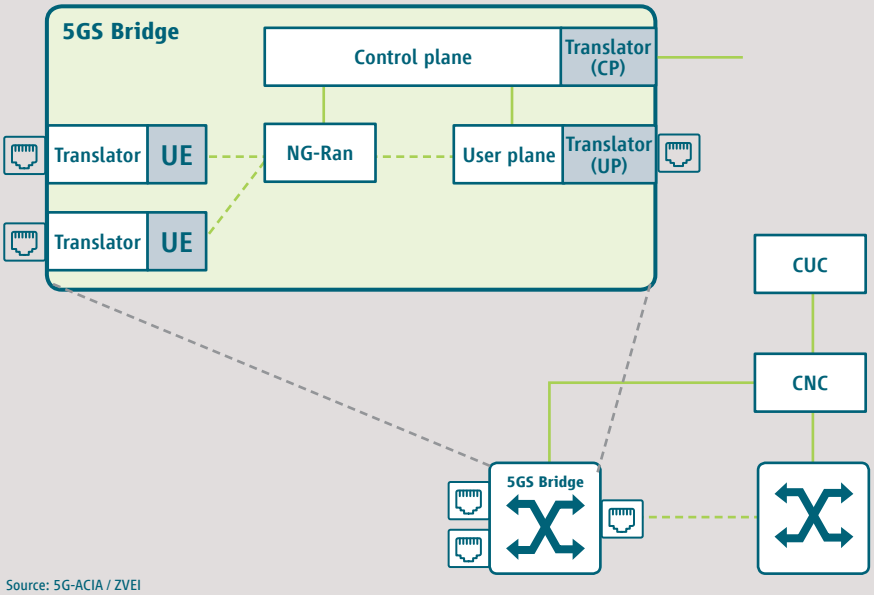
Finally, a single 5G network can serve multiple industrial Ethernet networks, e.g. VLANs, which are logically segregated, and therefore isolated in terms of QoS and privacy.

4.2.2 Integration models for bridged networks

There are two fundamental approaches to the integration of 5G network with a bridged industrial Ethernet network. The first option is to model the 5G network as an Ethernet link. In this case, the 5G network behaves like a "cable" with deterministic performance in terms of capacity and delay. While "5GS as a cable" is a feasible option, it also has a number of disadvantages and constraints. These include the inability to adequately model

the dynamic QoS performance inherent to a wireless system, the very limited ability to influence the scheduling of industrial Ethernet network connections (e.g. TSN streams), and limitations on mapping individual user equipment radio connections towards the wired Ethernet network.

Fig. 6: High-level (logical) integration architecture based on 3GPP TS 23.501 for integration with TSN



Source: 5G-ACIA / ZVEI

Alternatively, the 5G network can be modeled as a logical bridge, as illustrated in Figure 6 for the specific example of integration with TSN as it is currently standardized by 3GPP. This option has several advantages and can more effectively exploit the features and QoS capabilities of the 5G network. For instance, mobile network-specific functionality remain hidden from the Ethernet network, simplifying management, i.e., as an operator of a factory only a "5GS bridge" with ports co-located with UE and user plane functions is visible, while the internal complexity of the 5G network is not visible (unless explicitly explored through 3GPP management systems).

To achieve this invisibility with regard to the industrial Ethernet network and to resemble any other bridge, the 5G network bridge provides Ethernet ports on the user plane (UP) using the "translator" functions at user equipment (UE) and using the user plane function (UPF) on the core network side. For each port of this logical bridge, the 5G network needs to support the QoS requirements of the industrial Ethernet network.

Moreover, the 5GS bridge is compliant with the corresponding IEEE 802.1 standards, e.g. IEEE 802.1AS for time synchronization (gPTP), IEEE 802.1AB for network discovery (LLDP), IEEE 802.1Q and P for VLAN/PCP tagging, and further relevant amendments for real-time traffic. Similarly, integration with any industrial Ethernet network protocol requires dedicated "translator" functions in order to enable the necessary interoperability.

4.2.3 Integration with non-bridged networks

Integration of non-bridged Industrial Ethernet networks with 5G networks is more challenging due to the way the two are operated. For instance, with Sercos III, a single Ethernet

frame is sent by a master device to a slave segment consisting of one or multiple devices. Each slave modifies part of the Ethernet frame before it forwards it to the next slave; this continues until the Ethernet frame is returned to the master.

A ring topology of this kind is very different to the structure of a 5G network, which divides communication into downlink (from base station to terminal equipment) and uplink phases (from terminal equipment to base station). A ring topology cannot be directly mapped to a 5G network. However, as detailed in section 6.3, ring or daisy-chain topologies may be partly supported by 5G networks, e.g., by wirelessly connecting the first and last device of a ring, or by utilizing the downlink phase to distribute frames in a ring and capture feedback in uplink phases. Nevertheless, these deployment scenarios are not currently subject to 3GPP standardization and therefore would, to a degree, require proprietary solutions.

Physical deployment models are explained in more detail below.

4.3 Deployment models

4.3.1 General options

When a 5G wireless network is to be integrated with an existing industrial communications infrastructure, as shown in Figure 1, the implementation of the user plane function (UPF) is of vital importance as the UPF (according to [2]) is the anchor point and therefore the logical and physical entity where the user data enters and exits the 5G network to the data network e.g. the IT/OT network or Internet. Given the multi-layer IT communication infrastructure outlined in section 3, the UPF can be deployed in any of the domains as depicted in Figure 7, Figure 8 and Figure 9.

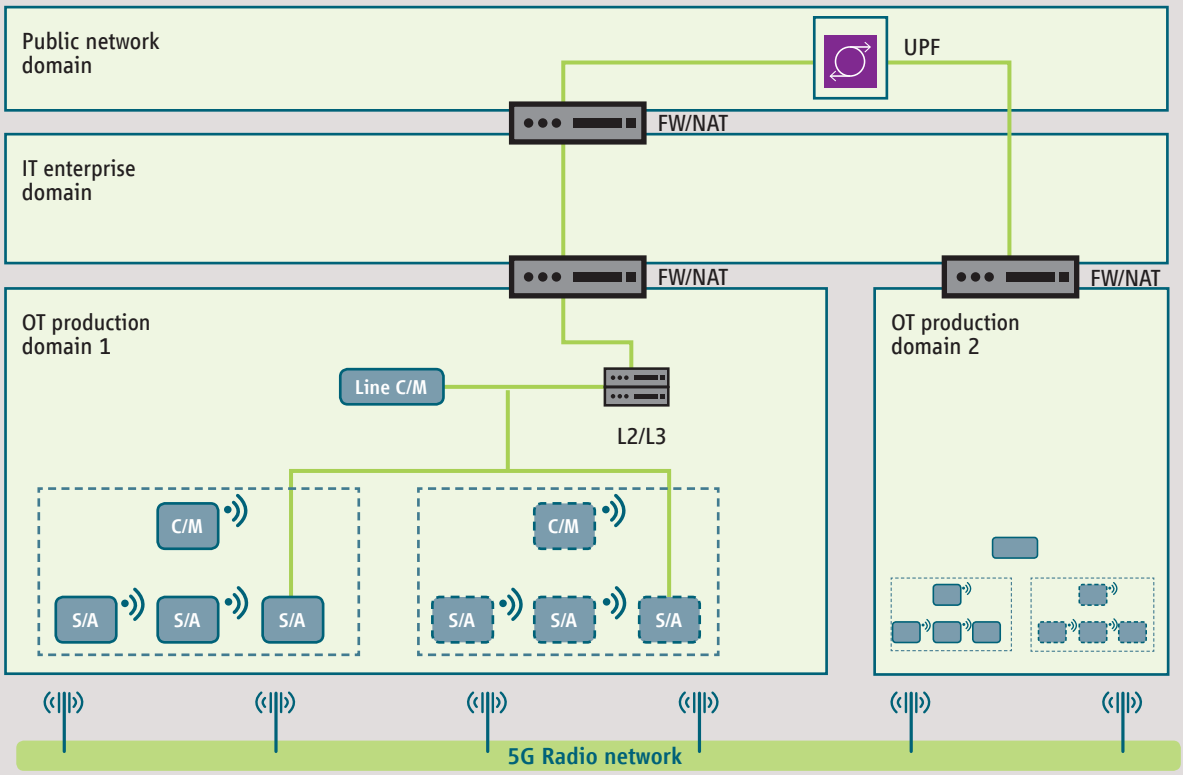
Compared to a wired legacy infrastructure, some controllers and sensors/actuators are connected wirelessly resulting in a mix of wireless/wired infrastructure in line with an expected real-life scenario.

With all deployment options given here, it is assumed that adequate radio coverage (signal strength, bandwidth, etc.) is provided either by small indoor cells or via outdoor macro cells. The radio access network (RAN) represents an independent layer with coverage for all domains that need wireless connectivity. RAN planning and design (antenna sites and characteristics, power, bandwidth etc.) is a process that must be performed on a case-by-case basis and requires knowledge of the specific enterprise building topology.

Deployment of the UPF in the public network domain as shown in Figure 7 corresponds to the 5G network scenario “NPN hosted by the public network” as described in the 5G-ACIA 5G Non-Public Networks for Industrial Scenarios white paper [3]. This option reflects the conventional network layout typically offered by a national service provider– the UPF is placed within the public network and therefore outside the enterprise premises. The advantage is moderate cost and rapid implementation, as no major additional effort is needed for the integration into an enterprise IT/OT network. The drawback is that any wireless traffic from/to devices located in the OT production domain must traverse (at least) two layers of security gateways – which may lead to complicated firewall configurations.

In most cases, a 5G network will be integrated with existing wired infrastructure. Therefore, some connections between S/As and controllers are wired (green lines) and some are wireless.

Fig. 7: Deployment option: NPN in a public network

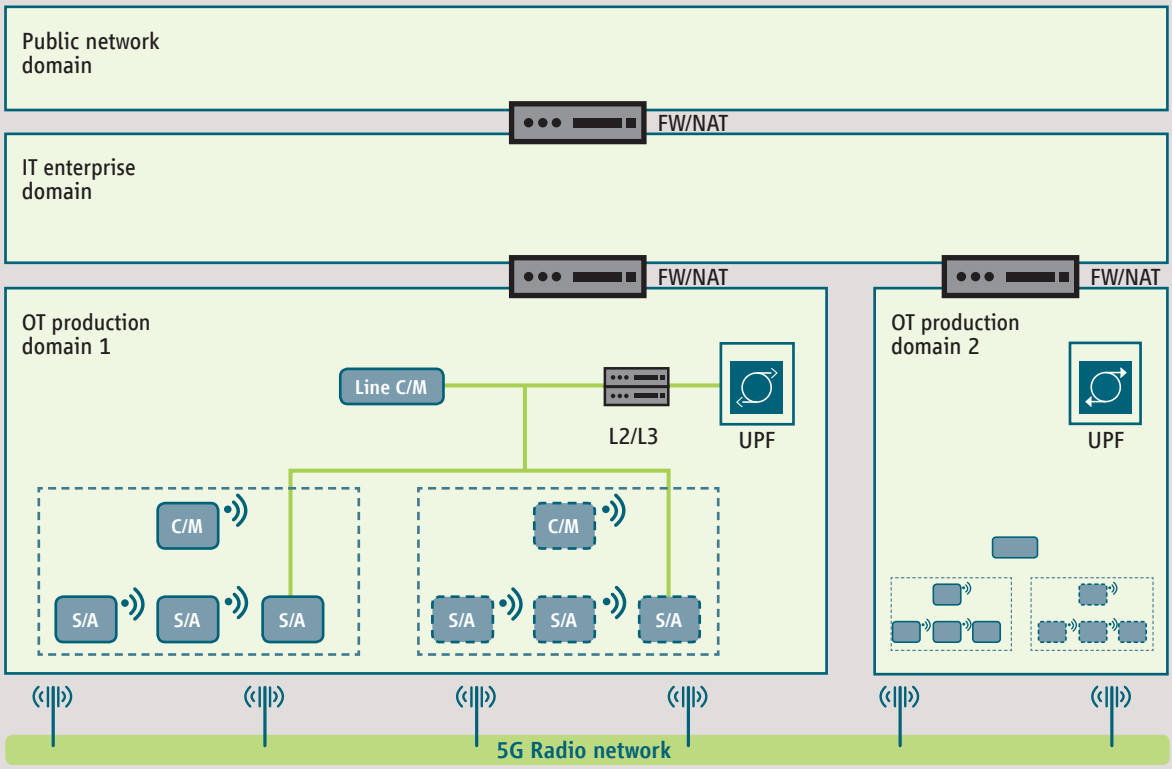


Source: 5G-ACIA / ZVEI

Deployment of the UPF in the public network may also lead to latency and availability values that are only suitable for use cases that do not require real-time communication and/or ultra-high reliability. These use cases might include, for example, remote maintenance and asset condition monitoring.

Deployment of the UPF in the OT production domain, see Figure 8, allows user plane traffic to be transmitted to the existing L2/L3 shop-floor infrastructure and does not have to traverse any security gateway. If an enterprise has multiple segregated OT production domains, a UPF node may be deployed in each domain. Alternatively, the OT production domains can be connected on L2 (Ethernet or L2 tunneling) to allow the deployment either of one UPF only, or to allow load sharing between the two UPFs, resulting in increased robustness and higher reliability. One or more UPFs are within the enterprise premises. Compared to deployment of the UPF in the public network, this option requires tight integration of 5G network elements into the existing IT/OT infrastructure. With reference to [3], this option corresponds to the scenarios "Standalone non-public network" or "Non-public network in conjunction with public networks" with shared radio access network (with or without shared control plane).

Fig. 8: Deployment option: NPN OT-UPF

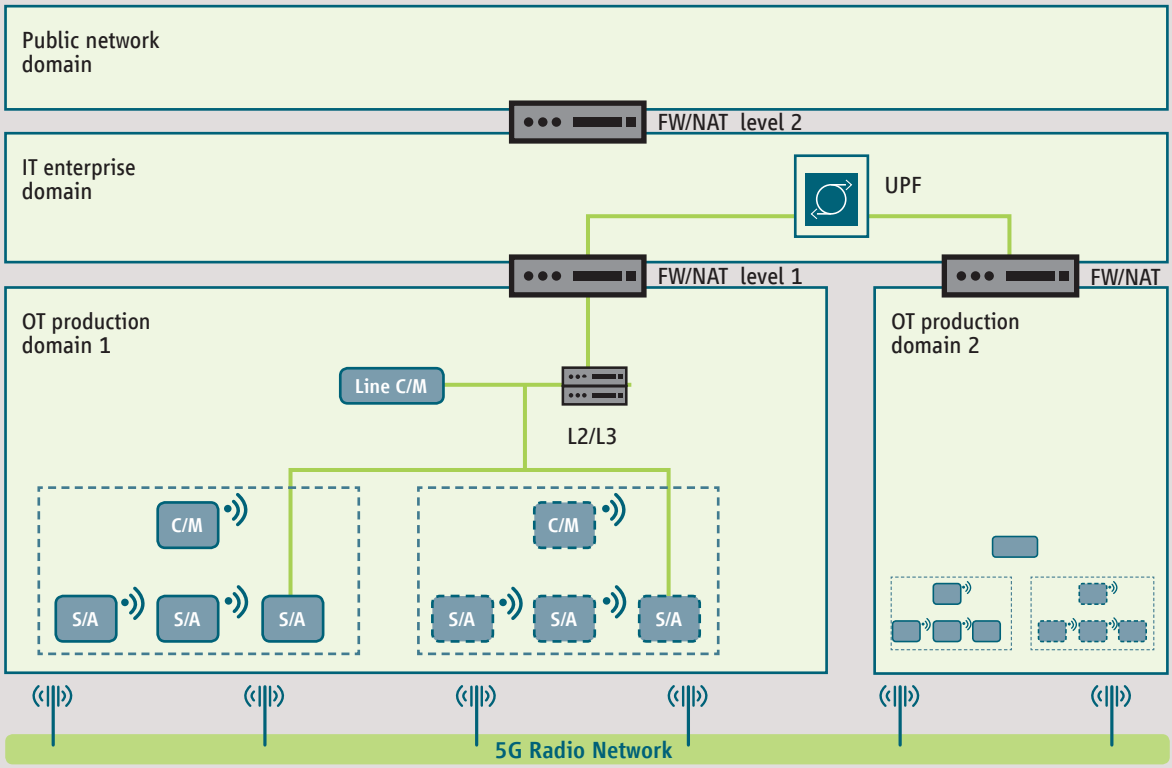


Source: 5G-ACIA / ZVEI

The best latency values and highest availability can be achieved by locating all user plane 5G network nodes (i.e. UPF) on premises and allowing it to be locally configured by the OT's engineering staff. This allows the implementation of mission-critical use cases, such as closed-loop controls and security-critical applications. This applies to both L3 and for L2 protocols presented in the previous section of this document.

The third option is to deploy the UPF node in the IT enterprise domain, see Figure 9. This constitutes a compromise between the two options given above. It requires 5G network integration with the IT layer only but user plane communication must traverse at least one security gateway (FW/NAT). This may be an acceptable approach for some smaller enterprises willing to implement and manage VPNs and L2 tunnels through the gateway. Many larger enterprises, possibly separate organizations for OT production and the enterprise IT domains, may experience greater benefit by separating the two domains more strictly.

Fig. 9: Deployment option: NPN IT-UPF



Source: 5G-ACIA / ZVEI

Use cases supported by this option are not restricted in any way. Short latency times and high availability can be achieved when the security gateways are not a bottleneck, neither from an administrative viewpoint nor from a performance and robustness perspective.

All three deployment options can be combined so that various UPFs may be spread across the OT production domains, in the enterprise IT domain and the public network domain. Depending on the required service quality and type, an application can select either of the UPFs. For instance, an application requiring critical communications characteristics can select the UPF in the OT production domain while a logistics application requiring global coverage can select the UPF in the public network domain instead. The preferred UPF can be selected by e.g. placing the UPF in a specific network slice or by assigning a specific APN/DNN for use by the application.

4.3.2 Integration of automation system topologies

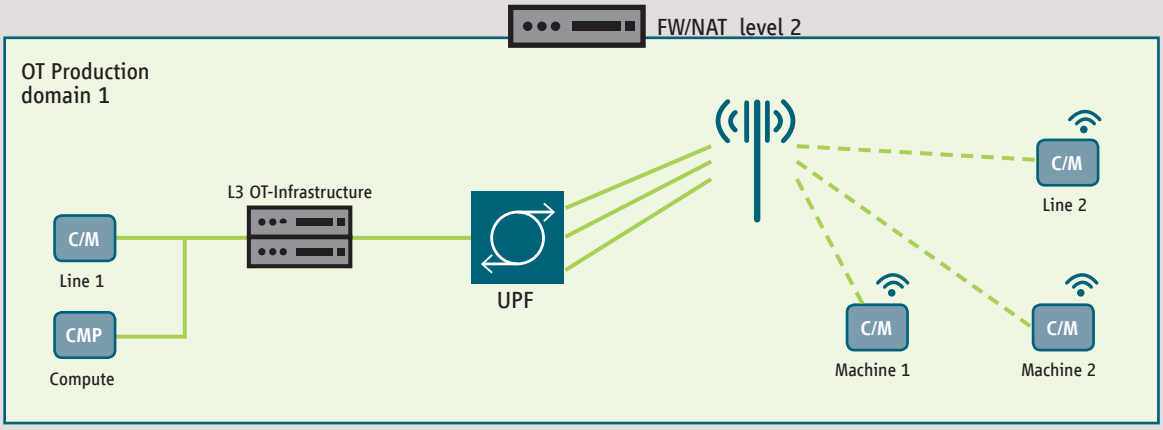
The second deployment option mentioned in section 4.3.1, where the UPF is deployed within the OT production domain (NPN OT-UPF), is assumed to be the network topology most frequently used at present for automation systems. This section considers the various communication topologies for this option in more detail. However, all considerations also apply to the other two UPF deployment options, with the corresponding drawbacks (in terms of security, GW transversal, latency and robustness).

In general, automation data traffic is transmitted between controllers (C/M) and between controllers and field devices (S/A) via IP (L3) or via Ethernet (L2), and in some cases via a mixture of L3 and L2. Some S/As have their own wireless communication interface while some do not, and need to be connected through an IO/GW. Similarly, some controllers may

have their own wireless communication interface, some may not and therefore need an IO/GW to communicate via radio links. Although this white paper does not address all possible combinations and permutations, the scenarios presented can be applied to specific use cases.

Figure 10 gives an example structure for a layer 3 connection between the C/M and the S/As. Annex A provides more details for this and the other options based on layer 2 and on layer 2 tunneling.

Fig. 10: Layer 3 connectivity



Source: 5G-ACIA / ZVEI

A specific challenge encountered within the automation industry is machine cloning, i.e. when machine configurations and parameters are copied from one machine onto many others. This may also entail the cloning of IP addresses. If multiple machines and their independent automation control systems are connected to a single UPF, IP routing will be rendered impossible if the S/As and C/Ms will be using the same address. IP address cloning is only possible in conjunction with a 5G network if the cloned machines (and the cloned S/As and C/Ms) use unique UPFs or unique APNs/DNNs.

4.3.3 Controller deployment options

In addition to the possibilities described in sections 4.3.1 and 4.3.2, the controller (e.g. PLC) can also be connected to a 5G UE, depending on the specific use case within the factory.

The various options for positioning the 5G UPF were discussed above. Another important consideration for integration of 5G into an existing or new production line is the location of the controller.

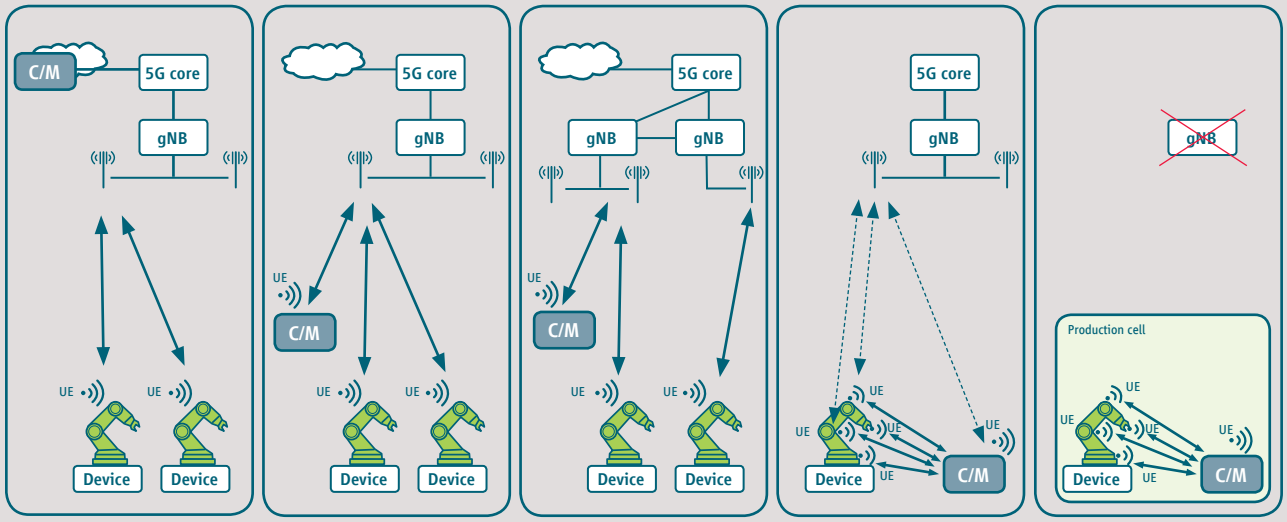
Clearly, the IO devices, drives, etc. are connected to the 5G UE. However, whether the (motion) controller should be connected to the UPF or to the UE depends on the location of the controller within the factory network architecture and how the logical functions of the UPF are distributed.

An existing controller would not usually be moved, especially when, for instance, upgrading an existing production line or cell to enable new functions, e.g. by adding wirelessly connected sensors. Moreover, it might not be possible to implement logical functions such as the UPF at the controller due to the limitations of the factory IT/OT network.

5G networks need to support flexible deployment of the controller and devices, even if, in the future, the controller will probably be located centrally in order to be part of the factory's edge computing resources. After all, there are many divergent architectures in factories and process industrial facilities – there is no single standardized architecture.

Figure 11 depicts various options for controller and device deployment. Controllers and devices may in some cases be installed within metal enclosures that can significantly degrade radio signal strength, such that critical user data is not received with the required reliability. At the same time, signal strength may be sufficient for control signals. In some cases, the production cell is fully shielded, rendering the 5G network completely undetectable. In these instances, it is possible to harness a 3GPP-specified feature called 5GNR sidelink. This allows mobile 5G devices to communicate directly with each other without any network infrastructure. The figure below depicts the latter two scenarios.

Fig. 11: Deployment options for the machine controller



Controller connected to 5G Core

Controller connected to 5G UE

Controller connected to 5G UE
Some devices are on 2nd gNB

URLLC Sidelink
Radio resources controlled by gNB

URLLC Sidelink
Simple deployment

Source: 5G-ACIA / ZVEI

4.3.4 Device-to-device deployment options using 5GNR sidelink

Based on the use cases listed in section 2 device-to-device (D2D) (5GNR sidelink) principles may seem appealing, however, their applicability to critical radio communication in factories has not yet been investigated by 3GPP.

From 3GPP Release 16, 5G will support vehicle-to-vehicle (V2V) communication for connected cars by means of the 5GNR sidelink. The sidelink was designed for the specific needs of V2V communication with the aim of enabling various levels of automated driving but still has similarities with communications with AGVs. The goal of 3GPP standardization efforts for the 5GNR sidelink is a reliability of 99.999% with a maximum latency of 3ms. Communication via the 5GNR sidelink is more predictable as most of the devices within any communication group are static and in close proximity.

The suitability of D2D for industrial communication depends on further research and standardization work (3GPP Rel 17/18) aimed at achieving the required degree of reliability.

4.4 Time synchronization

A given working clock domain, as depicted in Figure 3, will only serve a subset of the UEs within the industrial facility as a whole -- up to 300 [1]. Often, these subset UEs are connected to the same gNB. However, it is also possible for a working clock domain to span multiple neighboring gNBs. This depends on the actual use case and its vertical application, and on the deployment and wireless environment.

- The UEs can be synchronized with their respective working clock by using a 5G radio connection for transporting the time synchronization information. In this case, time precision will not fulfill industrial needs, due to uplink and downlink latency, and latency jitter of that connection.
- The UEs are synchronized using a native 3GPP-specified mechanism that ensures much greater time precision. To achieve this, the 5G network corrects synchronization messages based on known and/or measured uplink and downlink latency budgets.

Synchronization for industrial communication via the 5G system is specified in TS23.501 Release 16 [2]. This document describes 5G system features that support time-sensitive communication and allow the 5G system to be integrated invisibly as a bridge in an IEEE TSN network. This mechanism is based on a synchronized 5G system, where gNBs, UPFs and UEs are synchronized with a common 5G grandmaster clock. To this end, a mechanism has been specified by 3GPP for the transfer of the reference time of one or more clock domains over-the-air to the UEs, while gNBs and UPFs are typically synchronized via the underlying network. The time reference needed for industrial applications is provided by one or more grandmaster clocks that are located outside the 5G system. Time distribution is based on gPTP, which can be invisibly transmitted over the 5G network via the user plane. The 5G system acts as a time-aware system that corrects the gPTP time distribution at the egress port of the 5G user plane, according to the residence time that gPTP messages have spent within the 5G system. This 5G system residence time is determined by time-stamping the gPTP messages at the ingress port and egress port of the 5G system based on the 5G grandmaster clock.

In special cases, the 5G grandmaster clock can also act as a time reference for the industrial network, i.e. the TSN domain and the 5GS are synchronized based on a common clock provided by the 5GS.

A native 3GPP solution for 801.1AS support is therefore available and is the recommended solution for device synchronization.

5 Conclusion and outlook

This white paper describes the most important use cases and functional principles of industrial Ethernet networks as currently used in factory and process automation control systems. The second part of the paper outlines a reference model for integration of these systems with 5G networks. It explains how a 5G network can be logically integrated into existing wired networks and how various deployment options can physically support industrial use cases.

This section gives some key findings and recommendations. These are, on one hand, intended as guidance to system engineers and, on the other hand, expresses with regard to how 5G and other systems should ideally evolve with regard to the needs of industrial communications.

Support for the control application's quality of service requirements

The 5GS enables additional functions for factory use cases by replacing wired connections with wireless connections. Control applications and applications on the device side do not need to be aware of this change, and should continue to operate as before. The 5GS has to be configured with the information needed to establish the connections in conjunction with the required QoS. To allow automatic operation with TSN, 3GPP has defined translator functions to provide the required control information for establishing the 5G links. If similar automated operation is desired for legacy industrial Ethernet protocols, similar translator functions will have to be defined.

Radio coverage is a prerequisite

To ensure sufficient availability and reliability with a 5G network, a careful radio network planning activity is of paramount importance. Consideration must be given to specific building topologies, installed machines and other moving and stationary assets that might impact the radio waves. Radio coverage may need to be adjusted regularly.

Integrating on the IP layer (L3) is relatively simple

As the current 4G and initial releases of 5G provide IP connectivity, integrating a 5G network with an existing IT infrastructure is relatively straightforward. It requires simple IP address range coordination.

Integration on the Ethernet layer (L2) is more complicated

L2 tunneling integration and L3 integration are, in the short term, the recommended methods of integrating 5G with existing automation control systems as existing industrial Ethernet devices can be reused with no need for modifications. A further advantage is that 5G network elements do not need to be enhanced with specific functionality, such as interpretation of addressing schemes, forwarding of L2 frames sequences, analysis of frame type, etc.

L2 integration with the 3GPP PDU session of type Ethernet and the 5G LAN feature can be applied for some industrial Ethernet network types, but may not be universally applicable. Each traffic case must be analyzed and/or verified in a test network before implementation in a production environment.

In the longer term, L2 integration (applying the 3GPP-PDU session of type Ethernet and the 5G LAN feature) will become an integration option.

L2 native integration best with TSN

When 5G networks support native L2 connectivity (i.e. PDU session of type Ethernet) integration with industrial Ethernet systems can be achieved by means of TSN principles. The 5G system is then integrated into a TSN network as a TSN-compliant bridge that supports real-time streams, as required by industrial Ethernet networks; specific functions and procedures will then be transparent to the 5G network.

Broadcast domains in one UPF only

Deploying a single L2 broadcast domain for multiple UPFs causes additional, error-prone configuration effort, both within the 5G network and within the IT infrastructure, as explained in section 4.3.2.

D2D (5GNR sidelink)

Although appealing, the applicability of D2D principles to critical industrial communication needs to be assessed by means of additional research, similar to the activities performed previously for V2X communication. The research community is called upon to carry out such work.

6 Annex A: Integration of automation system topologies

This annex provide some greater detail on the various integration options mentioned in section 4.3.2, based on layer 3 and layer 2 communication and via layer 2 tunneling.

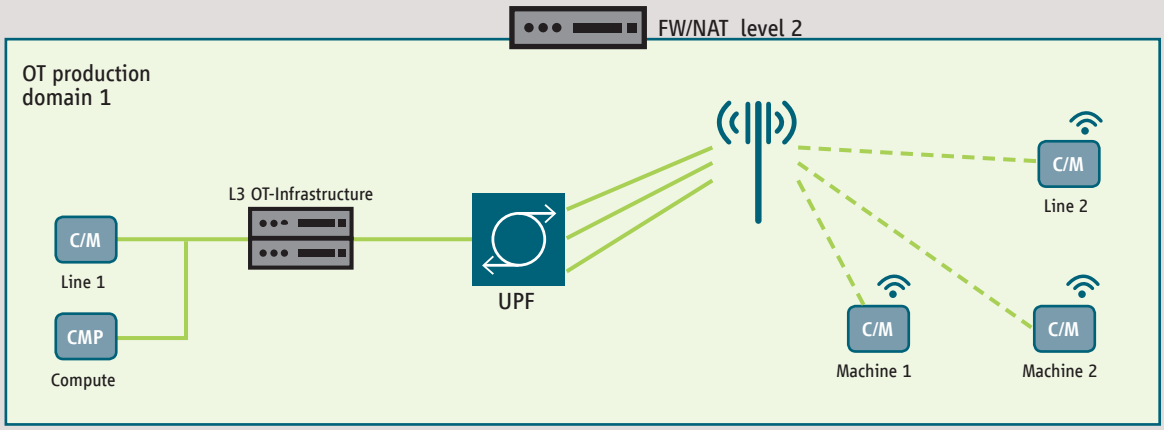
6.1 Integration with layer 3 automation control systems

Integration of automation systems on the IP level (L3) allows the 5G network to be used in the most flexible way: the UPF node represents an IP router that is added to and integrated with the existing L3 infrastructure. L3 integration enables use cases UC1, i.e. L2C, C2C and D2Comp, by means of the 3GPP concept "PDU session of type IP". A PDU session channels IP traffic from and to each UE into a dedicated 3GPP-native tunnel, while the UPF node connects like a router via IP to the existing network infrastructure. Each UE can have many concurrent PDU sessions with differing IP addresses and differing QoS attributes.

Deployments of this type are shown in Figure 12 and Figure 13 where dotted lines represent radio links.

Figure 12 depicts a scenario with a single OT production domain, i.e. all C/M nodes and or compute devices are connected via a single UPF node. However, it should be noted that the single antenna shown represents an entire radio network that may, in reality, comprise multiple antennas with differing characteristics, depending on the factory layout and topology.

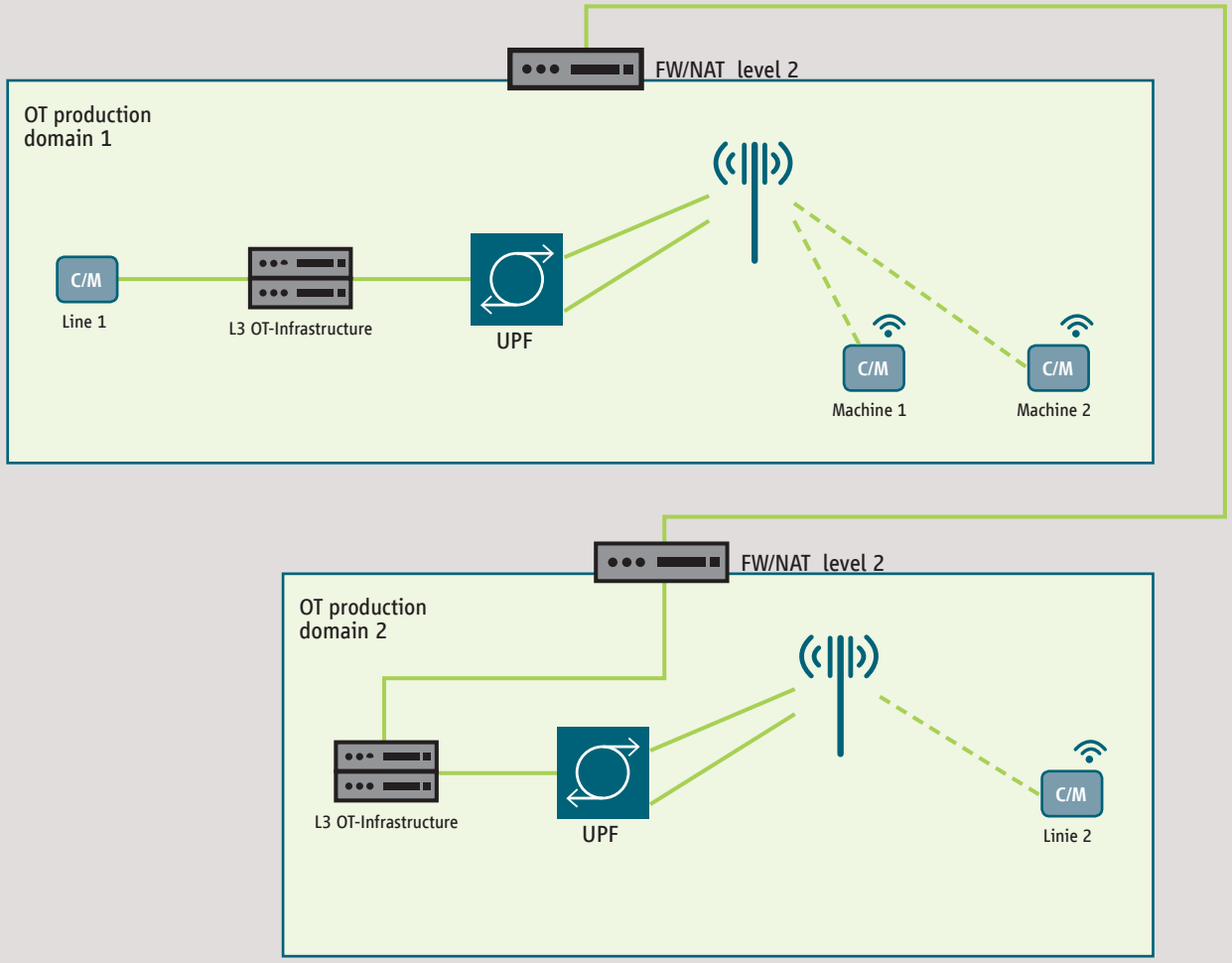
Fig. 12: Layer 3 connectivity



Source: 5G-ACIA / ZVEI

Figure 13 shows deployment across two OT production domains, where one line controller (Line 2 C/M) is connected via a radio link remotely to the first line controller (Line1 C/M).

Fig. 13: Layer 3 connectivity, multiple OT domains



Source: 5G-ACIA / ZVEI

Common to both scenarios is that all C/Ms and compute resources can communicate with each other on the IP level, with no constraints imposed by 5G. If logical segregation is needed (e.g. line 2 C/M should not communicate with machine 1 C/M), all IP network segregation and routing techniques can be employed, i.e. private IP addressing, subnetting, etc. In the two figures, dotted lines represent wireless connections and solid lines represent wired connections carrying L3 traffic (IP). All traffic from devices connected via the radio network is anchored by the UPF node – the UPF node routes the traffic into the existing L3 infrastructure. Private or public IP addressing schemes can be used in accordance with the OT domain addressing strategy.

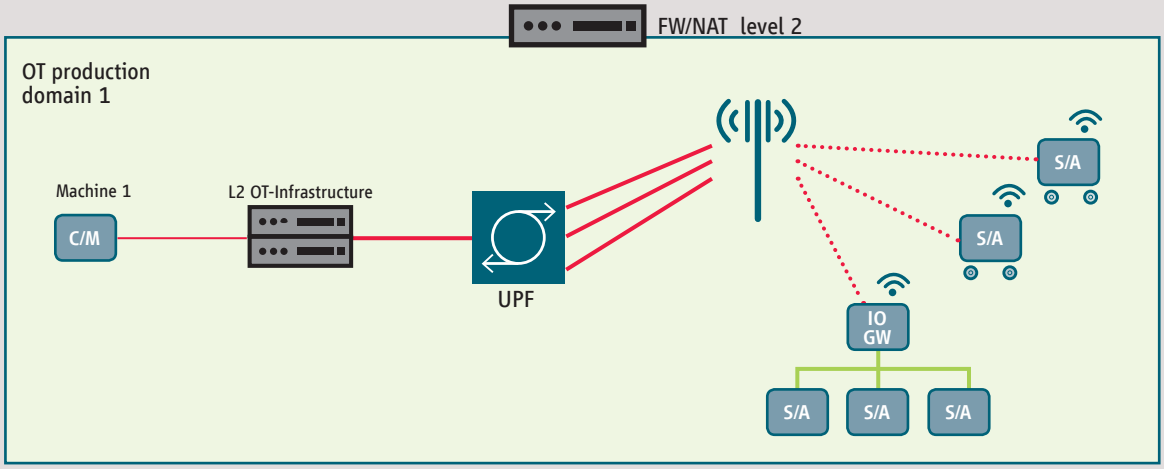
Depending on required latency, bandwidth and prioritization, each RAN-connected device can be assigned specific QoS settings, e.g. for particular latencies or priorities.

6.2 Integration with layer 2 automation control systems

In addition to L3 traffic, many use cases (e.g. UC2, i.e. C2D) require L2 transmission, as described in section 2.

Figure 14 shows deployment scenarios where controllers are connected via L2 links to an arbitrary number of S/As that either have their own mobile interface and are therefore connected directly to the 5G radio network, or are connected via an IO/GW node. When connected via an IO/GW node, only that IO/GW node will possess a 5G UE, i.e. the S/As attached to it will not be aware of the radio link at all. Similar to L3 traffic, L2 traffic is also transmitted via radio links (dotted red lines) and is anchored in the UPF node that channels L2 traffic to each UE into its own 3GPP native tunnel, i.e. PDU session of type Ethernet.

Fig. 14: Layer 2 connectivity, multiple L2 PDU sessions within a single broadcast domain



Source: 5G-ACIA / ZVEI

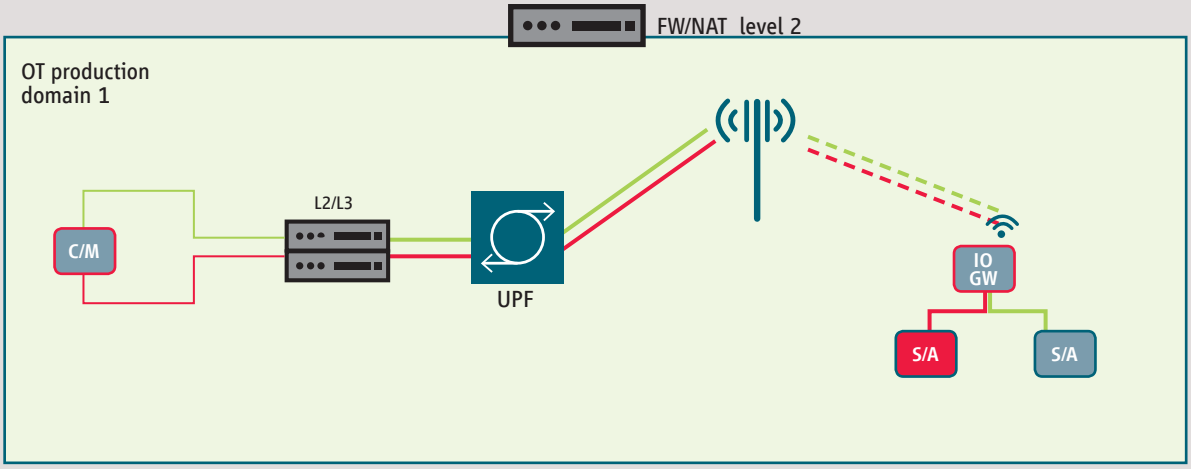
L2 traffic for multiple industrial Ethernet network systems can be multicast and broadcast between C/Ms and S/As, i.e. a controller can employ L2 multicast addressing to communicate with a large number of S/As that belong to e.g. a single automation control system (e.g. a machine). The UPF supporting native L2 access (3GPP PDU session of type Ethernet) must therefore ensure that S/As and C/Ms belonging to a single automation control system are within a unique broadcast domain. In real-world deployments, multiple automation control systems are connected via one UPF. In this case, the broadcast domains belonging to the various automation control systems must be segregated in the UPF. This segregation can be achieved e.g. by assigning a VLAN tag to each automation control system.

From a deployment perspective, an automation control system (i.e. a machine) can manage many field devices spread over a large geographical area and therefore may be connected via multiple UPFs in various OT production domains. In these cases, the L2 broadcast domain would have to be extended to include more UPFs so that all C/Ms and S/As belonging to a single automation control system can communicate with each other. Given the complexity of implementing and maintaining a network structure of this kind, it is advisable to avoid spreading field devices belonging to a single automation control system across more than one UPF.

It should be noted that some use cases (e.g. UC1 in section 2) may require distribution across more than one UPF, when e.g. machine controllers are communicating with a remote line controller, possibly located in a different OT production domain. In such cases, it is advisable to employ routed IP (L3) communication between these two controllers.

Since many industrial applications comprise both real-time, mission-critical traffic (with short cycles) and non-real-time traffic, some S/As and C/Ms may require L2 or L3 connectivity. In a configuration of this kind, the IO/GW connection to S/As may require multiple concurrent connections (i.e. PDU sessions), each with its own specific characteristics. Characteristics can be e.g. L2 or L3 type PDU sessions, but also QoS levels according to 3GPP TS 23.501 [2] (5QI values).

Fig. 15: Concurrent PDU sessions



Source: 5G-ACIA / ZVEI

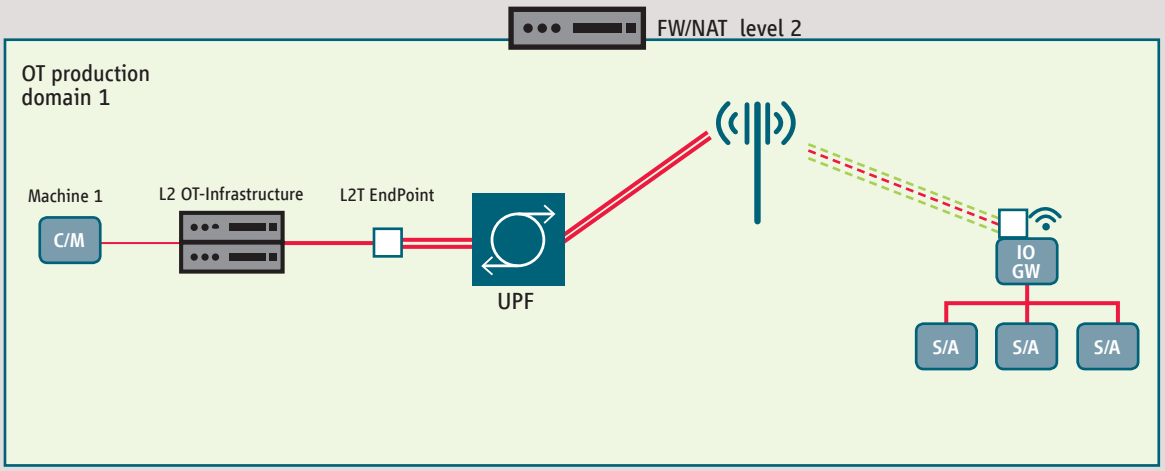
Figure 15 depicts a configuration where a UE acts as a wireless IO/GW connecting two S/As that do not have their own mobile interface. The red S/A requires L2 connectivity, while the blue S/A requires L3 connectivity – and they require differing latency and bandwidth. The IO/GW initiates two PDU sessions with differing characteristics.

Optionally, the IO/GW can only initiate a single PDU session of type Ethernet. It is then dependent on the application, whether or not a S/A uses IP communication via that single PDU session; this IP communication will be transparent to the 5G network.

6.3 Integration via layer 2 tunneling

A further way to integrate automation control systems with 5G networks before the complete 3GPP R15 feature set, including PDU session of type Ethernet, is commercially available, is to use L2 tunneling (L2T). This allows transmission of L2 frames via an L3 network. The L2 frames are encapsulated by the L2T end-point node to form an L3 connection that is established as per standard 3GPP procedure (see section 6.1) between the UPF and one or more S/As or IO/GWs. The peer L2T end-point positioned at the IO/GW then extracts the L2 frames from the L2 tunnel. It should be noted that L2T conveys the L2 frames in unmodified form, and completely transparently through the 5G network. L2 addressing and forwarding principles are not applied by the 5G network. L2T is depicted in Figure 16, with two L2T end-points; the red double-line represents the actual layer 2 tunnel.

Fig. 16: Integration with L2 tunneling



Source: 5G-ACIA / ZVEI

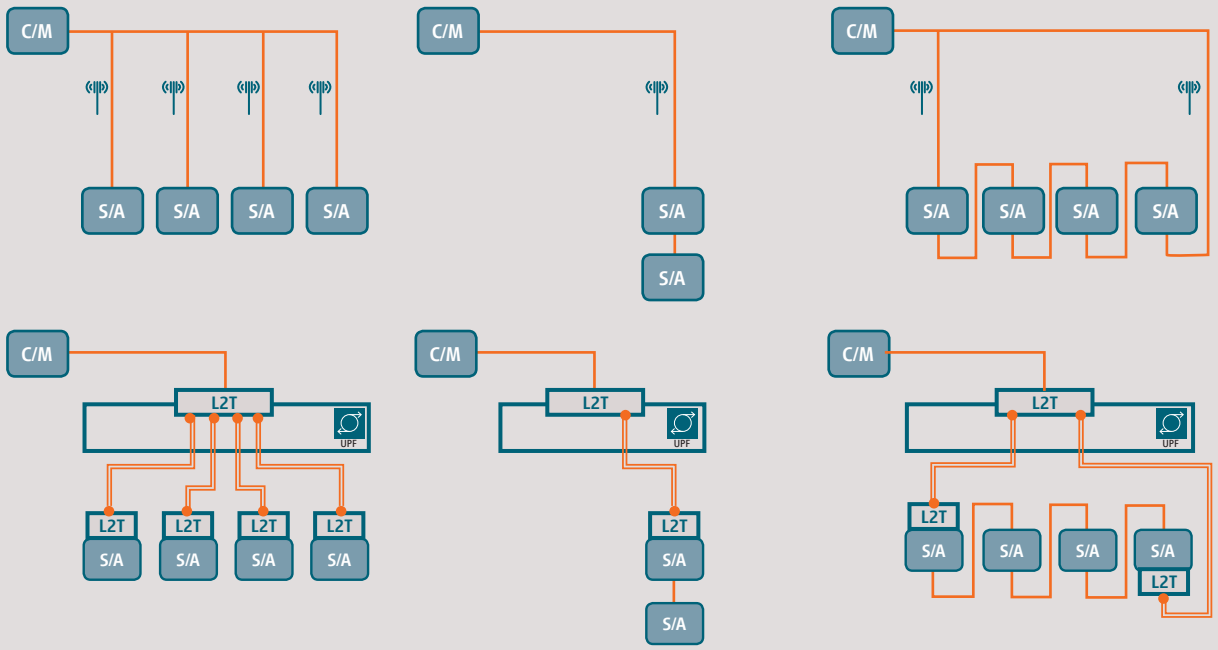
L2T enables a number of real-life automation control system topologies to be implemented across a 5G network. Figure 17 below shows three commonly used topologies. Again, red lines represent L2 connections and double red lines represent L2 tunnels that are conveyed via PDU sessions of type IP.

Figure 17 a) depicts a star topology where all four wired connections between the C/M and the S/As are replaced with 5G radio links. All four S/As need to be configured with a L2T end-point that communicates with the L2T end-point located in front of the UPF. L2 frames are conveyed invisibly between the C/M and the S/As. It should be noted that, in similar fashion to plain L3 integration, QoS attributes can be assigned with complete flexibility.

Figure 17 b) depicts a daisy chain topology where only the wired connection to the first S/A is replaced with a 5G radio link. This first S/A acts as an L2 bridge to its second (and possibly additional) S/A in the daisy chain. Again, the L2 frames are conveyed invisibly between the C/M and the first S/A.

Figure 17 c) depicts a ring topology where the wired connections between the C/M and the first and last S/A are replaced with 5G radio links, while the connections between the S/As remain wired.

Fig. 17: Control system topologies mapped to L2T



Source: 5G-ACIA / ZVEI

All three examples share a fundamental principle, i.e. the machine controllers (C/M) are remote from the machine which is (or parts of it are) mobile (non-stationary), while the field devices (S/As) are attached physically to the machine.

7 References, abbreviations and terms

7.1 References

[1] 3GPP TS 22.104; Service requirements for cyber-physical control applications in vertical domains

[2] 3GPP TS 23.501; System Architecture for the 5G System; Stage 2

[3] 5G-ACIA White Paper, 5G Non-Public Networks for Industrial Scenarios (White Paper), www.5g-acia.org

[4] 5G-ACIA White Paper, 5G for Automation in Industry (White Paper), www.5g-acia.org

[5] "The future of industrial communication: automation networks in the era of the Internet of things and Industry 4.0.," Wollschlaeger, Martin, Thilo Sauter, and Juergen Jasperneite; IEEE Industrial Electronics Magazine 11.1 (2017): 17-27. https://www.hs-owl.de/init/uploads/tx_initdb/IEEEMagazine.pdf

[6] IEC 61158 series; Industrial Communication Networks - Fieldbus specifications

[7] IEC 61784-2: Industrial Communication Networks - Profiles - Part 2: Additional fieldbus profiles for real-time networks based on ISO/IEC/IEEE 8802-3

[8] IEC/IEEE 60802; TSN Profile for Industrial Automation https://1.ieee802.org/tsn/iec-ieee-60802/

[9] TSN; IEEE 802.1 set of standards using synchronized services and time aware shaping, cyclic scheduling, pre-emption, ingress filtering or redundancy

7.2 Abbreviations

3GPP	3rd Generation Partnership Project
5G	5th generation cellular network
5GC	5G core network
5GS	5G system
5QI	5G QoS indicator
APN	Access point name
C2C	Controller to controller
C2D	Controller to device
C/M	Controller / master
CMP	Compute
D2CMP	Device to compute
D2D	Device to device
DCS	Distributed control system
DHCP	Dynamic Host Configuration Protocol
DN	Data network
DNN	Data network name
ERP	Enterprise resource planning
FW	Firewall
gNB	g-node B (5G NR base station)
GW	Gateway
IP	Internet protocol
L2	Layer 2 communication based on IEEE 802.3
L2C	Line controller-to-controller
L2T	Layer 2 tunneling
L3	Layer 3 communication, routed IP-based communication
LAN	Local area network
MES	Manufacturing execution system

NAT	Network address translation
NG RAN	Next generation radio Access Network
NPN	Non-public network
NR	New radio
OT	Operational technology
PDU	Protocol data unit
PLC	Programmable logic controller
QoS	Quality of service
RAN	Radio access network
S/A	Sensor / actuator
SoE	Sequence of events
TR	Technical report
TS	Technical specification
TSN	Time-sensitive networking
UE	User equipment
UPF	User plane function
URLLC	Ultra reliable and low latency communication
VLAN	Virtual LAN

7.3 Terms

Industrial communication network, also called an industrial Ethernet network
Fieldbus profiles for real-time networks based on ISO/IEC/IEEE 8802-3, see [7]. Examples are CC-Link, EtherCAT, Ethernet/IP, POWERLINK, PROFINET, Sercos III.

Communication cycle time
The communication cycle is the period between two consecutive transmissions or receptions of data.

8 5G-ACIA members As of November 2019

5G Alliance for Connected Industries and
Automation (5G-ACIA),
a Working Party of ZVEI
Lyoner Strasse 9
60528 Frankfurt am Main, Germany
Phone: +49 69 6302-424
Fax: +49 69 6302-319
Email: info@5g-acia.org
www.5g-acia.org