Deliverable D5.2

FIRST REPORT ON 5G NETWORK ARCHITECTURE OPTIONS AND ASSESSMENTS

The 5G-SMART project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 857008.

D5.2 First report on 5G network architecture options and assessments

Grant agreement number:	857008
Project title:	5G Smart Manufacturing
Project acronym:	5G-SMART
Project website:	www.5gsmart.eu
Programme:	H2020-ICT-2018-3
Deliverable type:	Public
Deliverable reference number:	D18
Contributing work packages:	WP5
Dissemination level:	Public
Due date:	2020-11-30
Actual submission date:	2020-11-30
Responsible organization:	Orange
Editor(s):	G. Madhusudan (Orange), Dhruvin Patel (Ericsson)
Version number:	1.0
Status:	Final

Short abstract:	This document elaborates on architecture models from deployment and operational viewpoint by considering the use cases identified within the 5G-SMART project. Also, it provides a deployment validation analysis of such deployment models considering a set of key technical enablers.

Keywords:	NPN, TSN, Edge computing, Industrial LAN

Contributor(s):	Krister Landeras (ABB) Markosz Maliosz (BME) Ahmad Rostami (Bosch) Jose Costa Requena (Cumucore) Finn Pedersen (Ericsson) György Miklós (Ericsson) Joachim Sachs (Ericsson) Marilet De Andrade Jardim (Ericsson) Stefano Ruffini (Ericsson) Kun Wang (Ericsson) Gabor Nemeth (Ericsson) Hubert Przybysz (Ericsson) Niels König (IPT) Pierre Kehl (IPT) Gabor Soos (T-System) Daniel Venmani (Orange) Olivier Le Moult (Orange)

Disclaimer

This work has been performed in the framework of the H2020 project 5G-SMART co-funded by the EU. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein.

This deliverable has been submitted to the EU commission, but it has not been reviewed and it has not been accepted by the EU commission yet.

Executive summary

5G standardization has introduced new architectural options such as standalone non-public network (SNPN) and public network integrated non-public networks (PNI-NPN) to enable smart manufacturing use cases. This report elaborates on some architecture models from deployment and operation point of view by considering the use cases identified within the 5G-SMART project. Furthermore, key technical enablers related to integration of 5G with Time-Sensitive Networking (TSN), support for LAN type services, network slicing and edge computing are analysed. Different deployment options are considered along with required technical enablers and a deployment validation analysis is provided. Further, this report lists the potential improvements or gaps observed with respect to the use case requirements.

With regards to interworking of the 5G system with Ethernet networks, a 5G network configuration solution is described. Also, shortcomings of existing Release 16 based 5G Virtual Network grouping mechanism are highlighted in the context of integrating with the existing industrial wired network. Different connectivity segments of industrial automation networks are described and the role of an integrated 5G-TSN network in the ecosystem is explored. Different deployment options for TSN integration with respect to NPN options are investigated. TSN integration is feasible for most of the 5G-SMART use cases when SNPN deployment options are used. For shared control plane (PNI-NPN) more investigation is needed on how control plane components are realized and integrated with a TSN system, to ensure secure and reliable functioning of the entire communication system.

An analysis is performed which shows some areas in which 3GPP release 16 specification to support Ethernet based industrial networks could be improved. Furthermore, solutions are proposed to address such shortcomings which are under discussion for release 17 specification. Primarily two areas are detailed: simplified network configuration and dynamic traffic-driven establishment of forwarding rules. Security zone allows segmentation of the traffic with different security requirements. The report explains how security zones in 5G can be enabled with VLAN configuration mechanism and network slicing functionality.

Furthermore, 5G-SMART use cases are examined considering native 5G connectivity for different deployment options. The examination is based on the reliability and latency requirement of the use cases. For use cases that do not necessitate very low latency, PNI-NPN options can be considered as potential deployment choices.

The report elaborates on different edge cloud service models suitable for the manufacturing use cases including container and functional level service-based abstractions. Recommendations on edge cloud infrastructure and resource management are made in order to meet low latency targets.

Overall, this deliverable provides a systematic study of network architectures from different perspectives. It includes an in-depth analysis on how technical enablers along with different architecture options can be deployed to ensure that the functional requirements of smart manufacturing use cases are met.

Contents

1 Introduction

Industry 4.0 use cases are usually demanding in terms of end to end (E2E), quality of service (QoS) – low latency combined with high reliability and availability. This leads to the consideration of various architectural options for industrial communication technology that are different from conventional mobile broadband services. Another major factor for consideration of various architectural options is the large investment in Operational Technology (OT) equipment done by end customers. These legacy systems are based on different industrial technologies and local network solutions. Often, new network architecture concepts need to co-exist and integrate with the existing solutions.

In addition to technical enablers in 5G such as low latency communication and network slicing, there are different architectural options that will accelerate the adoption of 5G into smart manufacturing ecosystem. These include 5G support for Ethernet industrial networks and edge computing that enables data processing close to the industrial equipment.

5G networks to be deployed as part of a private industrial communications infrastructure are defined as Non-Public Networks (NPN). There are different deployments of NPNs depending whether they are completely isolated as part of the industrial infrastructure i.e. Standalone NPN (SNPN), or the NPNs integrated with existing infrastructure of Mobile Network Operator (MNO) i.e. Public Network Integrated NPN (PNI-NPN). This document will analyse these deployment options and technical enablers required for smart manufacturing in line with 3GPP 5G specifications Release 16. A further gap analysis is performed considering Release 16 as baseline technology to support smart manufacturing use cases. This report is referred to as D5.2 in the rest of the document and the follow-up of this report is referred to as D5.4. D5.4 will be published by the end of the project.

1.1 Scope

The goal of this document is to study the impact of different NPN options on architecture and deployment choices in the context of 5G-SMART use cases. In particular the document highlights open issues and investigates different NPN options for:

- Operational models in the context of different NPNs,
- 5G interworking with Ethernet industrial networks,
- Integration with a local TSN domain including local time synchronization,
- Native 5G connectivity use cases,
- Support for different on-premise security zones.

The focus will be on 5G Stand Alone (SA) architecture. 5G Non-Stand Alone (NSA) may be considered in some contexts. 4G networks and other internet of things (IoT) variants such as LTE-M [1]and narrowband internet of things (NB-IoT) are out of scope for this document.

[1] https://www.gsma.com/iot/wp-content/uploads/2019/08/201906-GSMA-LTE-M-Deployment-Guide-v3.pdf

1.2 Relation to other work packages in 5G-SMART

In 5G-SMART, Work Package (WP5) analyses and designs 5G technical features for smart manufacturing. It takes a further leap from the trial work packages (WP2, 3 & 4) by deep diving in technical aspect beyond those that are already standardized in 3GPP. It also explores enhancements and their integration within the manufacturing ecosystem.

Figure 1 shows the overall workflow of WP5. WP5 takes input from the use case requirements defined in deliverable D1.1, D2.1 and D3.2 [5GS20-D11] [5GS20-D21] [5GS20-D32] and provides output to the dissemination activities in WP6. In WP5, new and future-looking 5G technical features are investigated and evaluated against the wide range of use case requirements. Figure 1 shows the three main pillars of work undertaken in WP5. The gap analysis provided in this deliverable serves as input to standard developing organizations (SDOs) and industry fora.

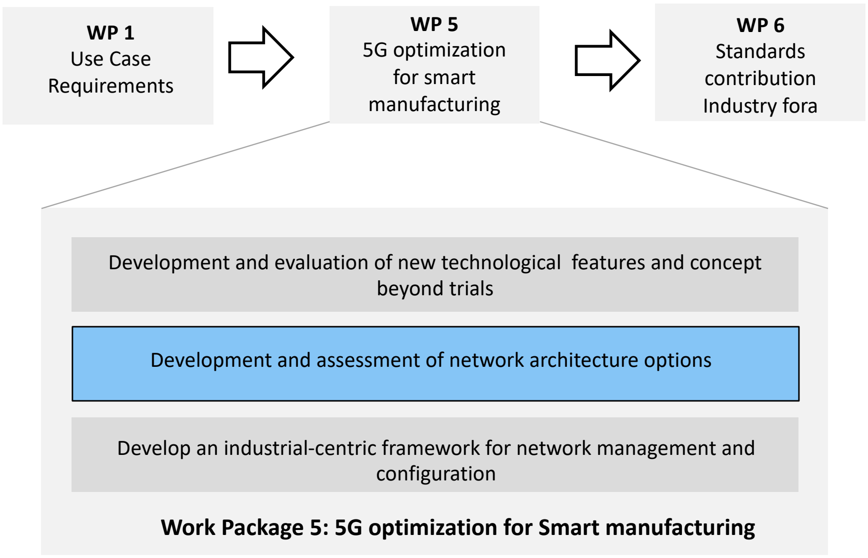

WP 1
Use Case Requirements
WP 5
5G optimization for smart manufacturing
WP 6
Standards contribution Industry fora
Development and evaluation of new technological features and concept beyond trials
Development and assessment of network architecture options
Develop an industrial-centric framework for network management and configuration
Work Package 5: 5G optimization for Smart manufacturing

Figure 1 Workflow of WP5 and its relation to other WPs

1.3 Overview of the structure of the document

The network architecture activity in work package 5 (WP5) takes input from [5GS20-D11], where 5G-SMART use cases are defined. Figure 2 shows the overall structure of the document showing methodology chosen to achieve expected results. The deliverable starts with listing down functional requirements of the 5G-SMART use cases. Further, based on the requirements, 5G network architecture models are described considering deployment and operational viewpoint. Several non-public network (NPN) deployment options are listed down, and their characteristics are analysed. Stakeholders and roles are defined to understand how such deployment models will operate in the industrial eco-system. In the next section, 3GPP standardized technical enablers and system enablers such as edge computing are elaborated. Next section provides the main output of the deliverable on deployment validation based on the 5G-SMART use cases, technical enablers and different NPN options. In addition to this, a gap analysis is performed considering functional requirements of the use cases. Section 7 concludes the document.

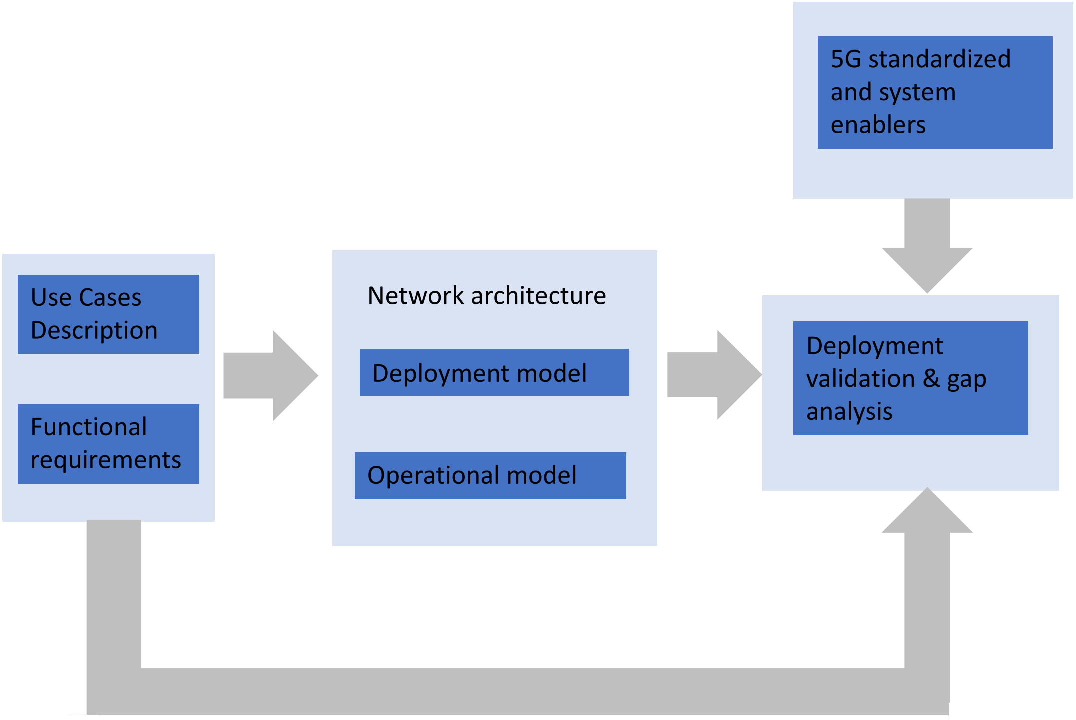


Figure 2 Structure of the document

2 Summary of requirements from 5G-SMART use cases

The use cases described in the 5G-SMART project explore the benefits of 5G in smart manufacturing. Within 5G-SMART, a wide variety of smart manufacturing use cases have been investigated in the deliverables [5GS20-D11] [5GS20-D21] [5GS20-D32]. The analysis of the requirements and of the use cases in these deliverables clearly show the need for a reliable, low-latency, high-performance wireless infrastructure in factories of the future [5GS20-D11].

Furthermore, the analysis highlights the complexity of introducing wireless technology in manufacturing. In addition to meeting challenging performance Key Performance Indicators (KPI)s, the wireless network infrastructure must meet functional requirements related to QoS, security and time synchronization as well as operational requirements on e.g. integration with existing factory infrastructure.

In this section, an overview of the functional requirements (as shown in Table 1 below) and their impact on network architecture for selected use cases are described. For a more comprehensive description of the use-cases, please refer to D1.1 [5GS20-D11]. A single 5G network architectural model might not be able to satisfy all the requirements shown in the table below. Hence, there is a need to investigate different architecture options where different stakeholders are present. This is also motivated by technical and business considerations.

Functional requirements	**Characteristics/Details**
Network status, capability exposure, and interface towards Operation & Management (O&M)	For an efficient integration of 5G system with the infrastructure of the factories of the future as well as with applications, some use cases require an exposure of the 5G communication status and configuration concerning device connectivity management (e.g. for configuring QoS or network slices).
Security	Guaranteeing confidentiality and integrity of information, while taking into account the common security measures like security zones in the factories, are among major requirements of most use cases. A 5G deployment architecture needs to ensure that such requirements are fulfilled when integrated with existing industrial network architecture [SEC20-5GACIA].
End-to-End QoS	Several targets use cases have stringent QoS requirements on the underlying communication system. Therefore, it is important that deployed 5G architecture models support these requirements in an end-to-end fashion, covering RAN, core network, and edge computing.
Isolation between different network functions (Network slicing)	5G network slicing is one the possible solution to realize several use cases having various QoS requirements on top of the same network infrastructure in a flexible, resource-efficient, and secure manner.
Time synchronization	Accurate time synchronization is necessary for some industry applications. In addition, it is an important factor in assuring that network communication delay variations (jitter) are within the requirements specified by applications. The network architecture needs to ensure that end-to-end time synchronization might be supported.

Layer 2 switching/TSN integration	Transporting TSN and industry LAN (I-LAN) traffic over 5G requires appropriate data-plane features (encapsulation/decapsulation) of TSN/I-LAN frames at the communication end points. The 5G network architecture needs to take care of the control-plane features (configuration and integration) with industrial network management system (e.g.TSN's Central Network Configuration CNC)
Seamless mobility	In order to avoid service interruption for the mobile devices on the manufacturing shopfloor. The network architectures should have a functionality to decrease the service interruption time.
Resource & Energy efficiency	5G communication system should provide resource handling (including energy) and efficient communication between application's endpoints, e.g. for applications with low traffic volumes or those in mobile devices.

Table 1 Functional requirements of the 5G-SMART use cases

3 5G architecture models

5G non-public network (NPN) provides communication services to specific organizations, which makes it more suitable for smart manufacturing compared to mobile network providing communication to general public. According to the 3GPP definition in [3GPP20-TS23501], "Non-public networks are intended for the sole use of a private entity such as an enterprise, and may be deployed in a variety of configurations, utilizing both virtual and physical elements". From an architecture point of view, there are different possibilities through which 5G NPN can be realised. Hence, it is of paramount importance to investigate the suitability of such options for the wide range of smart manufacturing use cases.

The current section provides information on the different 5G architectural models based on the deployment and operational viewpoints. Different deployment options are listed with its characteristic and operational models are defined based on the stakeholders and roles considered in the eco-system.

5G system architecture is based on the service-based architecture principles, it is specified in 3GPP technical specification [3GPP20-TS23501]. 5G system architecture includes network functions performing various tasks to ensure data connectivity of UE with external data network. Figure 3 shows a simplified 5G system architecture with relevant interfaces. The data network is connected to 5G system via N6 interface. For smart manufacturing case, the data network would be either TSN or Ethernet Industrial network. Packet Data Unit Session (PDU) is established between User plane function and UE. User plane function (UPF) interact with core control plane network functions (Session Management Function SMF) over N4 interface to configure PDU sessions between UE and UPF.

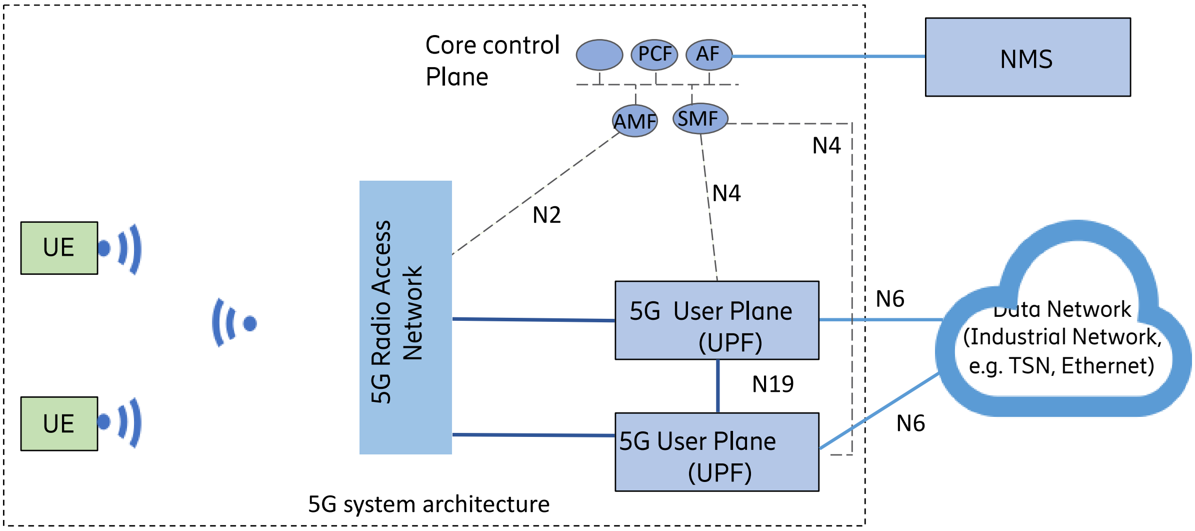


Figure 3 5G system architecture [3GPP20-TS23501]

3GPP in Release 16 distinguishes between two type of NPN deployments:

1. Stand-alone Non-Public Network (SNPN),
2. Public network integrated NPN (PNI-NPN).

The former does not rely on network functions provided by a Public Land Mobile Network (PLMN), while the latter is deployed with the support of a PLMN. According to 3GPP, PLMN is a network established and operated by an administration for the specific purpose of providing land mobile communication services to the public. In this report PLMN is referred to as Public Network (PN) in contrast to an NPN. 3GPP in Rel-16 introduced new functionality to support NPNs, in order to fulfil requirements listed in [3GPP20-TS23501]. Architectural aspects of the different NPNs deployments are documented in [3GPP20-TS23501].

An SNPN is identified by the combination of a PLMN ID and a Network identifier (NID). The Next Generation- Radio Access Network (NG-RAN) node providing access to SNPNs broadcasts the following information:

1. one or more PLMN ids,
2. list of NIDs (per PLMN id) identifying the SNPNs. NID can be configured at the base station (gNB) for specific broadcast. NG-RAN node support broadcast of total 12 NID.

Another optional information is a human-readable network name, and information to prevent UEs not supporting SNPNs accessing the radio cell. UE refers to 5G functionality in a (mobile) end device that provides connectivity to the network [5GS20-CT]. It is possible for a UE that has been registered to an SNPN to perform another registration with a public network (PLMN), via the SNPN user plane, to be able to access PLMN services via the SNPN.

PNI-NPNs can be enabled by means of dedicated Data Network Names (DNNs) or by network slicing. Closed Access Groups (CAGs) may be used to apply access control. A CAG identifies a group of subscribers that are permitted to access CAG cells. CAG is an access control mechanism used to prevent improper connection of unauthorized UEs. The CAG associates a selected number of cells with the UEs that are allowed to access the NPN through those cells.

3.1 Deployment models

Table 2 below summarizes the deployment models possible for the NPN. Four possible variants of the deployment models are defined. These variants are also reflected in other reports [NPN19-5GACIA] [E2E19-5GNGMN] [5GS20-D41]. Radio access network (RAN) sharing is in the scope of the study. For the case of the SNPN, a combination of the PLMN ID and Network identifier (NID) identifies an SNPN which can be assumed to NPN identifier (NPN ID). NPN ID is identifier assigned to NPN [NPN19-5GACIA], it can be combination of the PLMN ID and network identifier (NID) according to [3GPP20-TS23501].

Deployment Option no.	NPN deployment options	Characteristic/details	3GPP NPN type
NPN 1	Standalone NPN	All NPN functions are on-premises. NPN is a fully separate physical network from the Public Network (PN) with dedicated NPN ID. However,	

		dual subscription with NPN and PLMN is possible. Access to PLMN services can be realized via an optional firewall connection and roaming agreement.	SNPN
NPN 2	Shared RAN	NPN is based on 3GPP technology with its own NPN ID. Only the RAN is shared with the PLMN, all other network functions remain segregated, also data flows remain local. It can be realized by: • Multi-Operator Core Network (MOCN), where two or more core network entities are sharing eNodeB/ gNodeB and spectrum • Multi-Operator RAN (MORAN), where two or more core network entities are sharing gNodeB with non-shared spectrum	
NPN 3	Shared RAN and control plane	NPN is based on 3GPP technology and RAN shared with the PLMN. The network control plane is hosted by the PLMN. Data flows remain local.	PNI-NPN
NPN 4	NPN hosted by the PN	NPN traffic is off premise but treated differently through Network Slice instances and dedicated DNNs.	

Table 2 NPN deployment model based on report [E2E19-5GNGMN]

3.1.1 NPN 1 (Standalone NPN)

The first deployment model is the standalone NPN, where all the network functions required to operate the network are physically located on premise as shown in Figure 4. These includes dedicated RAN, 5G core network components and on-premise cloud capability functions for local user data communication. This enables low latency. Coordination techniques can be applied between the PN (off premise) and SNPN (on premise) for example to improve coexistence and interference issues on the RAN. Seamless device mobility requires additional configuration to enable e.g. devices connected to multiple networks via e.g. dual subscriptions. Local interaction between the 5G System (5GS) control plane and the OT management system is possible. Local survivability defines the capability of the NPN to maintain its operation when the connection between the industrial site and an external network is lost. Local survivability for NPN 1 is provided by a local control plane on premise. Concerning data privacy, user data as well as subscription and connectivity related information is kept locally on premise. Figure 4 illustrate the case where UE registered to SNPN can also registered PLMN via SNPN user plane going through demilitarized zone (DMZ). DMZ is network segment that logically separate the internal and external industrial networks for controlling data flow between the networks[2].

[2] https://www.tuvit.de/fileadmin/Content/TUV_IT/pdf/Downloads/WhitePaper/whitepaper-iec-62443.pdf

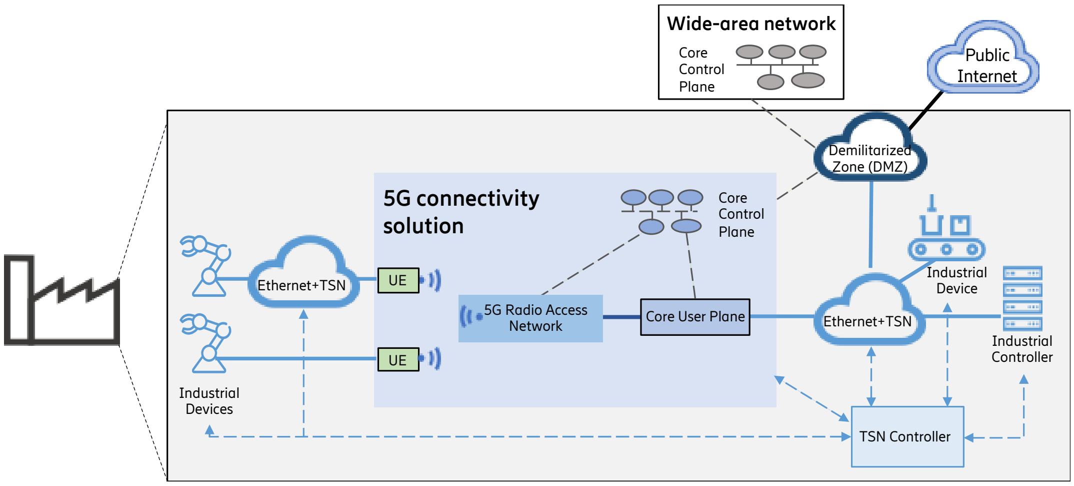


Figure 4 NPN 1 (SNPN)

3.1.2 NPN 2 Shared RAN (SNPN)

This deployment model considers a shared radio access network between public network and NPN, shown in Figure 5. The core network and also cloud capabilities are provided on-premise and all NPN traffic remains on-premise. The RAN is shared with the PN and enabled by MOCN or MORAN. Proper local radio planning and dimensioning is required to communication services requirements. For shared RAN interference caused by PN users entering the on-site NPN deployment is minimized as shown in reference [5GS20-D14]. Seamless device mobility requires additional configuration to enable e.g. devices connected to multiple networks via e.g. dual subscriptions. Local interaction between the 5GS control plane and the OT management system is possible too. Local survivability is provided by a local control plane on premise. Concerning data privacy, user data as well as subscription and connectivity related information is kept locally on premise, as the connectivity for NPN is used is established to the local core network nodes.

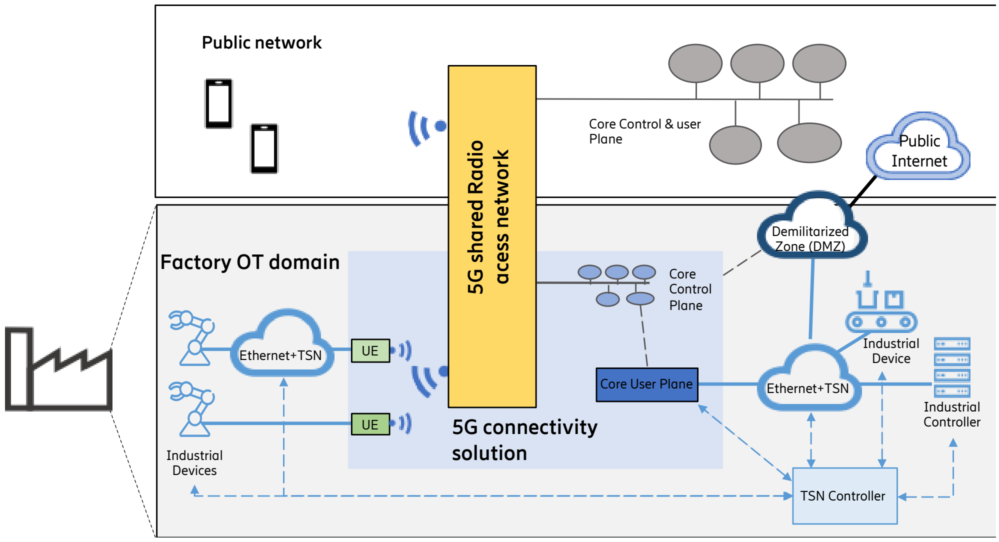


Figure 5 NPN 2 (SNPN)

3.1.3 NPN 3 Shared RAN and core control plane (PNI-NPN)

Compared to NPN2, in this deployment model the 5G control plane is shared by the PN and the NPN (see Figure 6). The cloud capability on premise and the local user data breakout enable deterministic low latency of the User Plane. Segregation can be realized via network slicing. Seamless device mobility between PN and NPN is implicitly enabled, as the devices, i.e. UEs, are also PLMN subscribers. Interaction with the shared control plane hosted in the PLMN and the private OT network management system will be required for wide range of manufacturing use cases. Local survivability may not be achieved if the connection to the PN Core Network is lost. Concerning data privacy, user data do not leave the factory premises however the users' subscription data, as well as connectivity related session and mobility information, are present in the operator's premises.

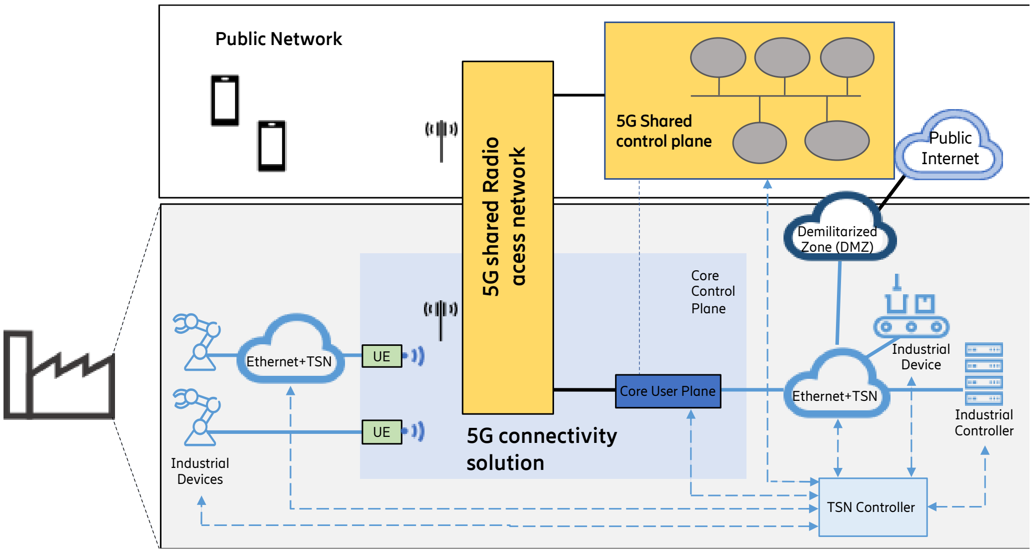


Figure 6 NPN 3 (PNI-NPN)

3.1.4 NPN 4 NPN hosted by the public network (PNI-NPN)

For NPN hosted in the public network, all network functions are hosted by the public network. All NPN users are PLMN subscribers and the separation between public network and NPN can be done via implementation of the network slicing.

3.2 Hybrid NPN network

Hybrid network refers to a combination of the above-described NPN deployment options. For example, NPN 1 can be utilised as primary mode for communication and an NPN hosted by PN acts as standby or fall-back option. Analysis on how it affects the availability and efficiency (in terms of cost) still needs to be investigated.

This aspect will be investigated in D 5.4.

3.3 Operation models

In the previous section, deployment models are presented without discussing roles and responsibilities of various stakeholders in the operation of the 5G network. Operational models are a way to take into account the roles of different stakeholder involved in operating an NPN. So far, only deployment models are described in 3GPP and industrial for a such as NGMN, 5G-ACIA, this report leaps furthers into defining operational models taking into account certain assumptions.

This section considers the operational aspects of 5G networks and analyses various relevant scenarios. An operation model defines the roles and responsibilities of various stakeholders in setting up and operating an NPN. For this purpose, we consider the following stakeholders and roles:

Stakeholders:

1 MNO is the stakeholder which owns and manages a PLMN.
2 Industrial Party is the stakeholder which requests NPN services for performing a industrial task.
3 3rd Party any stakeholder which cannot be categorized as MNO or industrial party, e.g., a network vendor or other third-party supplier, which can provide the NPN user (defined below in the roles) with services such as network deployment, integration and management.

Roles:

1 NPN Owner is the role of owning the NPN infrastructure, which includes both hardware and software components.
2 Spectrum Owner is the role of having the right to transmit radio signal at a certain frequency band.
3 NPN Integrator is the role of deploying and configuring the NPN according to a chosen architecture in order to make it ready for use.
4 NPN Operator is the role of operating and managing the NPN on a day-to-day basis.
5 NPN User is the role of using the services offered by the NPN.

3.3.1 Assigning Roles to Stakeholder

An operation model specifies the assignment of all the roles to the stakeholders. Table 3 lists all possible roles a stakeholder can take. In principle, except for the role of NPN user, which is exclusively assigned to the Industrial party, all other roles can be taken by any of the three stakeholders.

We should note here that, for the sake of simplicity, the term *ownership* is used in a more generic sense, and we do not distinguish between direct or indirect ownership. For instance, in case of spectrum ownership we do not differentiate between the ownership through direct licensing or through leasing. The right to transmit on particular spectrum depends upon the regulation from the local authorities. There are multiple ways in which spectrum can be accessed.

	NPN owner	**Spectrum owner**	**NPN integrator**	**NPN operator**	**NPN user**
MNO	Yes	Yes	Yes	Yes	No
Industrial party	Yes	Yes	Yes	Yes	Yes
3rd Party	Yes	Yes	Yes	Yes	No

Table 3 Possible Allocations of Roles to Stakeholders

According to Table 3 and if we put the role of NPN user aside, in theory we can identify at least 3 (NPN Owner) x 3 (Spectrum Owner) x 3 (NPN Integrator) x 3 (NPN Operator) possible combinations. That is to say, 81 distinct operation models. Nevertheless, not all the combinations, which can be derived from Table 3, would be meaningful and likely in practice. In fact, different factors such as business interests of different stakeholders, or different regulations in different geographical regions and countries make certain operation models more attractive (and accordingly more likely to be materialized) than others. The following two examples provide more details on such cases:

- If the MNO is the Spectrum Owner, it is likely that the MNO also takes the role of NPN operator.
- A stakeholder who is NPN owner is likely to take at least one more role (e.g., Spectrum Owner, NPN Integrator or NPN Operator).

Accordingly, the total number of practically meaningful operation models might be much lower than 81. Below we present a few examples of possible operation models.

3.3.2 Operation Model 1

In this model the industrial party is both provider and consumer of the NPN services. That is, all the roles are directly taken by the industrial party (as depicted in Figure 7).

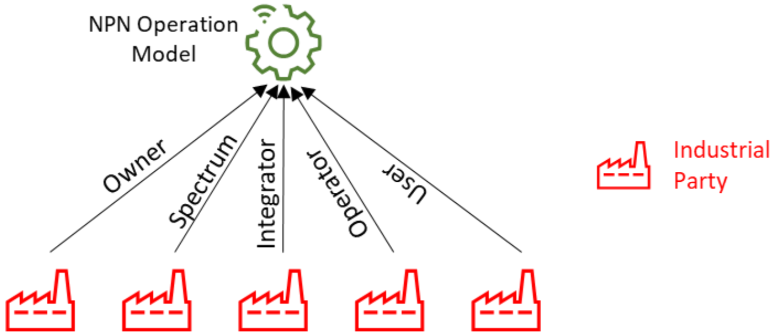


Figure 7 Operation Model 1

3.3.3 Operation Model 2

In this model the roles are distributed among all possible stakeholders, as depicted in Figure 8. In particular, the industrial party—as the user of the NPN—owns the NPN infrastructure, which is being deployed by a 3rd party integrator. Once the NPN is deployed and integrated, it will be operated by an MNO, which also takes care of providing required spectrum licenses. That is, in this case the industrial party has to deal with more than one external stakeholder for providing the NPN services.

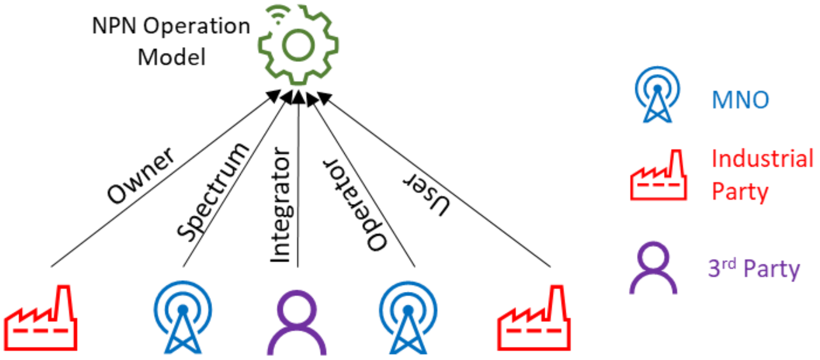


Figure 8 Operation Model 2

3.3.4 Operation Model 3

In this model, except for the 'NPN User' role, all roles are assigned to an MNO (as depicted in Figure 9). That is, two stakeholders are involved: an industrial partner and the MNO. While the former one is only the user of the NPN services, the latter one is responsible for all tasks of providing such services including deployment, integration, ownership, providing spectrum as well as operation and maintenance of the NPN.

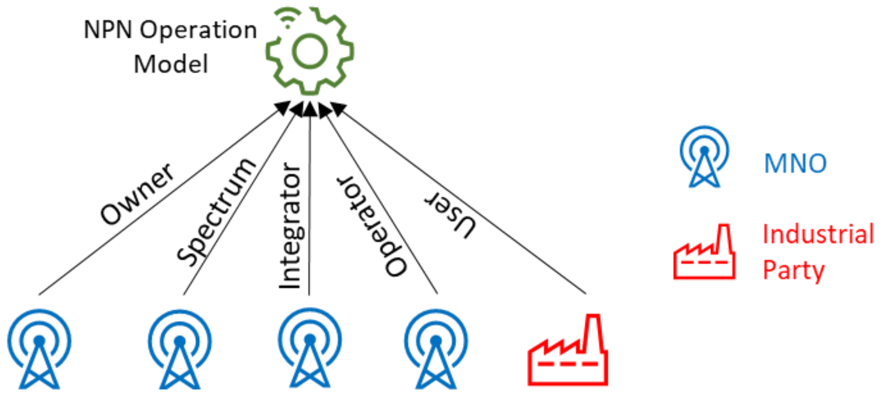


Figure 9 Operation Model 3

Above we have elaborated on various possibilities for designing an operation model. Several factors might affect the choice of operation models. These factors include usage simplicity, service continuity, liability, security, lock-in [AR5GF19], deployment flexibility and total cost [AR5GF19]. A comprehensive analysis of these aspects will be presented in a future deliverable.

Also, in defining the operation models above, we have not made any assumptions regarding the deployment models and possible interrelations between the deployment and operation models. In practice, however, there are some interdependencies between the two. For instance, in case of an PNI-NPN deployment, operation models 1 and 3 can be adopted, and operation model 2 would not be feasible. Also, in case of S-NPN deployment, all three operation models (operation models 1-3) can be applied.

4 Standardized 3GPP technical enablers

5G aims to support a wide range of verticals including smart manufacturing. In this section we deep dive into selected technical enablers based on the functional requirements of 5G-SMART use cases. The 5G system offers a number of services and features that are useful or even essential in the context of smart manufacturing use cases as listed below. D5.2 covers the investigation of the first four enablers. The final two will be handled in D5.4.

1. Network Slicing
2. 5G-LAN Type services
3. Time Sensitive Communication
4. 5G system exposure enabler towards OT system
5. QoS Management
6. Mechanisms for reliability

4.1 Network Slicing

A Network slice provides a logical network that runs on a common physical infrastructure and satisfies the service level agreement (SLA) for a certain category of service. It is important to keep an end-to-end perspective and an E2E network slice involves the Radio Access (RAN), the user/data plane and the 5G control plane.

Network slicing is motivated by either business or operational aspects:

1) A business agreement motivates to provide a dedicated slice, with functionality isolated from other slices and access to a guaranteed level of computer, storage, communication resources. An example is the PNI-NPN of type NPN3 or NPN4.
2) From an operational aspect, different service type categories have different building practices. E.g. a massive IoT service type has low data volumes and high cost pressure and a centralization of user/control plane (i.e. NPN4) is the most cost-efficient operation model with negligible transport-network costs for the traffic.

Network slicing has been defined as part of the 5G architecture in 3GPP. A UE may access multiple slices that are linked to the NPN where the UE is registered. The slices are associated to given Service-level Agreement (SLA) based on bit rate, latency, and packet loss. Figure 10 provides an overview of network slicing in 3GPP.

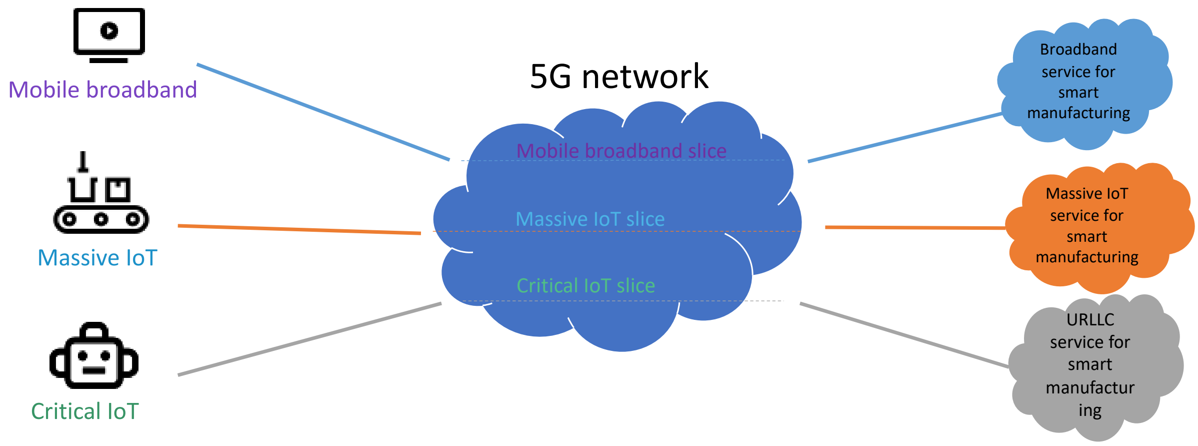

Mobile broadband
Massive IoT
Critical IoT
5G network
Mobile broadband slice
Massive IoT slice
Critical IoT slice
Broadband service for smart manufacturing
Massive IoT service for smart manufacturing
URLLC service for smart manufactur ing

Figure 10 Network slicing overview

In Release 16 the following basic Slice/Service Types (SST) have been identified

Slice/Service type	SST value	Characteristics
eMBB (enhanced Mobile Broadband)	1	Slice suitable for handling of 5G enhanced Mobile broadband, useful, but not limited to, the general consumer space mobile broadband applications including streaming of High-Quality Video or large file transfers.
URLLC (ultra-Reliable Low Latency Communications)	2	Slice suitable for handling of ultra- reliable low latency communications.
MIoT (Massive IoT)	3	Slice suitable for handling of massive IoT.
V2X	4	Slice suitable for the handling of V2X services.

Table 4 Network slicing type and characteristics

The network slicing as specified in 3GPP documents describes the process for creating network slices based on the application or service requirements.

4.2 5G support for Industrial LAN services

Communication among machines and production lines within a factory is based on industrial LAN networking. Different networking technologies are often deployed in a factory. Bridged Ethernet is the general LAN networking technology commonly used for this purpose. Larger subnetworks of the factory are interconnected via IP routing. Ethernet LAN is often complemented with fieldbus technologies or real-time industrial Ethernet variants, which can provide deterministic performance, e.g., on bounded latency. Over the recent years, a lot of efforts has been put into developing a converged, standardized industrial Ethernet technology under the umbrella of TSN, which introduces time-sensitive features into standard Ethernet. TSN is expected to replace over time legacy real-time Ethernet variants, and thereby the former “local real-time network segments” will be merged into the general Ethernet network.

Benefits of 5G are identified in use cases where 5G communication can be introduced not only in enterprise-wide networking, but also on the field level. This means that the 5G interworking with the industrial LANs, and in particular Ethernet, is essential.

There are several technology enablers defined by 3GPP that are accelerating the integration of 5G support with industrial LAN. These are

1. Ethernet connectivity support,
2. Time Sensitive Communication support for TSN integration and
3. Virtual Network groups that help in building a virtual LAN.

4.2.1 Ethernet connectivity support

The 5G supports the “Ethernet” type Packet Data Unit (PDU) session [3GPP20-TS23501]. With this, a UE can communicate via the 5G network directly with Ethernet frames. Enhancements have been made in the 5G core network to forward and handle such Ethernet frames towards the external data network.

4.2.2 Time Sensitive Communication

In Release 16, 3GPP has considered the entire 5G system to act as a TSN-enabled bridge. This is done in order for the 5G system to allow time sensitive traffic to pass through it, while it is integrated into the TSN-enabled Ethernet based industrial networks. By doing so, the 5G system guarantees reduced latency and end-to-end synchronization for applications relying on 5G URLLC services.

5G provides support for time-sensitive networking, including the following functionalities:

- Configuration of traffic handling and traffic forwarding for TSN (and non-TSN) Ethernet streams in the 5GS.
- Time-synchronization over the 5GS.

The 5G functions to support TSN have already been described in detail in [5GS20-D51], [IG+20] and [JBG+19].

4.2.3 Virtual Network Group

A 5G Virtual Network (VN) group consists of a set of UEs using private communication for 5G LAN-type services. A 5G VN group can be used for both IP- or Ethernet-based services, in e.g. private, residential, enterprise or industrial environments [3GPP18-22821] [3GPP20-22261]. It should be noted that 5G VNs are defined as a private group of 5G UE devices and it does not include non-5G devices. Figure 11 depicts the VN group mapping with external data networks identified with their corresponding Data Network Names (DNN) as one of the possibilities to realize the functionality, two DNNs as two VN group, each group contains set of UEs (orange and green).

The 5G VN group configuration is either provided by O&M or provided by an Application Function (AF).

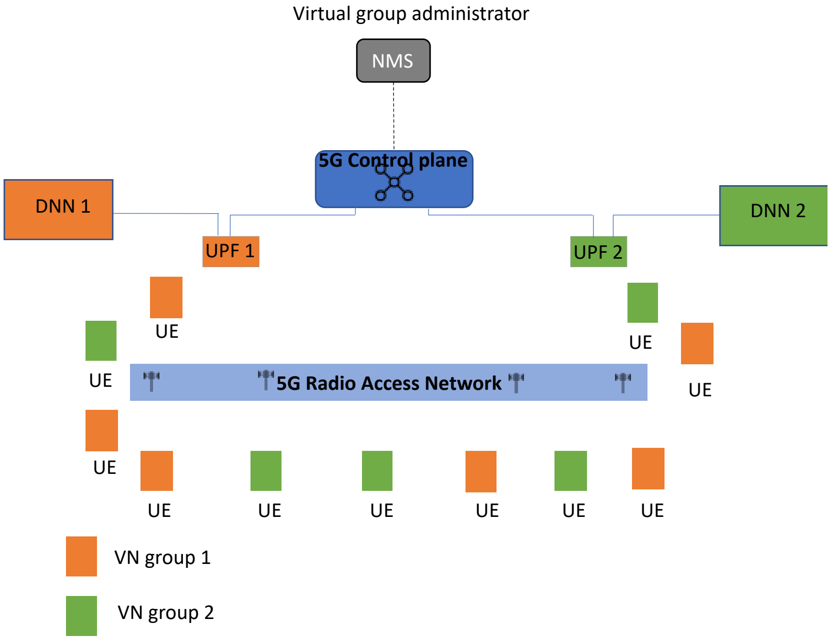


Figure 11 Mapping between DNN and 5G VN group

5GS provides switching and forwarding mechanisms based on using 5G VN group. Switching of the traffic can be done utilizing different mechanisms and at different interfaces in the 5G system. It might be local switching within user plane functions or can be between two user plane functions. An example for local switching for UE-to-UE communication is shown in Figure 12.

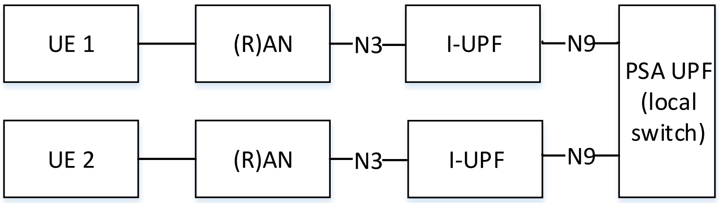


Figure 12 Local switch-based user plane architecture in non-roaming scenario [3GPP20-TS23501]

4.3 5G service and capability exposure enablers towards OT system

To enable seamless integration of 5G in smart manufacturing, service exposure and capabilities are seen as one of the important functional requirements also highlighted in table 1. 5G-ACIA has recently published a paper [SE5G20-5GACIA] on functional requirements imposed on 5G exposure interfaces specifically for device connectivity management and network monitoring.

As the 5G system is envisioned to support a range of verticals, 3GPP has specified a generalized service enabler architecture (SEAL) for a wide range of the verticals including industry 4.0 [3GPP20-23434]. SEAL architecture provides the northbound application programmable interface (APIs) for flexible integration of the OT application in case of the smart manufacturing verticals. In the architecture as shown in the Figure 13, a SEAL client is realized on a UE which communicates with a SEAL server. The

SEAL client supports interactions with a vertical application client. Above the SEAL layer, there is a vertical application layer (VAL) which communicates with the SEAL layer to expose functionality and information towards the vertical application (on client side with SEAL-C and on server-side SEAL-S). The SEAL server also interacts with the 3GPP core network via network interfaces as shown in Figure 13. SEAL enables a simplified implementation and flexible integration of the vertical application (OT application in our case).

Below are the main common capabilities defined by SEAL for the vertical application layer:

- Group management service.
- Configuration management service for UE configuration, User profile.
- Location management service for cell, service area and geo coordinates.
- Identity management service for user authentication and authorization.
- Network resource management service for switching unicast/multicast resource quality detection and service continuity.

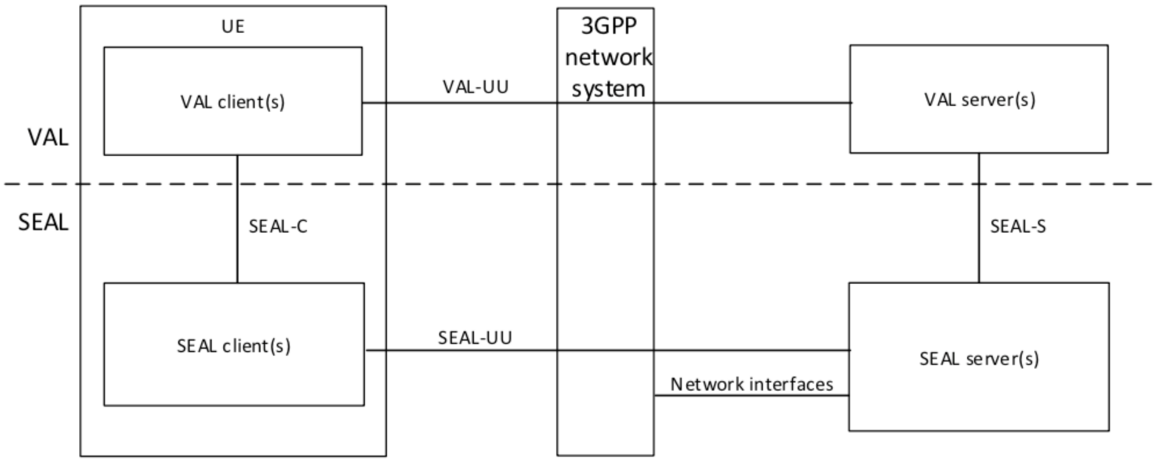


Figure 13 3GPP specified SEAL architecture [3GPP20-23434]

Considering the SEAL enabler for the smart manufacturing applications and integration with the OT system on the factory shopfloor, there is a need for the additional functionality such as QoS coordination, QoS monitoring, 5G-LAN group management, TSN support, clock synchronization, user authorization, device monitoring and integration with existing operational technologies (e.g. OPC-UA). Figure 14 shows a realization of the 5G system with the SEAL architecture for integration with existing smart manufacturing applications and the OT system. The detailed solution on how the service enabler layer will enable such exposure of the functionality is under investigation in Release 17.

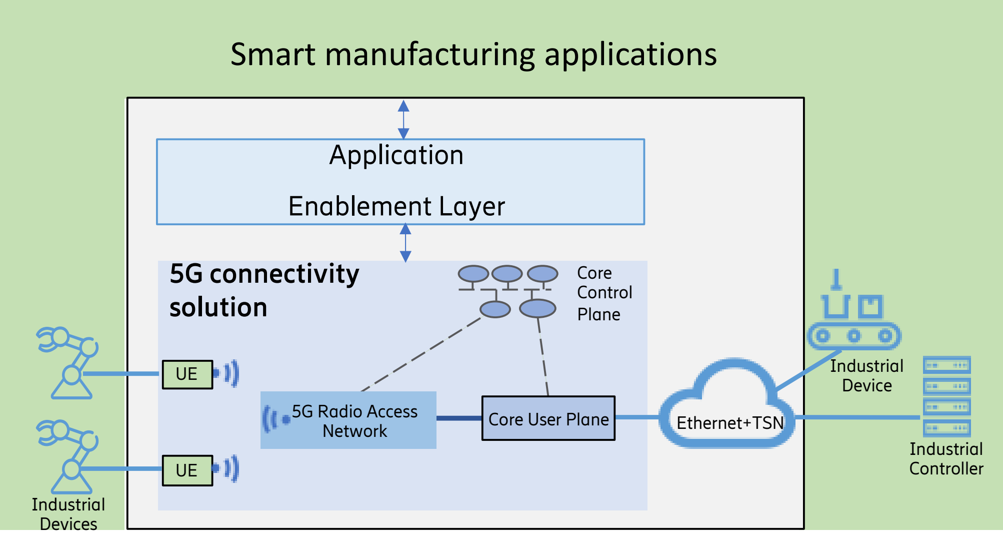


Figure 14 SEAL architecture integrated with 5G system for smart manufacturing

Edge cloud enablers

This section provides some insights how edge cloud architectures and industrial smart manufacturing can cooperate from the software architecture point of view. In this deliverable the interrelations with the 5G network architecture, and how edge computing integrates with 5G is not investigated, it is planned in D5.4.

4.4 Cloud Service Models for Smart Manufacturing

The primary typical application of cloud systems in smart manufacturing is to offload computation and storage intensive tasks into the cloud. A progressive next step is to offload not only the computation and storage intensive but all control and storage functionalities of the devices (e.g., robots, automated guided vehicles, or device controllers) into the cloud. This leverages the advantages of the cloud system where computing tasks are dynamic, and resources are elastic and available on-demand. This approach can also be used to enhance, and update on-demand the control capabilities of a device and allows building more cost-effective devices. Consequently, not only the computationally complex tasks can be offloaded from the devices requiring less powerful processors on board, but also the software development and maintenance costs are decreased due to the reduced functionality on the devices. If all control components run in the cloud, then the cooperation of the controlled devices can be easily realized within the cloud in between the control modules.

Cloud computing has three fundamental service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). IaaS allows the consumer to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. With PaaS the consumer is able to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. SaaS provides complete applications running on a cloud infrastructure and accessible over the Internet and licensed on a subscription basis.

More lightweight service models have emerged, such as the Container as a Service (CaaS) and Function as a Service (FaaS). CaaS is considered as a subset of IaaS because the basic resource for CaaS is a container, rather than a Virtual Machine (VM) in IaaS. CaaS is often running on top of IaaS, but this is not strictly necessary. FaaS provides service-hosted remote procedure calls that enable the deployment of individual functions in the cloud that run-in response to events.

The control functions that are to be run in the cloud are software components. The service models that let users deploy and manage such software components independently from the service provider are IaaS, CaaS, and FaaS.

4.4.1 Evaluation of Cloud Service Models for Smart Manufacturing

With IaaS virtual servers, the operating system, storage, and networking must be managed in addition to the software. These additional tasks can be a burden and means additional cost for the user and the VM images must be built for the applications. However, if the software application has some unique requirements that needs to be fulfilled only with specialized Virtual Machine images – like legacy control VMs for brown-field deployments – then this is the only way to proceed.

With CaaS the software and all of its dependencies are packaged into containers that are more lightweight than Virtual Machines. Because of the reduced overhead, more containers can be provisioned on a host than VMs and each service running in a container could have its own container image, its own release cycle, and its own rolling upgrades, allowing for smaller teams to develop them in parallel. However, the container images still need to be built and the orchestration of the containers is in the control of the user.

With FaaS even the repetitive tasks can be avoided, like building container images for every new application. In most of the cases there are still containers running in the background to serve the functions, however the user does not have to manage the infrastructure at all.

As the functions are actually event handlers and they react to triggers this kind of operation fits with smart manufacturing control applications. Typical FaaS systems impose restrictions on functions like maximum run-time and stateless operation, which makes complex long-running applications hard to build, using them, and that must be considered when designing the control system.

The right platform ensures the balance between flexibility and simplicity, allowing to build software faster and operate it automatically without being too constrained. Because of the operation of event based remote procedure calls, the FaaS platforms are not designed to support a wide variety of workloads and the flexibility provided by CaaS platforms, therefore the combination of CaaS and FaaS seem to be currently the optimal platform for smart manufacturing.

CaaS provides managed containers and the most wide-spread open-source container orchestration platform is Kubernetes. Kubernetes is becoming the de facto orchestration for enterprise containers. Kubernetes (K8s) is an open-source system for automating deployment, scaling, and management of containerized applications. Kubernetes has become a cloud-native standard and facilitates deploying complex applications using lightweight and portable containers. It supports workload abstraction, such as Deployment, Service or Job, etc.; and is capable of rolling upgrade and rollback of applications. Most of the open source FaaS platforms are also built on top of Kubernetes

4.5 Edge Cloud Scenarios

By the division of edge and central cloud clusters the locations of the cloud platform components are separated and also the hardware environment the cloud platform components are running on can be different as shown in Figure 15. Still there is a need for a federated management system for handling several cloud and edge clusters from a unified management interface irrespective of the physical hardware. Such a federated management can provide application-level software migration and portability between clusters with fault tolerance as well. The combined operation of edge and central clouds provides intelligent collaboration and flexible migration of services between the cloud and the edge. There are alternative architectural concepts that will be demonstrated on Kubernetes based implementations in the followings.

The presented options assume exclusive control authority of a single owner -- or fully coordinated set of owners -- and does not explore interoperability scenarios where different parts or subsystems are owned and controlled by different organizations, e.g., service providers.

We show some examples on how these options can be mapped to industrial use cases:

- A factory has an edge cloud (in one server room) and another edge or central cloud for redundancy in another room.
- A factory has an edge cloud per building/shopfloor, and a central cloud for the site with many buildings.
- An industrial user has one edge cloud per factory and one central cloud for multiple factories. This central cloud could be private cloud operated by the industrial user, or it could be a private cloud service provided by a cloud provider.

Further investigation will be elaborated in D5.4.

4.5.1 A fully functional cluster at the edge

One way for edge computing is to operate on its own, without interconnections towards remote central data centers/clouds, the other way is being connected with remote central data centers/clouds. A fully functional cluster at the edge need not address the federation of edge and central clouds and has no dependencies from a central cloud. It solves everything locally, by implementing a lightweight Kubernetes compatible edge solution. (Figure 15 upper part)

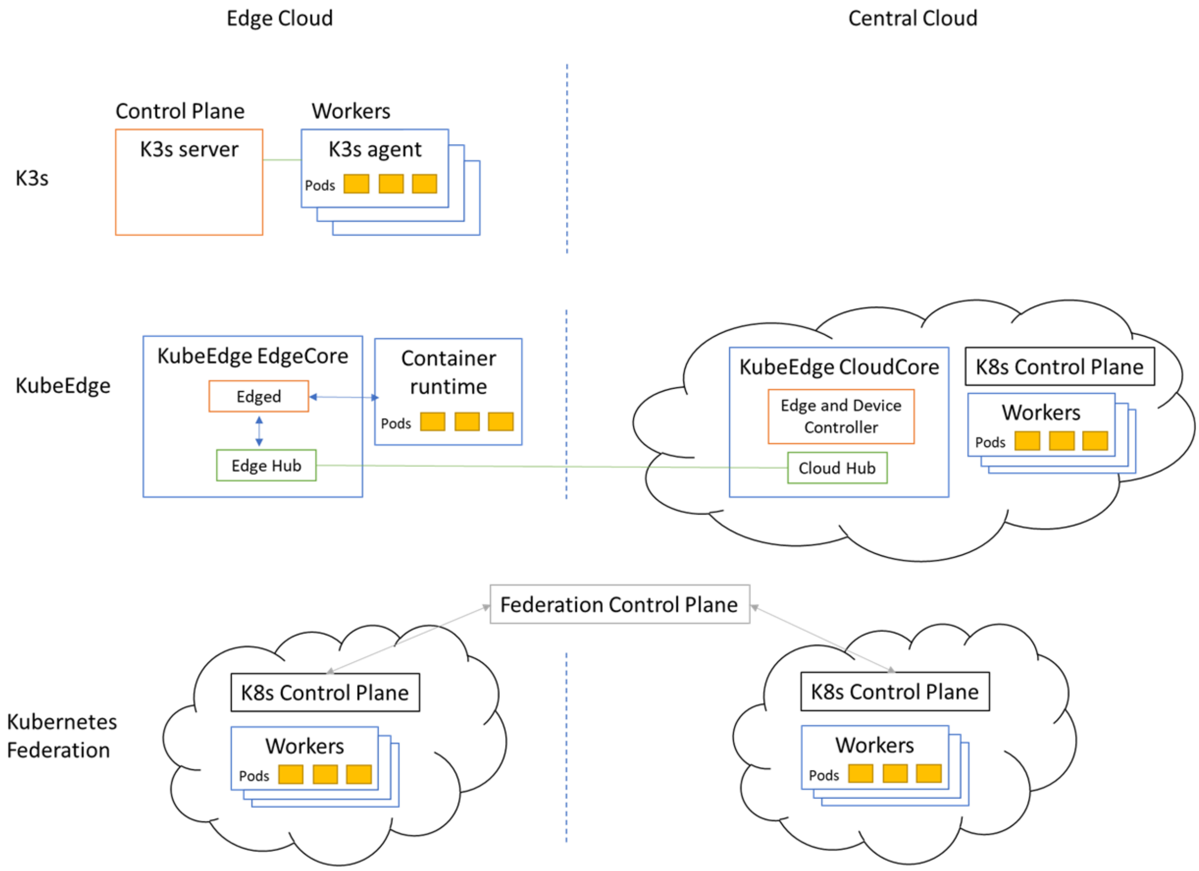


Figure 15 Edge cloud scenarios

K3s as reference architecture

K3s[3] by Rancher [K3s] is a lightweight, certified Kubernetes compliant distribution. The developers aimed to run Kubernetes in low, fixed resource environments. The k3s is packed into a single binary which is only 40 MB. It contains the main Kubernetes, container runtime and main system functions. The Kubernetes functions can be easily started with the *k3s server* or the *k3s agent* command, which not only starts the core Kubernetes components, but it also starts networking components (e.g. Flannel, CoreDNS and Traefik ingress). Normally a K8s cluster needs etcd server to store cluster data, the k3s uses SQLite instead of etcd. Due to the main focus is the IoT (Internet-of-Things) market, the k3s is optimized to run ARMv7 and ARM64 architectures, but it can be run on x86_64 also. A typical k3s architecture consists of two types of Kubernetes components. The server contains the SQLite database, kube-proxy, scheduler tunnel proxy and controller manager. The agents (workers) run tunnel proxy, flannel, kube-proxy, kubelet and containerd. As the K8s is aimed to build a High Availability (HA), production-ready cluster, one can build a HA, production grade cluster with the k3s also. It can have multiple servers and agent components, use external database, and have external load balancer also. K3s is designed for production environments, can be scaled out and it runs on any Linux OS.

As all components of K3S run on the edge, therefore no cloud-side collaboration is involved. If K3S is to be used in production environments, there should be a cluster management solution on top of K3S that is responsible for cross-cluster application management, monitoring, etc.

4.5.2 Cluster Control Plane in Central Cloud

The reference implementation for this kind of architecture is KubeEdge[4]. KubeEdge is built upon Kubernetes and provides core infrastructure support for networking, application deployment and metadata synchronization between cloud and edge. KubeEdge is made to build edge computing solutions to extend the central cloud. (Figure 15 middle part) KubeEdge consists of a cloud part and an edge part, both edge and cloud parts are opensourced. While the control plane for the cluster resides in the central cloud, the edge part can work in offline mode too. The edge cloud part is lightweight (66MB footprint and ~30MB running memory) and thus can support lower computing power hardware, both x86 and ARM. In this approach, the control plane resides in the cloud (either public cloud or private data center) and manages the edge nodes containing containers and resources. It can help to save setup and operational costs for edge cloud deployments.

As the Kubernetes control plane runs in the cloud, users can directly manage edge nodes, devices and applications from the cloud via Kubernetes tools.

Edge and Core Cloud components

KubeEdge builds interfaces that are consistent with Kubernetes, either on the cloud side or the edge side. In the central cloud the CloudCore component is deployed that contains the Edge and Device controller modules. They operate as an extended Kubernetes controller which manages edge nodes

[3] K3s Rancher, K3s – Lightweight Kubernetes, https://rancher.com/docs/k3s/latest/en/

[4] KubeEdge Project, https://kubeedge.io/en/

and pods metadata so that the data can be targeted to a specific edge node. The Cloud Hub module watches for changes on the cloud side and caches and sends messages to EdgeHub.

At the edge cloud the EdgeCore component is built up from the Edged module, an agent that runs on edge nodes that manages containerized applications at the edge, and the communication is provided via EdgeHub module with the CloudCore.

Regarding maturity level, KubeEdge (v1.4) is currently at the incubating project level in the Cloud Native Computing Foundation (CNCF) project list, which corresponds to be used by early adopters.

4.5.3 Kubernetes Federation

Kubernetes multi-cluster management focuses on giving a single view to interact with, and report on each separate cluster under management.

Kubernetes Cluster Federation (KubeFed for short) [5]allows you to coordinate the configuration of multiple Kubernetes clusters from a single set of APIs in a hosting cluster. (Figure 15 lower part) KubeFed aims to provide mechanisms for expressing which clusters should have their configuration managed and what that configuration should be.

When multiple clusters are federated then they share pieces of their configuration which is managed by the so-called host cluster. The benefit is that any resources configured to take advantage of the federation will treat all member clusters as a single distributed cluster.

Federation is used to run applications across multiple clusters, in this environment in edge and central clusters. The goal of the federation control plane is to simplify service and application administration over the clusters.

KubeFed is currently at alpha maturity level and moving towards initial beta.

4.5.4 Evaluation of Edge Cloud Scenarios

The preferred platform in general and also for smart manufacturing depends on the requirements on maturity, ease of deployment, available resources, operation and maintenance (see Table 4.). As Kubernetes itself, all these edge platforms are continuously over development and the maturity levels are around the same for all platforms (k3s is operational at the edge, however a cluster management component is missing that interconnects it with a central cloud, KubeEdge is at incubating project level and KubeFed is alpha). The order regarding the ease of deployment starts with k3s then KubeEdge and ends with KubeFed. From low to high resources requirements in the edge environment the order is k3s, KubeEdge and KubeFed at the end if there are enough resources to run full Kubernetes at the edge.

The operation and maintenance of Kubernetes and therefore KubeFed is complicated, compared to k3s or KubeEdge, but in central clouds there are usually dedicated operation and maintenance teams to handle this.

[5] Kubernetes Cluster Federation, https://github.com/kubernetes-sigs/kubefed

	K3s	KubeEdge	KubeFed
maturity	only edge components	incubating project	alpha
ease of deployment	easy	medium	hard
resource requirements	low	medium	high
operation and maintenance	simple	average	complex

Table 4. Summary of alternative edge cloud scenarios

4.6 Low latency options in the cloud stack

Critical control functions in a factory are time-sensitive services. When moving the execution of such functions into an edge cloud, a deterministic and low latency execution must be ensured. Cloud platforms are complex software stacks leveraging virtualization technologies; thus, the latency of cloud execution is influenced by both the edge cloud infrastructure and the resource management allocating processing and memory resources to execution units.

4.6.1 Infrastructure

It is recommended to use physical servers instead of Virtual Machines for the Kubernetes cluster. Physical servers do not have the hypervisor overhead layer that is common to virtual machines. As such, running containers directly on bare metal (physical servers) should offer faster performance. The actual difference in latency must be measured by the application communicating over the network but depending on the hypervisor type the difference can be even multiple times larger in case of VMs.

4.6.2 Resource management options

In Kubernetes, when specifying a pod, it can be optionally specified how much resource a container need. The most common resources to specify are CPU and memory (RAM). If the node where a pod is running has enough of a resource available, it is possible (and allowed) for a container to use more resources than its request for that resource specifies. However, a container is not allowed to use more than its resource limit.

Kubernetes, and other container orchestration systems using Docker, applies CFS (Completely Fair Scheduler) quotas to enforce CPU limits. CFS is the default process scheduler in Linux. CFS bandwidth control allows the specification of the maximum CPU bandwidth available to a group or hierarchy. The bandwidth allowed for a group is specified using a quota and period. Within each given “period” (microseconds), a task group is allocated up to “quota” microseconds of CPU time. Once all quota has been assigned any additional requests for quota will result in those threads being throttled. Throttled threads will not be able to run again until the next period when the quota is replenished. Setting CPU limits are enforced via CPU quota and, unfortunately, this can lead to unnecessary throttling, which leads to higher latency.

As an example, an application is running on a CPU with *cgroup* constraints. This application needs 200 milliseconds of processing time to complete a request. Unconstrained, the request is completed in

200 ms. However, if we assign a CPU limit of 0.4 CPU to the application, this means the application gets 40ms of run time for every 100ms period—even if the CPU has no other work to do. The 200ms request now takes 440ms to complete.

If there is no CPU resource limit set, one pod can cause the instability of the host and impact other pods, because it may be using all the CPU and networking resources for a heavy workload, this effect is also called the noisy neighbor problem. Although it is generally recommended to set resource limits (CPU, memory) for pods, turning off CPU quota might be the best approach to achieve low latency in most Kubernetes clusters running trusted workloads. In addition to turning off CPU quota, the pod can be pinned to one or several dedicated CPUs (cpuset), which limits its resource usage only to those CPUs.

These aspects show that even if the 5G network fulfils the latency requirements the setup and parameter settings of the cloud software stack for a given network deployment needs to be also considered when designing and end-to-end 5G system for low latency application.

4.7 Robot Operating System in containerized cloud environment

Robot Operating System (ROS) [6]provides libraries and tools to help software developers create robot applications. It provides hardware abstraction, device drivers, libraries, visualizers, message-passing, package management, and more. ROS is licensed under an open source, BSD license and is one of the most popular prototyping platforms for developing robots.

ROS creates a peer-to-peer network of processes (potentially distributed across machines) that are loosely coupled using the ROS communication infrastructure. ROS implements several different styles of communication, including synchronous RPC-style communication over services, asynchronous streaming of data over topics, and storage of data on a Parameter Server. ROS currently only runs on Unix-based platforms.

ROS has two versions, ROS1 and ROS2. ROS1 has become popular among the open source robotics community. ROS2 is a complete refactoring of ROS as ROS2 was developed from scratch not to break ROS1 stability for implementing the important missing features that were accumulated during ROS1 development, such as multi-robot systems, real-time operation and production level stability. Thus, the two versions are not compatible and differ significantly.

A ROS node is a process that performs computation. Nodes are combined together into a graph and communicate with one another. Nodes are meant to operate at a fine-grained scale, controlling separate parts of a robot, e.g. laser scanner, wheel motors, etc. Thus, a robot control system will usually comprise many nodes.

Kubernetes Pods are the smallest deployable units of computing that one can create and manage in Kubernetes. A pod is a group of one or more containers, with shared storage/network resources, and a specification for how to run the containers. If a pod runs a single container then the pod works as a

[6] Robot Operating System, https://www.ros.org/

wrapper around a single container; Kubernetes just manages pods rather than managing the containers directly.

The straightforward mapping is that a ROS node is encapsulated into a container, then into a pod. Because of the architectural differences between ROS1 and ROS2 the integration with Kubernetes differs in several points.

4.7.1 ROS1

In ROS1, there is a ROS master that provides naming and registration services to all the other nodes. The role of the master is to enable individual ROS nodes to locate one another. Once these nodes have located each other they communicate with each other in a peer-to-peer fashion. Thus, first each newly connected node communicates with the master, and after that it can communicate with other nodes directly.

For this communication bi-directional connectivity is required between the nodes.

The reachability of ROS master is defined by URI, e.g. if one has a more complicated ROS setup, such as a ROS master running on another machine, the ROS_MASTER_URI or ROS_IP environment variables need to be changed.

At start-up, ROS nodes check that the configured server name and the associated IP address match with its own address, if not, it refuses to start. To resolve this, a so-called headless service can be created in Kubernetes because it creates a domain name that points to the IP address of the Kubernetes pod. It is also necessary to set a ROS_HOSTNAME environment variable, the domain name which will be sent to the master as well as to the other components.

If ROS nodes running inside a Kubernetes cluster communicate with the outside world, e.g. a robot, the ports of the ROS nodes must be exposed as external service from the cluster.

The distributed mechanism of ROS1 requires a good network environment to ensure data integrity, and the network does not have data encryption, security protection and other functions. Any host in the network can obtain the message data issued or received by the nodes.

4.7.2 ROS2

The networking operations significantly differ between ROS1 and ROS2. The ROS1 data transport protocol uses TCPROS/UDPROS, and communication is highly dependent on the operation of master node. At the core of ROS1 is an anonymous publish-subscribe middleware system that is built almost entirely from scratch. Communication in ROS2 is based on DDS (Data Distribution Service) standard as a middleware layer, enhancing fault tolerance capabilities. ROS2 can work with different vendors of DDS like FastRTPS, RTI-Connext, OpenSplice, and more. A consequence is that there is no master node in ROS2. Compared to ROS1, ROS2 is in the early stage of development.

As DDS is based on sending multicast UDP packets the Kubernetes cluster networking has to support multicast also, e.g. WeaveNet supports it, but Flannel or Calico does not. DDS has its own discovery service, therefore DDS pods can discover and establish connections with each other by topics, and therefore the Kubernetes service object might not be needed.

To communicate between the outside world and the cluster due to DDS, the ROS2 nodes must be in the same network, as the nodes send multicast UDP packets to the network. In this case, it is necessary to establish a VPN (Virtual Private Network) connection, on which the sending of multicast packets is allowed.

While in the case of ROS1 it was necessary to set the availability of the master server for all nodes, this is no longer needed here. Instead, an integer must be specified in the ROS_DOMAIN_ID variable, and from this number, the ports used by ROS2 nodes are calculated by given formulas.

Further, cloud related analysis and impact to it 3GPP technical enablers will be investigated in D5.4.

5 Deployment validation and analysis

The current section provides a deployment validation analysis of the technical enablers explained in the sections above, when realized in different deployment models. Also, functional requirement and use cases are taken as basis for validation analysis. Gaps are identified in the current 3GPP Release 16 solution.

The section covers a wide range of topics, as shown in Figure 16 deployment validation and gap analysis. Following topics are elaborated

- Analysis on how native 5G connectivity (5G system not optimized for industrial IoT) can support 5G-SMART use cases with different NPN deployment options,
- Investigation on how 5G Ethernet techniques interworks with IEEE 802.1 Ethernet networks is performed. Solutions are proposed for the seamless support of 5G integrated Ethernet network configuration and 5G to support end host to host communication,
- 5G support for TSN is examined, how integrated 5G-TSN can support 5G-SMART use cases is shown. Investigation on how 5G-TSN integration with different NPN deployment options can be realized,
- Analysis of deployment options while enabling end to end 5G time synchronization and positioning,
- Security zoning aspects in integrated 5G Ethernet industrial network with VLAN configuration and network slicing.

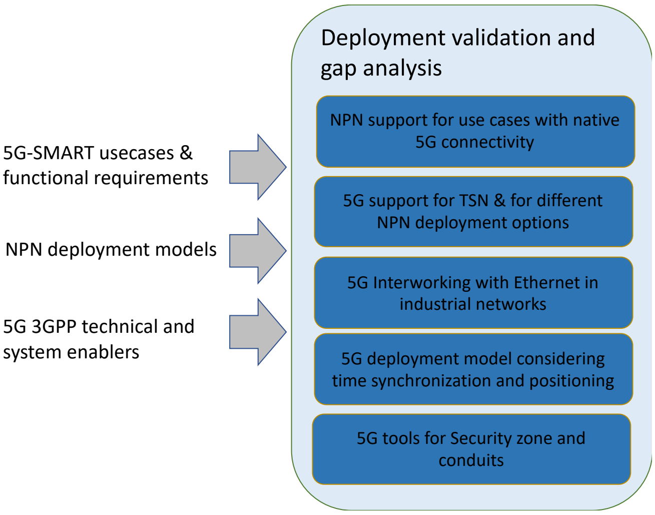

Deployment validation and gap analysis
5G-SMART usecases & functional requirements
NPN deployment models
5G 3GPP technical and system enablers
NPN support for use cases with native 5G connectivity
5G support for TSN & for different NPN deployment options
5G Interworking with Ethernet in industrial networks
5G deployment model considering time synchronization and positioning
5G tools for Security zone and conduits

Figure 16 deployment validation and gap analysis

5.1 Use cases with native 5G connectivity

This section describes how industrial use cases that are not based on TSN can be realized and are natively connected to 5G. By native 5G connectivity it is meant for the 5G system which are not optimized for the industrial IoT use cases needing support for Ethernet and TSN. For this we consider the use cases 1 to 6 from D1.1 [5GS20-D11] as reference as shown in Table 5. The main criteria that drive deployment consideration are E2E latency, security/privacy, E2E reliability and availability.

Use case no.	5G-SMART Use case
UC 1	5G-Connected robot and remotely supported collaboration
UC 2	Machine vision assisted real-time human-robot interaction over 5G
UC 3	5G aided visualization of the factory floor
UC 4	5G for Wireless acoustic workpiece monitoring
UC 5	5G versatile multi-sensor Plattform for digital twin
UC 6	Cloud-based mobile robotics

Table 5 5G-SMART use cases

The E2E latency is a very important consideration for deployment choices. Note that we are talking about E2E latency from device to the edge application. This clearly has a big impact on the user plane. We may broadly classify by E2E latency into 3 classes.

Class	Range
Low latency	< 3 ms
Medium latency	3-10 ms
High latency	> 10 ms

For instance, really low latency requirements around 1 ms can only be fulfilled in 5G deployment with a local user plane (UPF). At the other end high latencies allow some leeway to move to a PLMN hosted model even for the user plane. None of the principal use cases of D1.1 [5GS20-D11] (use cases 1 to 6) require low latency. Most of them fall into the medium latency category and a few into the high latency category. For the medium latency use cases, when considering a solution purely from the latency point of view it is feasible to use NPN options 3 and 4 if the PLMN operator deploys the network functions (NFs) including UPF close to the network edge in order that medium latency can be reliably supported.

Security/privacy related issues rank at the same level as E2E latency in their influence on deployment and operational models that are appropriate for different services. In most cases the edge cloud/application has high requirements of security and privacy. Hence, they either need to be located on premises or the MNO/NPN operator needs to provide hosting that guarantees the security and privacy via service level agreement (SLA) which may necessitate dedicated resources. The NPN options where the user plane is local i.e. NPN1, NPN2 and NPN3, should closely support requirements concerns by OT regarding security and privacy. The control plane does not pose the same kind of challenges as TSN and the LAN does not come into play in these use cases in such a tight manner.

Reliability in an E2E context is complex to evaluate as it depends on the number of elements in series (which reduces the reliability) and the parallelism/redundancy that is added to augment the reliability. For high reliability the number of elements in the end-to-end chain should not be too high and this works in favour of either having a local or edge deployment. From the cost point of view, the redundancy that is needed to achieve very high reliability targets favours sharing the resources with a PLMN. For instance, at the RAN level, providing redundancy of gNBs may be cost effectively achieved by MOCN or MORAN style sharing. Future work is planned to build models that can provide quantitative figures and guidance for the deployment choices against 5G-SMART use cases.

Another issue to be kept in mind is that real use cases are often a combination of URLLC, eMBB and sometimes mMTC communication services. Use cases 1, 2 and 3 involve wall mounted cameras sending video image and depth data for building a map of the factory floors in addition to stationary and mobile robots. This is a combination of eMBB and URLLC type of service categories. If we use a SNPN for the URLLC type data flows, will this SNPN also support the eMBB camera traffic? Here again quantitative analysis needs to be done. Hence hybrid NPN options with SNPN for URLLC and NPN3 for eMBB could be considered. D5.4 will investigate this issue. Two other aspects have an impact on the deployment choices are time synchronization and positioning as discussed in section 5.5.

5G system seamlessly support IP communication services including QoS. Considering the requirement of the use cases, there is need for the integration of the 5G system to support layer 2 communication. Next section provides further details on the same.

5.2 Interworking with Ethernet in industrial networks

It is of utmost importance that 5G technology support for Ethernet must be compatible with the IEEE 802 standards so that the 5G Ethernet techniques enable seamless integration with existing Ethernet deployment e.g. LAN integration and TSN deployments for industrial automation. It is understood that IEEE 802.1 Ethernet bridging principles equally apply to generic LAN integration and TSN traffic in practical deployments. It has been already agreed in Release 16 that the 5G system is integrated into the TSN network as bridge on a per UPF granularity [5GS20-D51]. The current section deep dives into three points essential to enable interworking with Ethernet in industrial networks.

- 5G integrated Ethernet network configuration industrial networks,
- 5G-LAN type services,
- 5G support to support end host to end host communication (UE to UE communication).

An example illustration of an Ethernet bridged network including 3GPP components is shown in the Figure 17. Based on the IEEE Ethernet standard 802.1Q all end stations are capable of communicating with each other through bridges. Details on bridge forwarding process is provided in the appendix. In this example there are four Ethernet bridges, out of which two are realized by 3GPP 5GS bridges. In Release 16, support has been added to make the 5G system look like a bridge to the TSN network management system.

Note that the 5GS Ethernet bridge is per UPF. The example in Figure 18 illustrates a tree Ethernet topology of the active links; there may be other physical links besides the links in the tree which are inactive. Additional Ethernet bridges may be present on both the sides of the 5GS TSN bridge. In general, 3GPP should support arbitrary Ethernet topologies. The forwarding in an Ethernet network may be realized by the general Ethernet flooding mechanism in combination with MAC learning. Alternatively, the forwarding may also be set by a central network controller such as the Network Management System (NMS) / Centralised Network Configuration CNC, in which case the NMS/CNC populates the Filtering Database (FDB) with the entries that are used for frame forwarding.

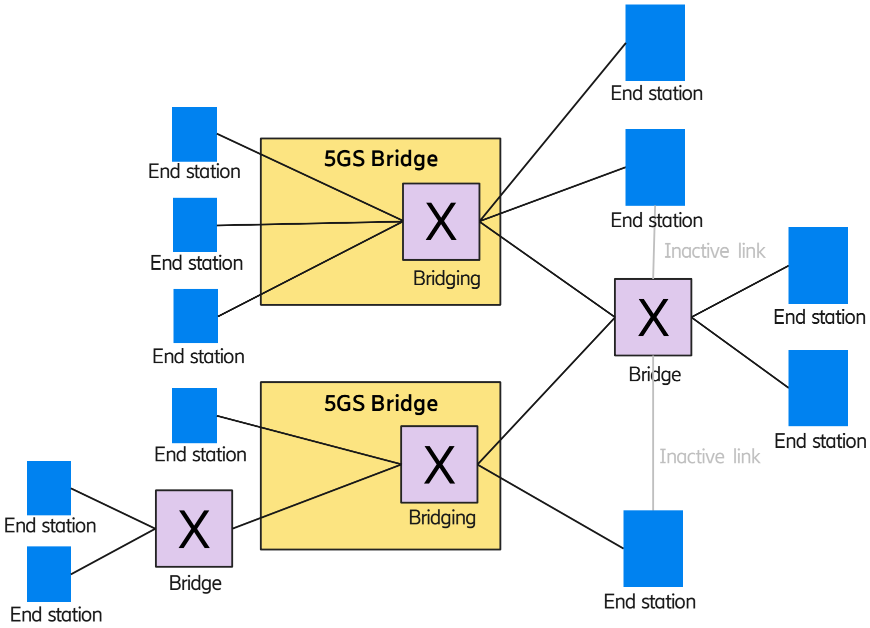


Figure 17 5G integrated Ethernet network architecture

5.2.1 5G integrated Ethernet network configuration for industrial networks

In 3GPP release 16, 5GS supports the fully centralized TSN configuration model, where the same central controller needs to be able to configure both Ethernet and 5G bridges as shown in Figure 18 as a unified network [5GS20-D51]. Otherwise, the administrator of the industrial network would need to configure distributed bridges individually using different mechanisms for the network configuration of the 5G bridges and the fixed bridges, which can significantly increase the operational cost and can make it difficult to achieve a consistent network configuration. Besides bridges that are managed by centralized controller, in an industrial network there may also be legacy Ethernet bridges that do not support centralized configuration, therefore in order for the 5G system to support the whole industrial network, both MAC learning and flooding based forwarding as well as the static forwarding configured by the central controller need to be supported.

Industrial networks require a dynamic virtual LAN (VLAN) configuration. Such VLAN separation can help to logically partition the network, e.g., for better security protection and for easier management; or to make it easier to establish dedicated user plane paths for specific streams. For a centrally managed Ethernet network, VLAN configuration has to be centrally managed, so that it can be harmonized with forwarding rules and QoS settings; furthermore, dynamic changes and updates in VLAN set-up should be allowed to be made by the central configuration.

In a summary, a 5GS bridge requires the following mechanisms to support industrial communication:

- flooding,
- MAC learning,
- IEEE 802.1x security e.g. X.509 certificates, Access Control Lists (ACL),
- forwarding traffic between any two bridge ports without pre-configuration,
- The NMS/CNC has the ability to configure static filtering entries for all bridges and for all communication directions, including the 5GS bridge, as specified in IEEE 802.1Q,

- The NMS/CNC has the ability to configure the VLAN handling for all bridges and all ports, including the 5GS bridge, as specified in IEEE 802.1Q

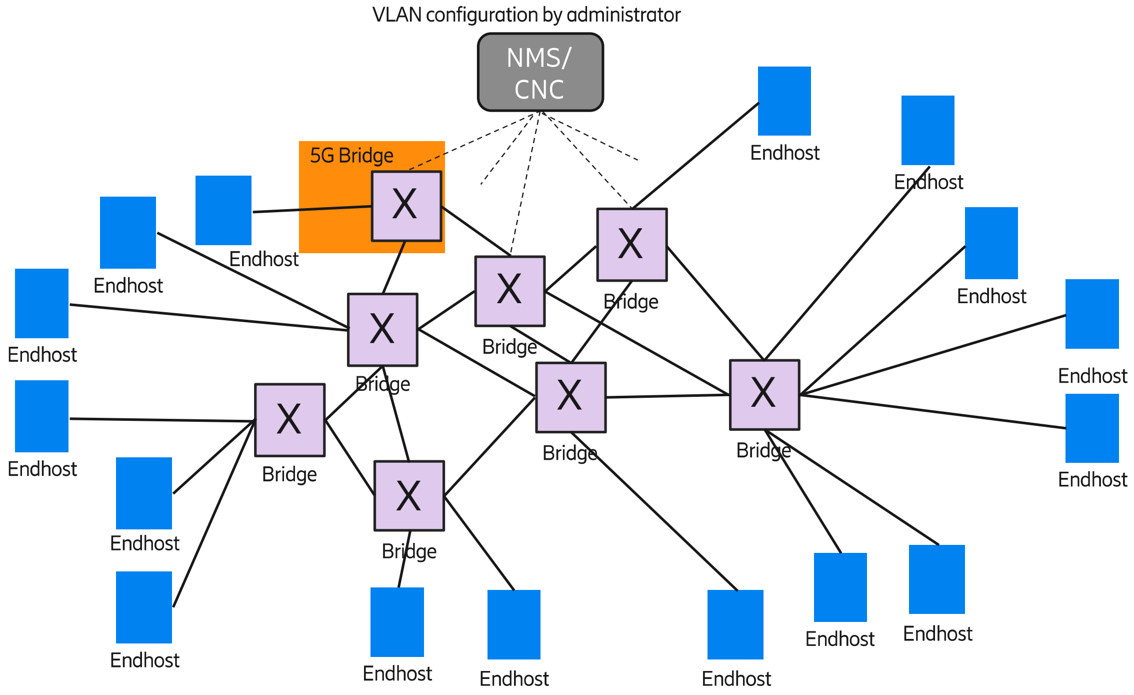


Figure 18 Simplified VLAN configuration in integrated 5G Ethernet network

Gap analysis

The following aspects are proposed to enable simplified configuration of 5G bridge, which have been proposed and are under discussion in 3GPP SA2 Release 17 at the time of this deliverable.

1. provisioning of Time Sensitive Communication Assistance Information (TSCAI)

   In the Release 16 specifications the 3GPP system obtains TSN stream information indirectly using IEEE 802.1Qci per-stream filtering and policing (PSFP) information (see section 5.3 and [5GS20-D51]). The CNC does not communicate directly characteristics of applications or TSN Streams towards the bridges, and instead provides PSFP information that the 5GS uses to extract traffic characteristics of every TSN stream. As a consequence, this information is only available when PSFP is configured by the CNC. There is a new proposal [FBM20] to specify a communication between the 5GS and the CNC via the network management protocol using subscribe/notify messages such that the CNC directly provides the assistance information to the 5GS, among other parameters related to the TSN stream. [FBM20] is being considered in the ongoing IEEE P802.1Qdj standardization activity on configuration enhancements for TSN. Also, this solution is under discussion in 3GPP SA2 Release 17, at the time of writing this deliverable.

2. Dynamic VLAN configuration

In Release 16, only statically preconfigured VLAN settings are defined. For a centrally managed Ethernet network, it is required that the NMS/CNC can configure the VLAN handling for all bridges and all ports, including the 5GS bridge, as specified in IEEE 802.1Q. In 3GPP SA2 Release 17, one of the solutions (solution 20 in [3GPP20-23700]) under discussion at the time of this deliverable precisely proposes that the VLAN configuration is provided by the CNC and is set on a per port granularity.

3. Forwarding traffic between any two bridge ports without pre-configuration.
   - The 3GPP Release-16 VN group-based traffic forwarding has very limiting restrictions, and therefore it is not applicable for general use in Ethernet networks. The realization of packet replication in Ethernet networks is implementation specific in the general case. In the user plane, the 5GS bridge performs frame forwarding as specified in IEEE 802.1Q. The frame forwarding should include the ability for flooding and MAC learning,

5.2.2 5G-LAN type services

As described above, interworking with existing Ethernet based industrial network is very crucial for integration of 5G networks into the OT domain. Ethernet is very widely adopted technology in the industrial networks, 3GPP has already done efforts to realize the Ethernet communication services by specifying Ethernet PDU session and LAN type services (section 4). The 5G-LAN type services as specified in 3GPP need to be investigated with regard to their applicability for manufacturing use cases. Listed below are the main gaps identified for the 5G-LAN type services and shortcomings that have been observed.

Gap analysis

Identified gaps in 5G VN solutions	**Shortcoming in interworking with Ethernet in industrial networks**
With realization of the 5G VN group where session management function (SMF) has full control of the Ethernet network topology, no interworking is defined with the fixed Ethernet domain.	This makes it hard to integrate the 5G network with existing industrial networks where Ethernet is widely used.
How to handle the Ethernet user plane forwarding is not fully specified in the current specification.	Not applicable for Ethernet which requires flooding for unknown addresses.
There is no support for forwarding a broadcast/multicast packet with source address not known to SMF/UPF.	Not applicable for Ethernet which requires broadcast/multicast to be supported irrespective of the source address.
Each UPF supports one N6 interface instance towards the data network, or only supports N19-based forwarding without N6.	Not applicable for TSN networks which can have multiple N6 interfaces
Multicast group formation of selected members of a 5G VN is not described in the Release 16 of the 5G specification.	Not aligned with TSN networks which often use multicast destination addresses but flooding needs to be avoided.

Table 5 5G Virtual Network groups gaps and short coming in interworking with Ethernet in industrial network

5.2.3 5G to support end host to end host communication (UE to UE communication)

Section 4.2.3 discusses support for 5G-LAN type service. A natural way to do UE-to-UE communication would be to use this service. As long as the UEs belong to the same 5G VN group the end devices can communicate using either local switching or N19 switching (see Figure 3). Note that in both cases there is a double traversal of the 5G network – from UE1 → gNB → UPF [→ UPF] → gNB → UE2. This has implications on the latency, reliability and time synchronization.

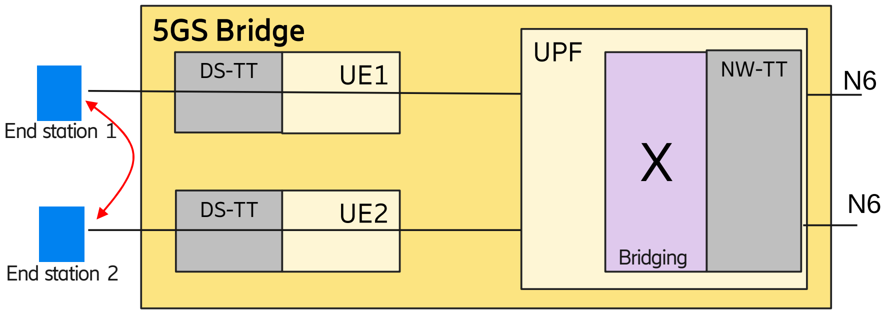


Figure 19 UE to UE communication

Gap analysis

5G VN groups are based on a 1:1 mapping between DNN and a group (Release16). The support of groups involves a common SMF controlling the UPFs for the 5G VN group. Ethernet type PDU(s) containing a VLAN tag shall be switched only within the same VLAN by a PDU Session Anchor. In a general UE to UE communication cases, the end devices will not necessarily be on the same VLAN.

However, when considering the TSN network integration with 5GS, the 5G system acts as an Ethernet Bridge on a per UPF granularity, as illustrated in the Figure 19. An end station may connect via Device Side-TSN Translator (DS-TT's) directly, or via one or more Ethernet bridges. In order to handle the UE-UE communication in time sensitive communications, different solutions are being evaluated in 3GPP SA2 Release 17 at the time of this deliverable. One of the solutions, (solution #10 in [3GPP20-23700]) is to use the fact that CNC provides static forwarding rules to the 5GS, and the UPF will feature the bridge forwarding functionality for traffic from any port to any other port of the 5GS bridge.

As described in the appendix, in an Ethernet network all end stations can always communicate with each other. It follows that end stations 1 and 2 as shown Figure 19 can always communicate; in other words, UE to UE communication is possible without any restrictions. For the 5GS bridge, this corresponds to communication between two bridge ports. So, in an Ethernet network it is neither necessary nor possible to restrict communication between UEs; such restrictions are not compliant with Ethernet bridging principles.

Considering 5GS integrated in the Ethernet network as bridge, there are mechanisms to configure communication between two end stations, which are:

- Utilizing DNN, it is possible to control UEs connecting to a given Ethernet network. This is can be realized by group management defined in 5G VN. It is possible to use session based DN authorization of the connectivity
- It is also possibility to separate the network into VLANs. End stations in the different VLANs cannot communicate with each other.

5.3 5G support for TSN

Time Sensitive Networking is foreseen as prominent standard for industrial wired communication for deterministic service. 3GPP has been working on enabling TSN support, detail analysis is available in deliverable D5.1 [5GS20-D11]. Integrated 5G-TSN can enable holistic communication from field devices to the cloud covering all the connectivity segments across industrial automation networks. An overview of the 5G interfaces regarding the integration with industrial networks is shown in the Figure 3.

5.3.1 Integration of TSN in 5G based industrial automation networks

TSN is a LAN technology which extends Ethernet to make it more deterministic in order to support time sensitive applications. Its use is typically for factory floor applications, which focus on deterministic communication and time synchronization rather than large throughput.

TSN is a set of open standards (more than 10) specified by IEEE 802.1 primarily for Ethernet 802.3 which means it can be used on a standards-based Ethernet switch. TSN can be seen as a toolbox providing bounded latency, time synchronization, reliability, traffic shaping and resource management.

A centralized configuration models for TSN is specified in IEEE 802.1 Qcc, a TSN bridge can have separated control and data planes. However, there are many interactions between the two in order to setup filtering, traffic shaping, scheduling, time synchronization, etc. On the control plane the bridges interacts with the CNC/CUC. Figure 20 below shows the centralized configuration model of TSN management.

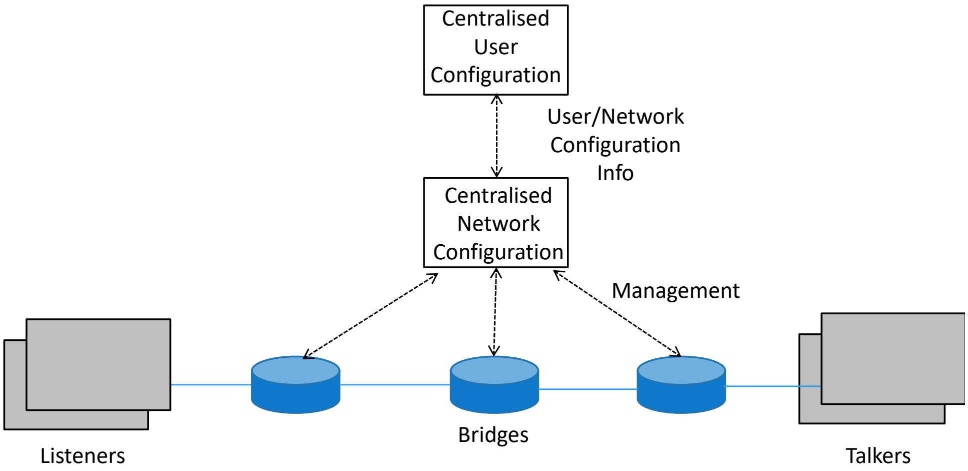

Centralised User Configuration
User/Network Configuration Info
Centralised Network Configuration
Management
Listeners
Bridges
Talkers

Figure 20- TSN centralized management

Various low level (L2) control traffic flows are realized between the bridges and the CNC. These include Link Layer Discovery Protocol (LLDP), Address Resolution Protocol (ARP) for the support of discovery of devices and their capabilities. It also includes neighbouring bridges information. These low-level protocols reveal a lot of information on the underlying network and are not secured and hence they are not supposed to leave the local LAN.

Typically, there can be three connectivity segments in the industrial automation network:

1 The devices on the factory floor,
2 TSN backbone and
3 Enterprise Ethernet network including a central room / edge cloud

The central room is where centralized control and management functions are located, like centralized programmable logic controllers (PLCs), CNC and automation data collection. The industrial communication network backbone is the TSN network that connects the machines on the factory floor to the central room. The industrial devices are in the OT domain. The TSN bridges essentially make up a network meant for the OT domain though it may have some overlap with the Enterprise Ethernet LAN. The controllers are in the central room. Figure 21 shows such a deployment.

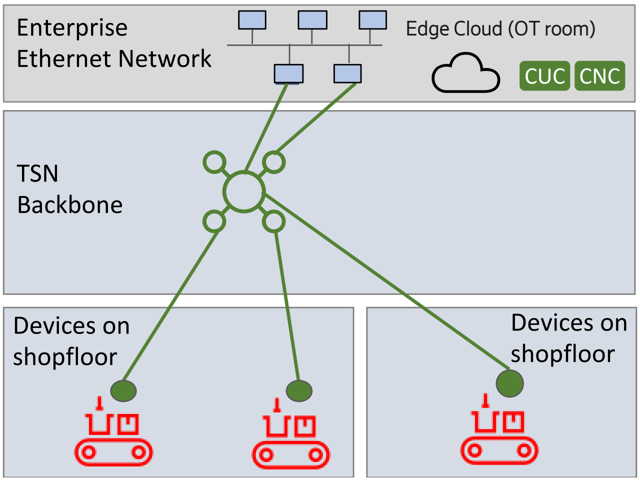


Figure 21 High level view of industrial network deployment

When 5G is deployed in a factory the radio access network spans over the entire shopfloor. With regard to the network segments in Figure 21, the 5G network is primarily providing a wireless backbone to the industrial communication, as shown in Figure 22. However, the 5G system can also provide connectivity directly between field devices. This corresponds to the connection between two device-side ports of the 5GS, according to the 5GS being represented as 5GS bridges towards the Ethernet/TSN network.

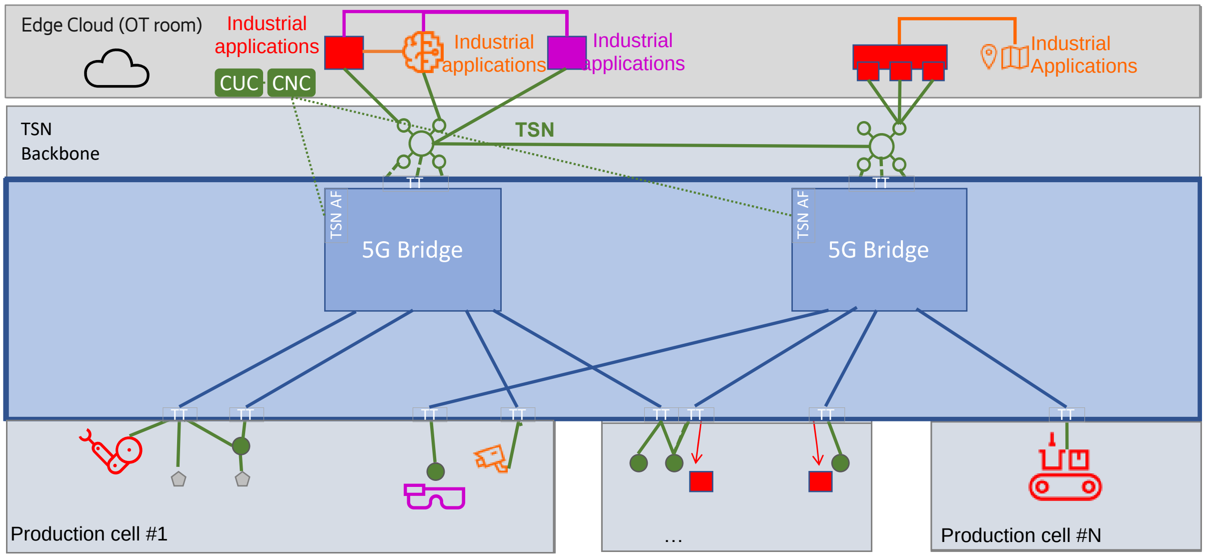


Figure 22 5GS as a bridge in a factory deployment.

For understanding how 5G networks needs to interwork with the industrial network and TSN, it is helpful to clarify how the industrial use cases are realized in a practical deployment. For this we look into the use cases described in [5GS20-D11] [5GS20-D21] [5GS20-D32]. A number of different use cases are depicted in Figure 23, Figure 24, Figure 25 and Figure 26. Different types of communications are used for the different use cases:

- Controller-to-device (C2D): connecting devices like sensors, actuators, drives to controllers
- Controller-to-controller (C2C): connecting diverse controllers for coordination
- Device-to-Compute (D2Cmp): connecting devices to the compute infrastructure

The types of communication correspond to different traffic categories with different requirements as discussed later (see e.g. Figure 29).

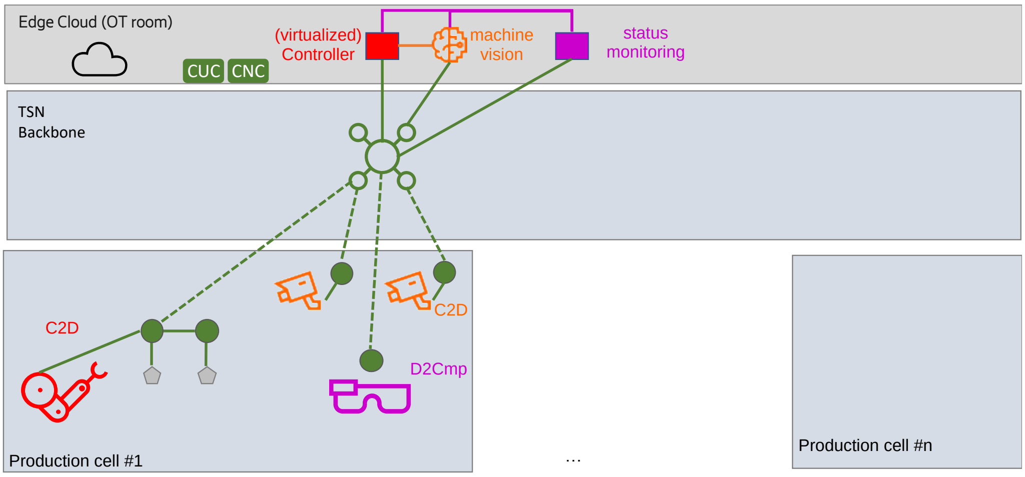


Figure 23 Deployment for use cases on industrial robotics for controller-to-device and device-to-compute use cases (use cases 1-3 in [5GS20-D11])

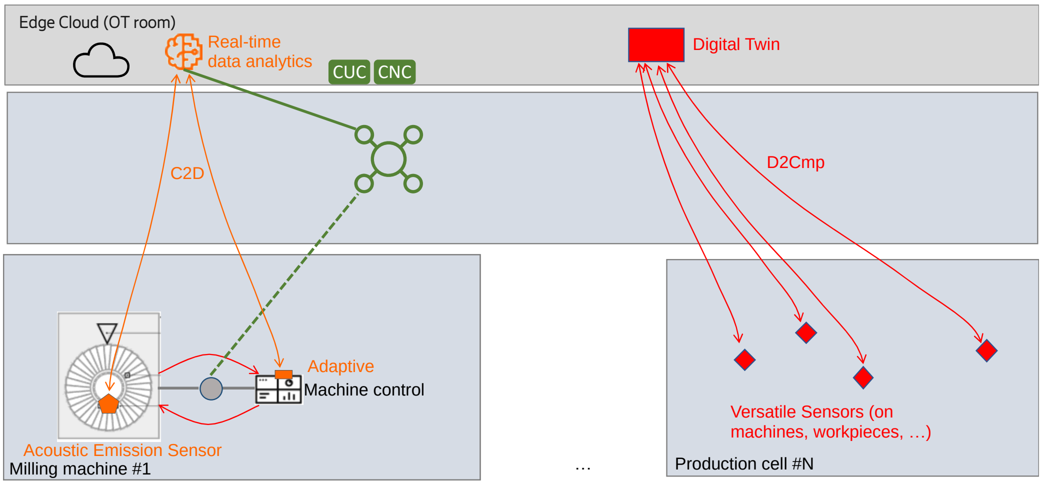


Figure 24 Deployment for use cases on enhanced machining (use cases 4-5 in [5GS20-D11]

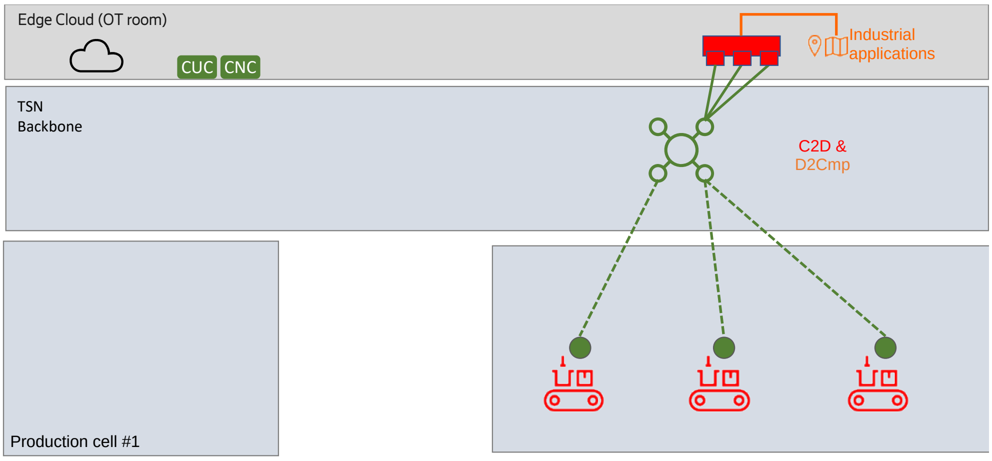


Figure 25 Deployment for the use case on cloud-based mobile robotics considering controller-to-device and device-to-compute communication (use cases 6 in [5GS20-D11]

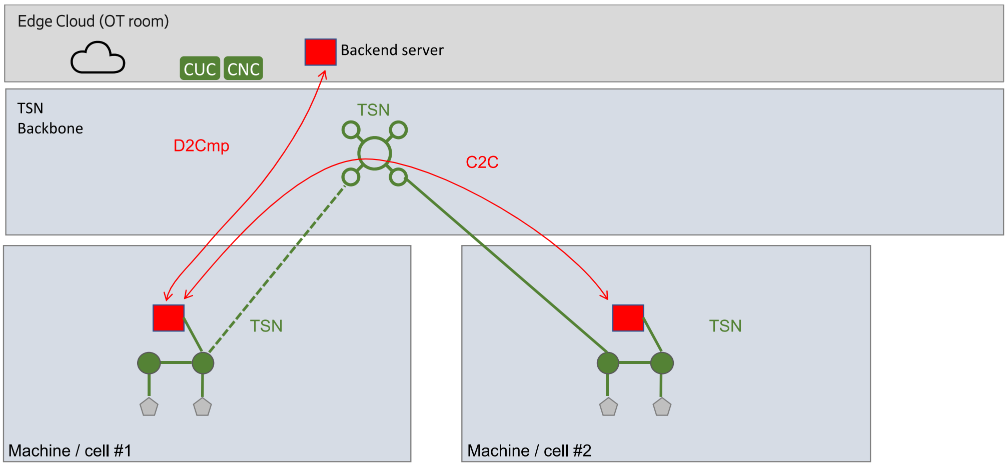


Figure 26 Deployment for the use case on industrial LAN & controller-to-controller communication (use cases 7 in [5GS20-D11])

5.3.2 Configuration of the 5GS as a TSN Bridge

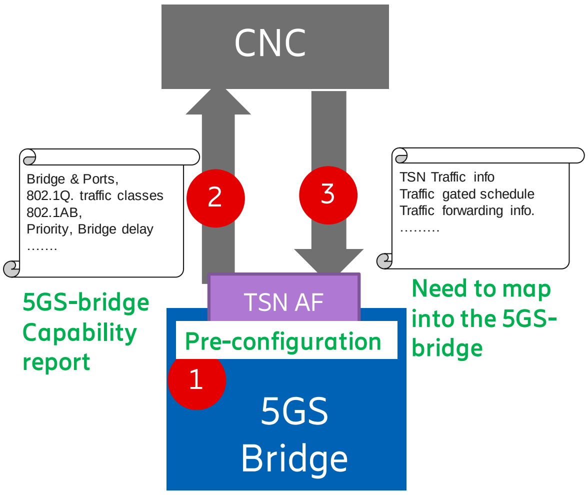


Figure 27 5GS configuration by CNC

Figure 27 shows a high-level view of the configuration by CNC of the 5GS as a TSN bridge. Note that the interactions between the CNC and the 5GS go through the TSN AF which maps between the TSN parameters and the 5GS parameters. The interaction between the TSN AF and the CNC is according to e.g. IEEE 802.1Qcc and is identical to the interactions that a CNC has to other TSN bridges.

The configuration proceeds in 3 phases:

1. Pre-configuration of bridge information

   This phase deals mainly with configuration of the 5GS bridge with static information such as bridge ID, and port numbers in the network side TSN translator (NW-TT). The TSN AF shall be pre-configured with a QoS mapping table. The mapping table contains TSN traffic classes and their relation to the preconfigured 5G QoS profiles. The 5GS bridge delay is configured and can be updated in phase 2 (e.g. with the residence time within a UE and DS-TT).

2. 5GS bridge capability report to the CNC

   Via the TSN-AF a 5GS bridge capability report is reported to CNC. As the 5GS can appear as multiple bridges (one per UPF) multiple of these reports are provided. This comprises information about the bridge name, number of ports, port numbers & addresses. Other information exposed is the topology information of the 5GS bridge based on IEEE 802.1AB LLDP. Furthermore, the TSN AF exposes its TSN capabilities, like the support for 802.1Qbv or 802.1Qci, in case that this is supported by all of the ports. The TSN AF also reports on the minimum and maximum delays between port pairs, including the residence time within the UE and DS-TT.

3. CNC bridge configuration commands to NW-TT

   The CNC provides a bridge configuration to the 5GS through the TSN AF, which contains, e.g., scheduling information as specified in IEEE 802.1Qbv, PSFP information as specified in IEEE 802.1Qci, or traffic forwarding information. The TSN AF maps the configuration information obtained from TSN network into 5GS QoS information (e.g. 5QI) of a QoS flow in a corresponding PDU Session for efficient time-aware scheduling. For the efficient transmission of the Ethernet frames over the 5G radio interface, it is helpful to obtain insights into the traffic pattern of the TSN streams; this is defined in 3GPP as TSCAI, which contains a traffic flow direction, a periodicity and a burst arrival time. The TSCAI is derived in the TSN AF and is forwarded to the 5G gNB where it may be used by e.g. the radio scheduler to optimize radio resource allocations. The information desired for TSCAI is not explicitly received by the CNC. However, if 802.1Qci PSFP is configured by the CNC, it contains a characterization of a TSN traffic stream. From this information the TSCAI parameters can be derived by re-engineering. It may not be possible to determine TSCAI, e.g. if PSFP is not configured by the CNC or if the information received is insufficient. In [FBG20] a configuration enhancement for TSN bridges has been brought to the IEEE 802.1 TSN standardization in the ongoing standardization project P802.1Qdj, which allows a TSN bridge to subscribe to traffic flow information from the CNC, so that a 5GS can determine TSCAI independent from a configuration of PSFP.

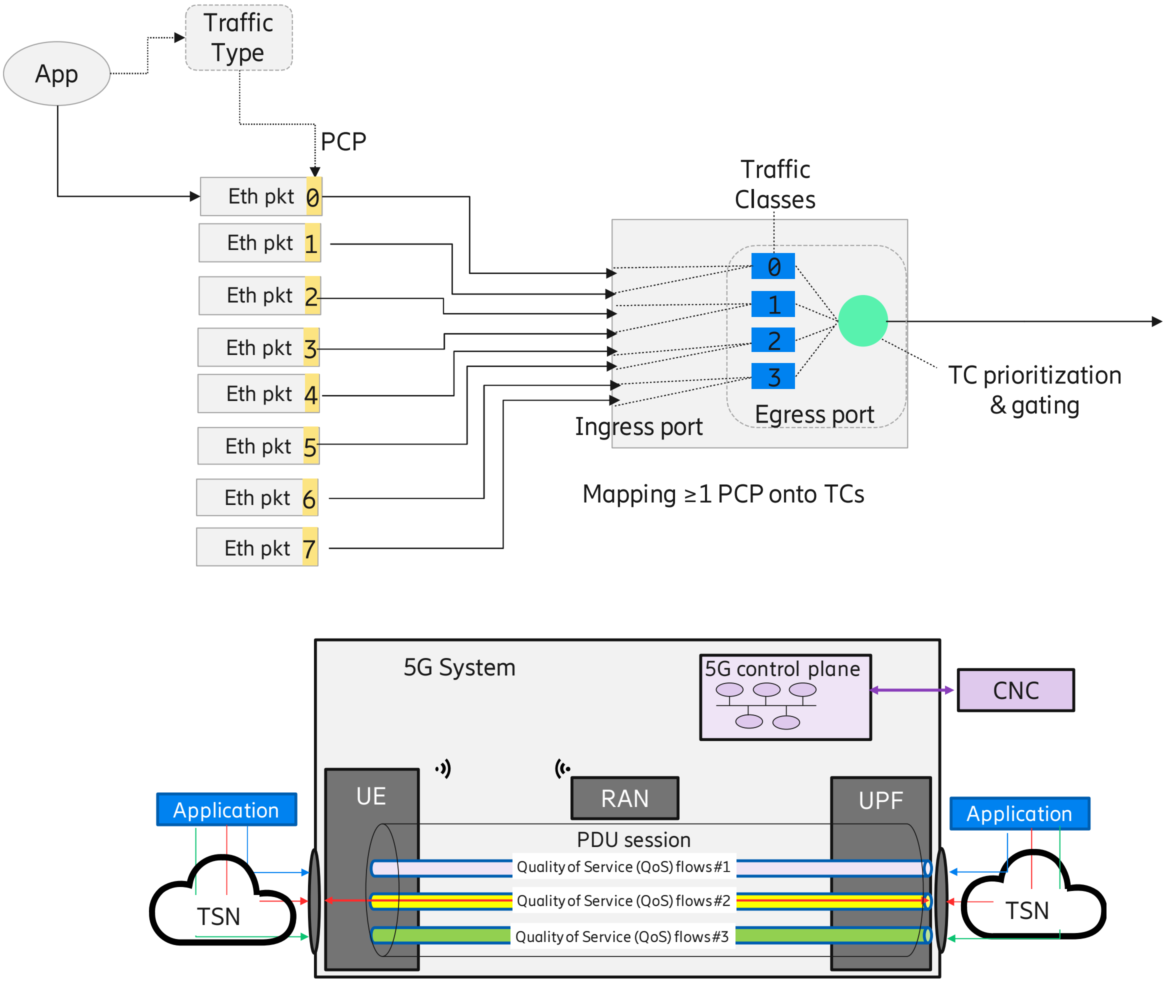
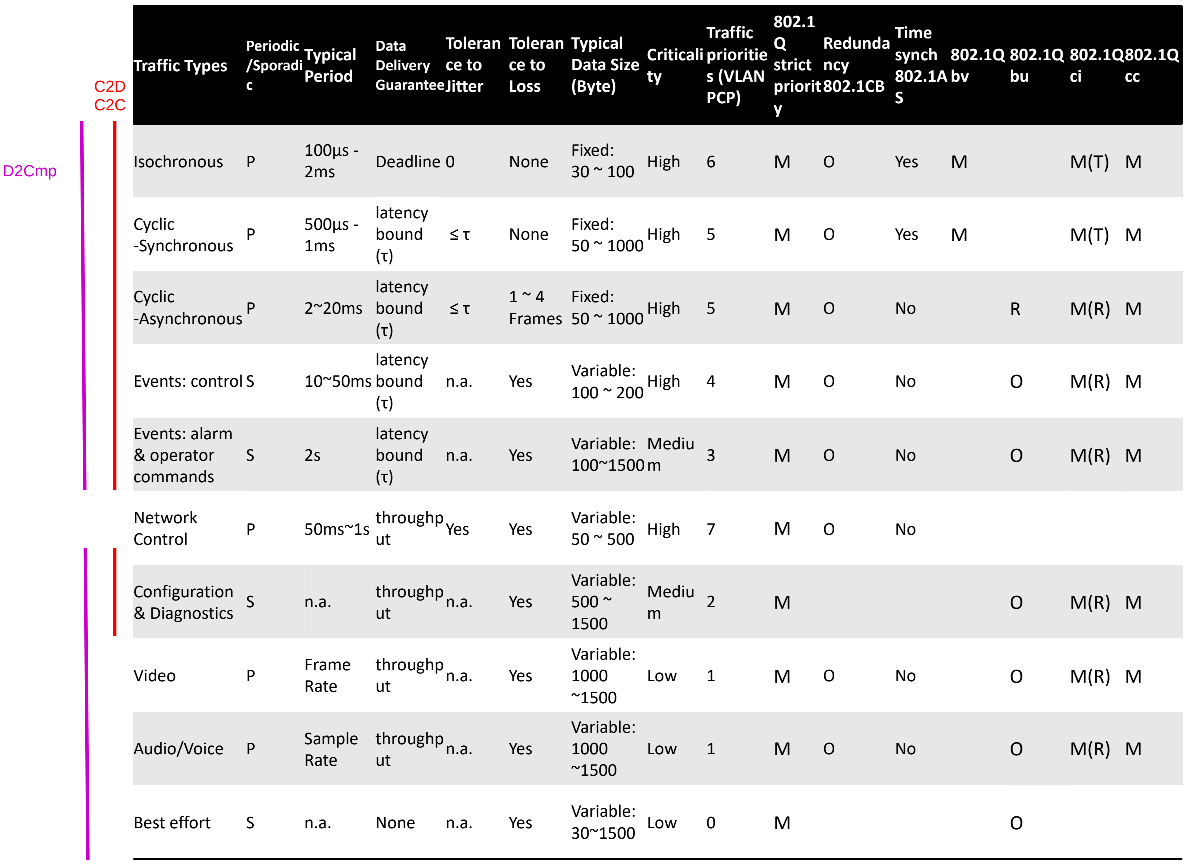


Figure 28 Traffic handling in Ethernet/TSN (above) and 5G (below)

In order to support QoS for Ethernet and TSN traffic, the traffic flows have to be mapped appropriately to the 5G QoS flows. Figure 28 shows the traffic handling for Ethernet / TSN and 5G. Ethernet traffic is tagged with a priority code point (PCP) within the 802.1Q VLAN header and is mapped in a bridge according to the PCP to a specific traffic class. The TSN configuration for the bridge specifies the traffic handling for those traffic classes, like providing 802.1Q priorities to different traffic classes or configuring the gates for traffic classes and their cycle times according to 802.1Qbv time scheduling. Various traffic types that are relevant for industrial automation are e.g. discussed in [3GPP20-22104] or [IEC/IEEE P60802]. Figure 29 shows examples of traffic types on the left and possible TSN configurations on the right.


C2D
C2C
D2Cmp

Traffic Types	Periodic /Sporadic	Typical Period	Data Delivery Guarantee	Tolerance to Jitter	Tolerance to Loss	Typical Data Size (Byte)	Criticality	Traffic priorities (VLAN PCP)	802.1Q strict priority	Redundancy 802.1CB	Time synch 802.1AS	802.1Q bv	802.1Q bu	802.1Q ci	802.1Q cc
Isochronous	P	100µs - 2ms	Deadline	0	None	Fixed: 30 ~ 100	High	6	M	O	Yes	M		M(T)	M
Cyclic -Synchronous	P	500µs - 1ms	latency bound (τ)	≤ τ	None	Fixed: 50 ~ 1000	High	5	M	O	Yes	M		M(T)	M
Cyclic -Asynchronous	P	2~20ms	latency bound (τ)	≤ τ	1 ~ 4 Frames	Fixed: 50 ~ 1000	High	5	M	O	No		R	M(R)	M
Events: control	S	10~50ms	latency bound (τ)	n.a.	Yes	Variable: 100 ~ 200	High	4	M	O	No		O	M(R)	M
Events: alarm & operator commands	S	2s	latency bound (τ)	n.a.	Yes	Variable: 100~1500	Medium	3	M	O	No		O	M(R)	M
Network Control	P	50ms~1s	throughput	Yes	Yes	Variable: 50 ~ 500	High	7	M	O	No				
Configuration & Diagnostics	S	n.a.	throughput	n.a.	Yes	Variable: 500 ~ 1500	Medium	2	M				O	M(R)	M
Video	P	Frame Rate	throughput	n.a.	Yes	Variable: 1000 ~1500	Low	1	M	O	No		O	M(R)	M
Audio/Voice	P	Sample Rate	throughput	n.a.	Yes	Variable: 1000 ~1500	Low	1	M	O	No		O	M(R)	M
Best effort	S	n.a.	None	n.a.	Yes	Variable: 30~1500	Low	0	M				O		

Figure 29 Examples of traffic types for TSN in industrial automation for controller to controller (C2C) Controller

Within the 5GS, different QoS flows are configured to provide a specific QoS, and traffic is then mapped to a QoS flow with the required QoS capabilities, as shown in Figure 28. The CNC configures for traffic handling in the 5GS bridge for the different Ethernet/TSN traffic classes according to the capabilities that have previously been reported by the 5GS bridge (e.g., data rate and latency). The 5GS has then to map the Ethernet/TSC traffic classes to the corresponding QoS flows. The PCF performs session binding using the UE Medium Access Control (MAC) address characterizing the DS-TT. The PCP value gets mapped to an appropriate *priority level* of a 5QI and the *packet delay budget* is used to calculate the 5GS bridge delay.

Gap analysis

The Ethernet traffic comprises a mix of different service categories, mixing e.g. TSN traffic streams with non-TSN Ethernet traffic. The 5GS needs to support the full range Ethernet traffic. In Release 16 only static configurations of Ethernet forwarding rules at the UPF are specified, which requires that all Ethernet traffic needs to be pre-configured in the 5GS or it will be discarded in the 5GS. A generalized approach for establishing forwarding rules is needed equivalent to the handling of forwarding rules in an TSN switch. This comprises central configuration of forwarding rules for e.g. TSN traffic streams via the CNC, as well as dynamic traffic-driven establishment of forwarding rules according to the MAC learning applied in Ethernet bridges.

5.4 5G integration with TSN for different NPN deployment options

5.4.1 Standalone NPN (SNPN)

The 5GS may be deployed using the SNPN option. Note that in a typical deployment there would be several UPFs acting as several 5GS virtual bridges in a factory as shown in Figure 30. There are also potentially several session management functions (SMFs) as the choice of an SMF depends on the DNN and there is likely to be several DNNs in an industrial environment.

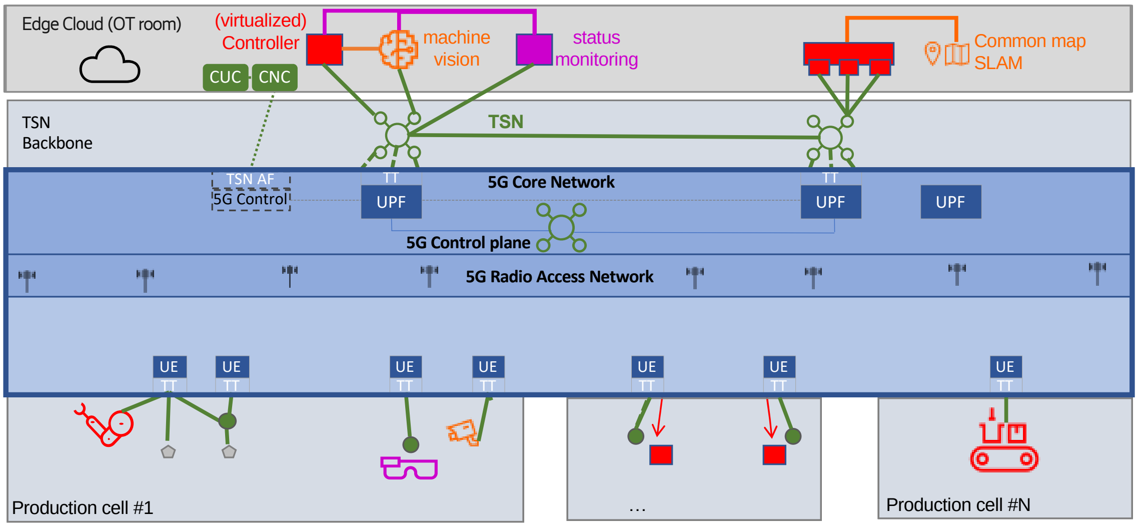


Figure 30 SNPN deployment overview

As for deployment choices for these different NFs, UPFs may be located on the RAN site and the SMFs in the central room. The other NFs of the 5G core may also be deployed in the central room (not shown in Figure 30).

The schema shown with the dedicated UPFs and SMFs may be viewed as a URLLC network slice deployed to satisfy the stringent requirements of TSN type applications. Other types of services may be deployed on the same network using a one or more different slice.

5.4.2 SNPN with RAN sharing

The lowest level of sharing of network components between two networks is 5G MOCN/MORAN where only the RAN is shared. A PLMN and a SNPN can share the NG-RAN. Adding RAN sharing to an SNPN does not change the analysis done in the previous section 5.4.1 for SNPN without RAN sharing. Within 5G-SMART, [5GS20-D14] provides details on RAN sharing and spectrum issues. Coexistence with PLMN networks in manufacturing environments involve cooperation/coordination for example synchronized TDD [5GS20-14]. The use of SNPN with shared RAN (NPN2) can facilitate such cooperation.

5.4.3 PNI-NPN: Dedicated user plane and shared control plane

A PNI-NPN with shared control plane allows the NPN to have dedicated user plane with UPFs, keeping all user plane traffic on premises, while some of the control plane functions are shared with the public network. Different options exist on what control functions are provided on-premises and which are located in the public network domain as indicated in Figure 31. The TSN AF corresponds to the controlling entity of the 5GS bridge(s). For a normal TSN bridge it is a function embedded into the bridge, and it has tight interactions with non-5G part of the industrial network including the CNC. The policy control function (PCF) and the SMF translate the TSN specific configuration of the traffic handling into corresponding 5G traffic handling procedures utilizing the 5G QoS capabilities. One possible solution is to have dedicated control-plane functions related to the TSN traffic handling for the NPN on premises, comprising the TSN AF, the PCF and the SMF. These can be associated to the NPN e.g. via a network slice that is provided to the NPN. Other control-plane functions like the access and mobility management function (AMF) may be shared between the NPN and the public networks. A further analysis of shared control plane functions in PNI-NPN and the impact on 5GS to interwork with the TSN will be addressed in a D5.4.

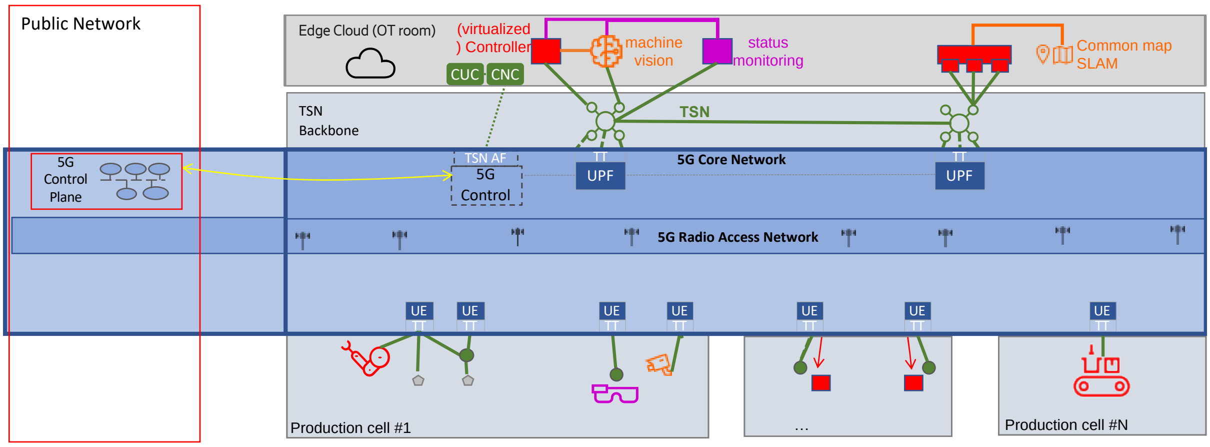


Figure 31 Shared control plane NPN option for the TSN case

5.4.4 NPN hosted by the PN (NPN 4)

This is the NPN option with the highest level of sharing where the 5G system belongs to the PLMN operator (MNO domain). It should be noted the 5GS acting as a TSN bridge interacts with the local LAN – bridges and devices. Hence a dedicated slice, is needed for TSN services. However, in manufacturing scenarios there exist also use cases that are not time critical such as configuration and software/firmware upgrades. A separate slice for these less critical use cases can make use of NPN4 type sharing.

5.5 Time Synchronization & Positioning

The section provides information on the time synchronization and positioning in context of the deployment models. With regard to time synchronization, there are two main fundamental steps from network architecture perspective that must be taken care of:

1. The first step is to make sure that the required time error limits are met within the 5G system.

In this regard, the time synchronization budget allocated to the 5G system was a major open question in TS 22.104 [3GPP20-22104]. After several discussions, 3GPP has concluded during the August 2020 SA1 plenary meeting that the time synchronization budget for the 5G system should be ≤900 ns, where the total end-to-end time error budget is 1 µs between the TSN GM and the TSN output of the end station, with 100 ns of time error budget to the TSN domain outside the 5GS separately. ITU-T Recommendation G.8271.1 has developed Hypothetical Reference Models (HRMs) which help network operators to design their synchronization networks by appropriately allocating time error budgets to different segments of the network.

In the current context of integrating TSN with a 5G system, as shown in Figure 32 below, the first step is to make sure that the time error budget of 900 ns is respected between the NW-TT input and the UE/DS-TT output in order to synchronize the TSN end-stations which is behind the DS-TT, keeping the total end-to-end budget between the TSN GM and the input at the end station within 1 µs. When the TSN GM is provided by the 5G system, this budget could be allocated between the primary reference time clock (PRTC) within the 5G system and the NW-TT in one side and the DS-TT on the other side

By carefully considering the various ITU-T clocks (such as Telecom Boundary clocks i.e. T-BC Class A, Class B, Class C or even Class D) and Primary Reference Time Clocks (i.e. PRTC, enhanced PRTC, etc.) which could transport and distribute timing packets within the 5G system, one can estimate the time error budget that could be achieved within the 5G system. Since time synchronization is an essential part of TSN, the 5G network can and should be deployed in such a way that the number of hops between the Grand Master (GM) clock and the NW-TT are minimal. The fewer the number of nodes, the lower is the time error cumulated within the 5G system and therefore the better is the time synchronization accuracy. In addition, the solutions for transporting timing packets from the gNB to the UE Over-The-Air techniques is currently under study in the scope of this project, including the appropriate budget to be allocated to the air interface, as well as to the end clock in the UE/DS-TT. The budget allocated to the air interface depends on various factors. In particular, the delay over the radio interface shall be compensated if this is not negligible (e.g., 10m corresponds to 30 ns).

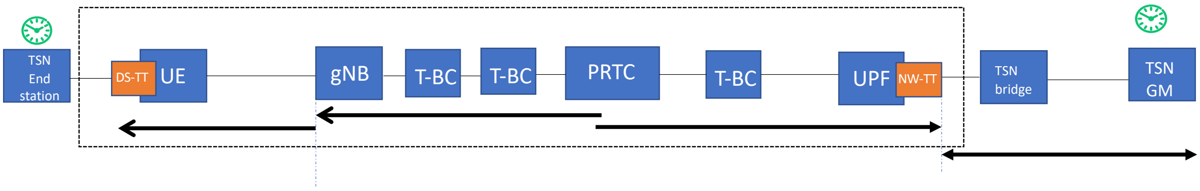


Figure 32 Time error budget in 5G system

2. The second step is to transport and distribute the timing reference from the 5G system through the clocks in the industrial networks

As described from a TSN point of view, the whole 5G system is seen as a single virtual bridge (or time-aware system), allowing to reach virtually any part of the shop floor with a very limited number of hops (as seen by the TSN network). The 5G time synchronization architecture is shown in Figure 33.

To satisfy the 1 microsecond of the time error application requirement which will include 5G system and TSN components, here according to SA 1 requirement 5GS is bound to have below 900ns of time error. A properly engineered solution integrating a 5G system into a TSN network can and should be designed so that the synchronization messages exchanged between a time source clock and an end device/application/end-station meets the time error application requirements as per ITU-T G.8271.1 and 3GPP SA 1 TS 22.104, (i.e. at most four hops outside of the 5G system might be considered, when the 5G system is in the path).

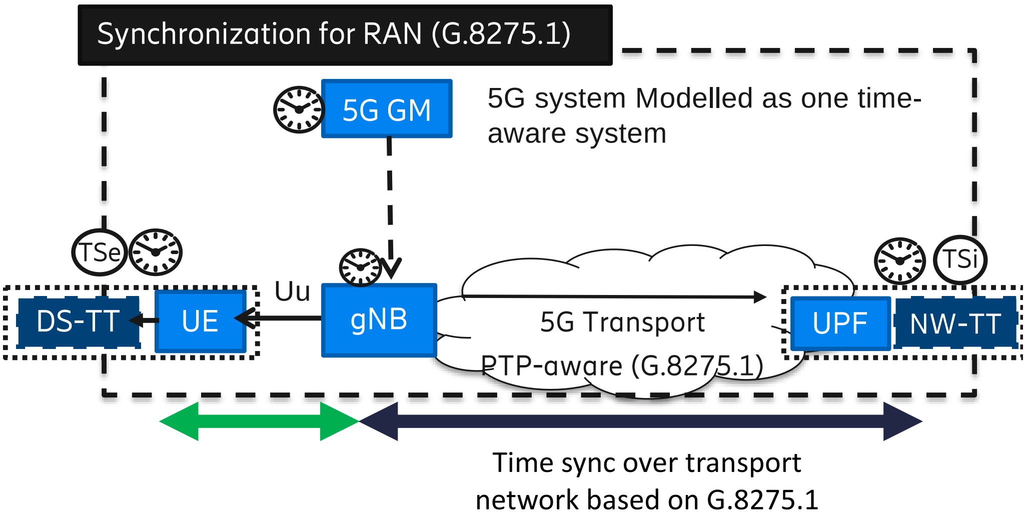

Synchronization for RAN (G.8275.1)
5G GM
5G system Modelled as one time-aware system
TSe
TSi
Uu
DS-TT
UE
gNB
5G Transport
PTP-aware (G.8275.1)
UPF
NW-TT
Time sync over transport network based on G.8275.1

Figure 33 5G time synchronization architecture

An early investigation for the different time synchronization architecture required for different NPN options provides a recommendation for different NPN deployment options as listed in Table 6.

NPN No.	Deployment options	Synchronization architectures of 5G system
NPN 1	Standalone network	Requires dedicated NPN synchronization architecture
NPN 2	NPN shared RAN with PN	Requires a synchronization architecture shared between NPN and PN. Dedicated synchronization clocks can be defined for NPN
NPN 3	NPN with Shared control plane and RAN	Requires a synchronization architecture shared between NPN and PN. Dedicated Synchronization clocks can be defined for NPN
NPN 4	NPN hosted by PN	Requires a synchronization architecture shared between NPN and PN

Table 6 Time synchronization w.r.t to NPN deployment options

Positioning

5G positioning has been described in detail in Deliverable 5.1 [5GS20-D51]. From the conclusions of D5.1 for indoor positioning, considering deployment model we can conclude:

- Need for adequate infrastructure deployment
- Connectivity requires good signals to and from one base station, but localization needs the involvement of at least 3 base stations in order to solve the position (x, y) and local time.

    - In particular, depending on the application, vertical 3D positioning may be required, necessitating the involvement of at least 4 base stations, placed in a suitably good vertical configuration in the building.
    - For all cellular positioning techniques except cell-ID, there must be line-of-sight paths to the base stations utilized for positioning measurements. This contrasts with connectivity where signals that have bounced off machinery or objects in the factory hall are also useful.
- Need for knowledge about the locations of the base stations (more specifically, of the base station antennas)
- Need for accurate synchronization of the base stations which is only applicable for Observed Time Different of Arrival (OTDOA) based methods.

The need for at least 3 base stations (more precisely the antenna of the BS) with line-of-sight to the UE is really a constraining factor for deployment. The placement of the BS is also critical as they need to provide good trilateration for all the concerned UEs. Only 5G-SMART use case 5 [5GS20-D11] lists positioning as a requirement and needs to take the above constraints into account.

5.6 Security zone and conduits

This section describes how the 5G VLAN configuration can be adopted for realizing security zones in smart manufacturing. In 5G communication security mechanisms are based on the global 3GPP specifications which are widely adopted and practiced in current 5G network deployments. For industrial wired network where wide range of the communication technology are observed today. The international standardization IEC has specified a security framework covering both organizational and technical aspect of security in IEC 62443. IEC 62443 standards define engineering measures that will guide an organization through the process of assessing the risk of particular Industrial Automation and Control System (IACS), also identifying and applying security countermeasures to reduces that risk to tolerable levels.

IEC 62443 provides a framework of functional and procedural requirements to address the issue of security for IACS. According to IEC 62443 both L2 (802.1X) and L3/L4 (cryptographic protocols for transport, TLS, DTLS) might be required for device authentication/authorization when connecting to fixed LAN infrastructure and application layer security. These standards also require user authentication and authorization, including roles, key and certificate management, access control mechanisms, and auditing and logging is required (e.g. Directory Services LDAP or similar, PKI and X.509 certificates). The standards define seven foundational requirements (FR) that must be met for securing each component in the system:

1. FR1: Identification and authentication control.
2. FR2: Use control.
3. FR3: System integrity.
4. FR4: Data confidentiality.
5. FR5: Restricted data flow.
6. FR6: Timely response to event.
7. FR7: Resource availability.

In addition to these seven requirements, different security levels are set by the IEC 62443 standard to ensure the security measures/techniques are well in place for the overall system.

5.6.1 Security zones and conduits

A typical smart manufacturing shopfloor consists of a large complex system and not every component of such a complex system will require the same level of security. Such differences are addressed by applying the concept of the security zone. According to ISA/IEC ISA2443, each security zone defines a logical grouping of physical, information or application assets sharing a common security requirement [SEC20-5GACIA]. Security conduits according to IEC 62443 is defined as “Conduits are the special type of security zone that groups communications that can be logically organized into a grouping of information flows within and also external to a zone”. Figure 34 shows the example of the security zone and conduits from the ISA-62443-1-1 draft. It shows 4 security zones (Enterprise network, industrial/Enterprise Demilitarized Zone DMZ, industrial networks). Conduits that interconnect all four zones are shown. DMZ are usually utilised to isolate the network traffic between IT and OT side.

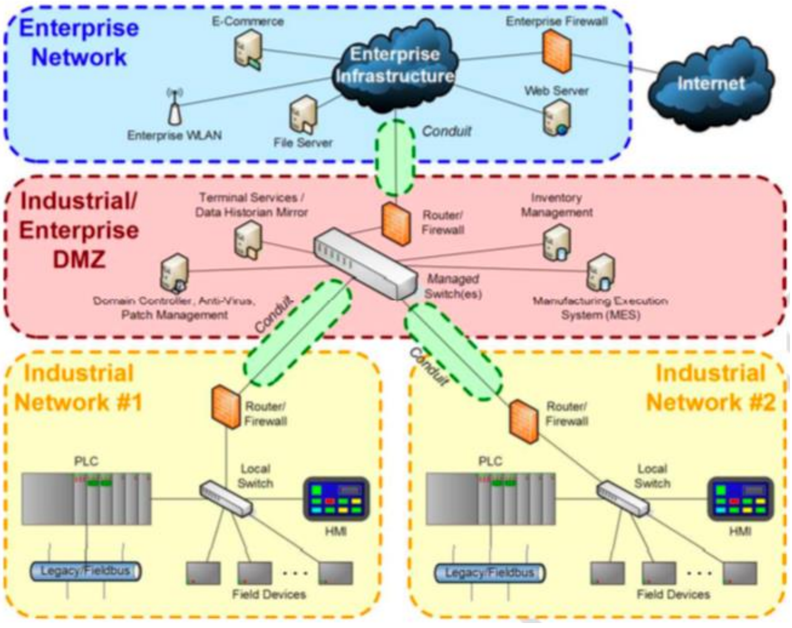


Figure 34 High level manufacturing example for zones and conduits [ISA 62443]

Figure 35 shows a high-level manufacturing example of security zones over an Ethernet network. It has three zones defined green, blue and green. A common practice for zoning in a factory is based on the Ethernet IEEE 802.1Q VLAN mechanism. VLANs are a mechanism to separate traffic and specify which end hosts receive the information of a certain VLAN. VLANs utilize tags in the Ethernet packet header to determine how they are handled by layer 3 device. Here bridges are transport nodes, VLAN traffic is switched, and end host should be connected to a single VLAN, i.e. be assigned as a member of a specific VLAN group, only based on the security zone it belongs to.

Security zones based on VLANs today are largely static with infrequent reconfigurations. With the transition towards increasingly flexible productions systems also network configurations become more dynamic. This necessitates a shift from e.g. command-based configuration towards more automated network configurations and an software defined networking (SDN)-based central control is foreseen (as shown Figure 35 in NMS) as a flexible tool to configure VLANs throughout the network.

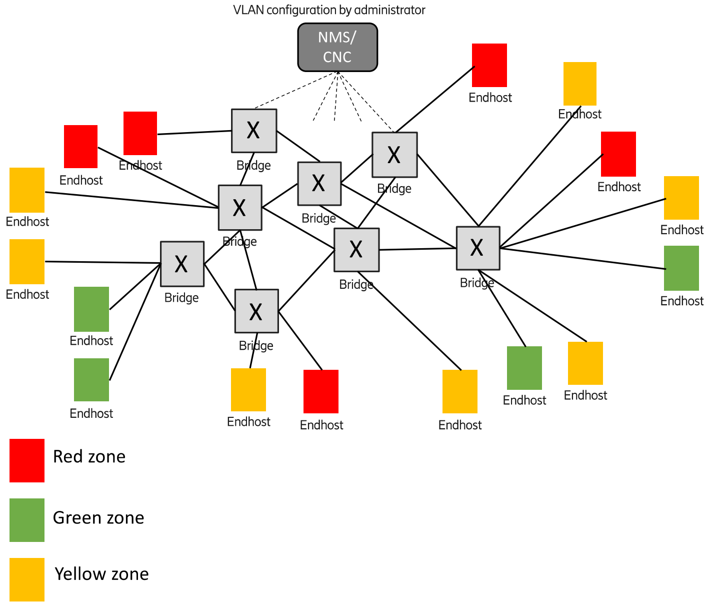


Figure 35 Security zone over industrial wired network

5G VLAN to support VLANs based security zones

In an 5G integrated industrial network, the 5G system is seen as a set of bridges identical to the other fixed bridges. The 5G bridge as described in the section 3 can be configured for the support of different VLANs in the same manner as a fixed bridge; no extra functionality is required for adopting the bridge/VLAN management for the 5GS.

As shown in the Figure 36, end hosts can be directly connected to the 5G system bridge either via a DS-TT port, which provides connectivity at the UE side, or via a NW-TT port. Also, further bridges can be connected to the 5GS on either UE or UPF side. The configuration of the VLANs similarly like any other fixed bridge is done by NMS or Centralized Network Controller (CNC) in case of the TSN.

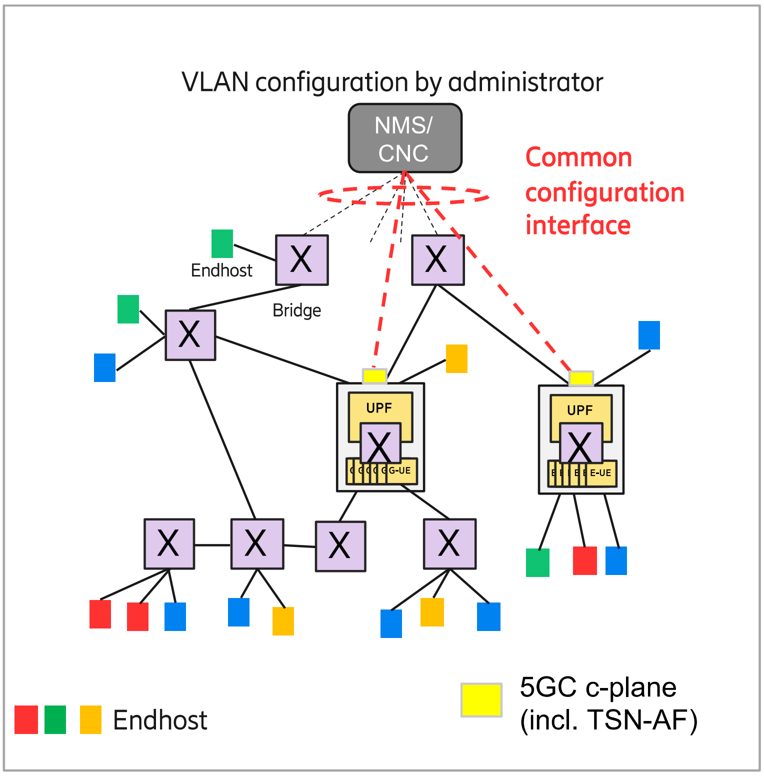

VLAN configuration by administrator
NMS/ CNC
Common configuration interface
Endhost
Bridge
X
UPF
G-UE
UPF
E-UE
🟥 🟩 🟨 Endhost
5GC c-plane (incl. TSN-AF)

Figure 36 5G VLAN to support security zone

Logical segmentation using virtual local area networks (VLANs) is another technique that could be used in conjunction with or in place of network slicing to implement security zones and conduits over the 5G network. Thus, each zone can be treated as a separate network slice.

Securing and isolating the slices will be a fundamental requirement. A network slice provides an isolated, end-to-end network, optimized for a specific business purpose. Specific security features necessary for services and applications can be built into the network slice architecture. When the UE is authenticated by the network upon initial access, a set of network slice selection assistance information (NSSAI) items denoting what is permissible is sent back to the UE. A request to access a network slice that is not part of the NSSAI item set is denied.

Authentication and authorization for slice access is incorporated into primary authentication. In addition to the primary authentication to the 5G network, a slice-specific authentication procedure using the Extensible authentication protocol (EAP) authentication framework (RFC 3748) can be used. UE receives a list of permitted network slices, indicated by their NSSAIs. The permitted network slices may further require slice-specific authentication by the slice Authentication, Authorization and Accounting (AAA) function.

6 Conclusion, summary and future work

Smart manufacturing has different functional requirements compared to traditional cellular applications. These functional requirements include security, time synchronization, layer 2 switching/TSN integration, exposure capabilities, interface towards operation & management, network slicing. This deliverable provides an analysis of the 5G network architecture from an operational and deployment point of view. A range of deployment options and its detailed characteristics are investigated in the document. Furthermore, actors and roles are defined to explain the operational view of the 5G system in the manufacturing ecosystem.

An overall deployment validation of the 5G technical enablers with respect to 5G-SMART use cases and their functional requirements is performed. A gap analysis is made between the 5G standard Release 16 and the desired functionality for interworking of 5G with Ethernet, TSN integration and native 5G connectivity use cases. Table 6 below summarized the technical enablers to address functional requirements and potential solutions and/or gaps are detailed below.

Functional requirements	**Technical enablers**
Industrial LAN support via Layer 2 switching/TSN integration	5G support for Ethernet Connectivity 5G support for TSN
Security Zoning	Ethernet LAN configuration service and network slicing
Time synchronization	5G support time synchronization [D5.1]

Table 6 Mapping of functional requirements to technical enablers

The following aspects are identified as potential enhancement to current standards, and these are under discussion in 3GPP SA2 Release 17

- 5G VN group solution have limitations including no interworking with fixed Ethernet domain defined, no support for forwarding a broadcast/multicast packets with sources not known to SMF and UPF.
- The manner in which Ethernet user plane forwarding is handled is not fully specified in the current standards.
- Instead of utilising static configuration rules for Ethernet forwarding, a generalized approach for establishing forwarding rules at the UPF is needed which is similar to the handling of forwarding rules in an TSN switch.

TSN integration is feasible for most of the 5G-SMART use cases when NPN1 or NPN2 (SNPN) deployment options are utilised. For NPN3 (shared control plane) more investigation is needed on how control plane components are realized and integrated with TSN system to ensure secure and reliable functioning of the entire communication system. Coexistence with PLMN networks in manufacturing environments involve cooperation/coordination for example synchronized TDD. The use of SNPN with shared RAN (NPN2) facilitates such cooperation.

5G-SMART use cases 1-6 can be realized without using TSN and using only native 5G connectivity. While low latency and high reliability (URLLC) type use cases are very important for smart manufacturing communication services, there is also a need for support for high bandwidth (eMBB) for camera/video type use cases and also support for a large number of sensors/actuators. Hence a number of slices with varying QoS needs will likely be deployed and used. Use of higher levels of sharing (shared control plane and NPN4) offer the additional resources needed to support a wide range of communication services (example URLLC and eMBB) in order to offer a complete smart manufacturing solution. In future work the 5G-SMART project will consider other aspects such as security and other functional requirements. Furthermore, impact on time synchronization and positioning is explored. An investigation on how 5G can support realization of the security zones is performed. 5G VLAN configuration technical feature is seen as suitable option to enable security zoning.

Concerning edge clouds for smart manufacturing, service models for edge clouds that currently seem to best fit with manufacturing use cases are container and FaaS abstractions. Edge clouds can operate separately and independently or can interoperate with central cloud components. Edge cloud also plays an important role for ensuring E2E low latency, the report described main factors (infrastructure and resource management options) to ensure low latency and provides recommendations.

For future work, we plan to extend the deployment and operational model study by evaluating different operation models. A detailed analysis of the interrelations between operation and deployment models is the topic of ongoing work.

Reliability mechanisms for different NPN deployment options will be investigated considering functional and performance requirements of the use cases.

Also, the impact on network deployment options (in particular NPN3) with TSN integration will be further investigated. Integration of edge cloud in the context of different NPN options will be explored based on the 5G-SMART use cases.

7 References

[3GPP18-22821] 3GPP technical report TR 22.821, “Feasibility Study on LAN Support in 5G”, June, 2018

[3GPP20-22104] 3GPP technical specification TS 22.104, “Service requirements for cyber-physical control applications in vertical domains”, October, 2020

[3GPP20-22261] 3GPP technical specification TS 22.261, “Service requirements for the 5G system”, March 2020

[3GPP20-23700] 3GPP technical report TR 23.700-20, “Study on enhanced support for industrial Internet of things in the 5G system (5GS)”, November, 2020

[3GPP20-23434] 3GPP technical specification TS 23.434, “Service Enabler Architecture Layer for Verticals (SEAL); Functional architecture and information flows” Service Enabler Architecture Layer for Verticals (SEAL), September, 2020

[3GPP20-TS23501] 3GPP technical specification TS 23.501, “System architecture for the 5G System (5GS)”, Release 16, August, 2020

[5GS20-CT] Berna Sayrac, 5G-SMART report , “5G common terminology report”, 2020 https://5gsmart.eu/deliverables/

[5GS20-D11] Ognjen Dobrijevic, et. al 5G-SMART deliverable D1.1, “Forward looking smart manufacturing use cases, requirements and KPIs”, June 2020. https://5gsmart.eu/deliverables/

[5GS20-D14] Kimmo Hiltunen, 5G-SMART deliverable D1.4, “Radio network deployment options for smart manufacturing”, December 2020. https://5gsmart.eu/deliverables/

[5GS20-D21] Ognjen Dobrijevic, 5G-SMART deliverable D2.1, “Design of 5G-based testbed for industrial robotics”, May 2020. https://5gsmart.eu/deliverables/

[5GS20-D32] Dirk Lange, Niels König, 5G-SMART deliverable D3.2, “Report of system design options for monitoring of workpiece and machine”, May, 2020. https://5gsmart.eu/deliverables/

[5GS20-D51] Dhruvin Patel, et. al 5G-SMART deliverable D5.1, “First report on new technological features to be supported by 5G standardization and their implementation impact”, May, 2020. https://5gsmart.eu/deliverables/

[AR5GF19] Ahmad Rostami, "Private 5G Networks for Vertical Industries: Deployment and Operation Models," 2019 IEEE 2nd 5G World Forum (5GWF), Germany, 2019.

[E2E19-5GNGMN] Dhruvin Patel, Joachim Sachs, NGMN, 5G E2E technology to support verticals URLLC requirement, October 2019.

[FBM20] János Farkas, Balázs Varga, György Miklós, ”Configuration Enhancements for 5G as TSN Bridge,” contribution to IEEE 802.1 TSN TG interim meeting, September 2020, https://www.ieee802.org/1/files/public/docs2020/dj-farkas-configuration-enhancements-for-5G-0920-v01.pdf

[IEC/IEEE P60802] IEC/IEEE 60802 TSN Profile for Industrial Automation, IEEE 802 and IEC SC65C/MT9,http://www.ieee802.org/1/files/private/60802-drafts/d1/60802-d1-1.pdf.

[IG+20] I. Godor *et al.*, "A Look Inside 5G Standards to Support Time Synchronization for Smart Manufacturing," in *IEEE Communications Standards Magazine*, vol. 4, no. 3, pp. 14-21, September 2020, doi: 10.1109/MCOMSTD.001.2000010.

[JBG+19] J. Farkas, B. Varga, G. Miklós , J. Sachs, 5G-TSN integration meets networking requirements, Ericsson Technology Review, August 2019.

[NPN19-5GACIA] 5G-ACIA, 5G Non-public Network for industrial scenarios, July, 2019, https://www.5g-acia.org/publications/

[SEC20-5GACIA] 5G-ACIA, Security aspect of 5G for Industrial Networks, May, 2020, https://www.5g-acia.org/publications/

Appendix

IEEE 802.1Q bridge forwarding process

The figure below is figure x from subclause 8.6 of IEEE Std 802.1Q-2018, which is the illustration of the general IEEE 802.1Q bridge forwarding process. The core part of the actual forwarding is in the frame filtering step. As the figure illustrates, the filtering database is an input to the whole process, where process, where the filtering database is set using a number of possible mechanisms, including MAC learning and CNC provided static filtering entries.

The logical view of the forwarding process assumes a single filtering database and a single unified forwarding process. The unity of the forwarding process is important as the filtering database may be populated by an entity that is external to the bridge (e.g. by the CNC provided static filtering entries). Besides, for the correct implementation of the flooding, the bridge needs to know when a destination address is known or unknown, for which the bridge also needs a single unified filtering database.

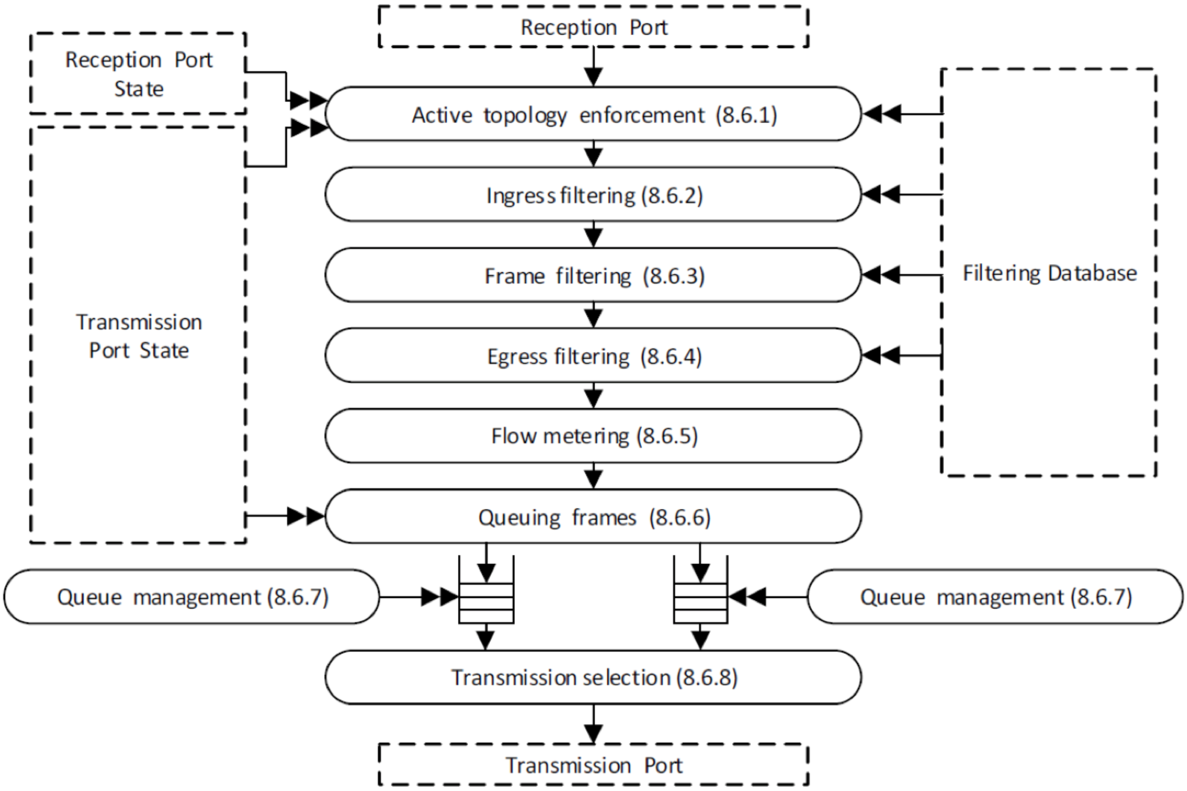


Figure 37 Bridge forwarding process based on IEEE 802.1 standard

Due to the fact that the forwarding process relies on a single process and a single database, we suggest that the 3GPP Ethernet forwarding model should also keep the forwarding into two parts, one based on Packet Detection Rules and another based on UPF internal mechanism.

It could be extremely difficult to run the Ethernet forwarding like that, one running in the UPF and one running based on SMF controlled rules[7], since the data for these two parts would reside at different entities (SMF vs. UPF). It could be extremely complex to harmonize the data and the processes for these two parts, whereas for correct IEEE 802 compliant Ethernet behaviour we need a single unified database and unified process.

The UPF implementations are of course always free to realize Ethernet frame forwarding based on the vendor's decisions which should not be restricted. The 3GPP standards should not limit the implementation by imposing special ways of splitting up the unified IEEE 802.1Q forwarding process. The use of the PDR/FAR rules as components in the forwarding are not excluded depending on how a UPF vendor realizes the bridge forwarding. But the standard must not force specific way for splitting up the bridge forwarding in the implementations.

Flooding as the default Ethernet forwarding mechanism

The default forwarding mechanism in Ethernet is flooding which make sure that frames are delivered to all other end stations in the networks. The figure below illustrates the flooding, whereby an end station sends a frame which is forwarded on all other active links of the bridges that constitute the spanning tree except where the frame was received. In this way, the flooding mechanism guarantees that the frame reaches its destination in the Ethernet network. Flooding applies to broadcast frames, and to unicast and multicast frames when the location of the destination is unknown, i.e., there is no entry in the Filtering Database corresponding to the destination MAC address. Flooding and the Ethernet bridging model is described in IEEE 802.1Q in detail. Flooding is also described in [3GPP20-TS23501] section 5.8.2.5.3 on a high level.

[7] I.e. packet detection rules (PDR) and forwarding action rules (FAR).

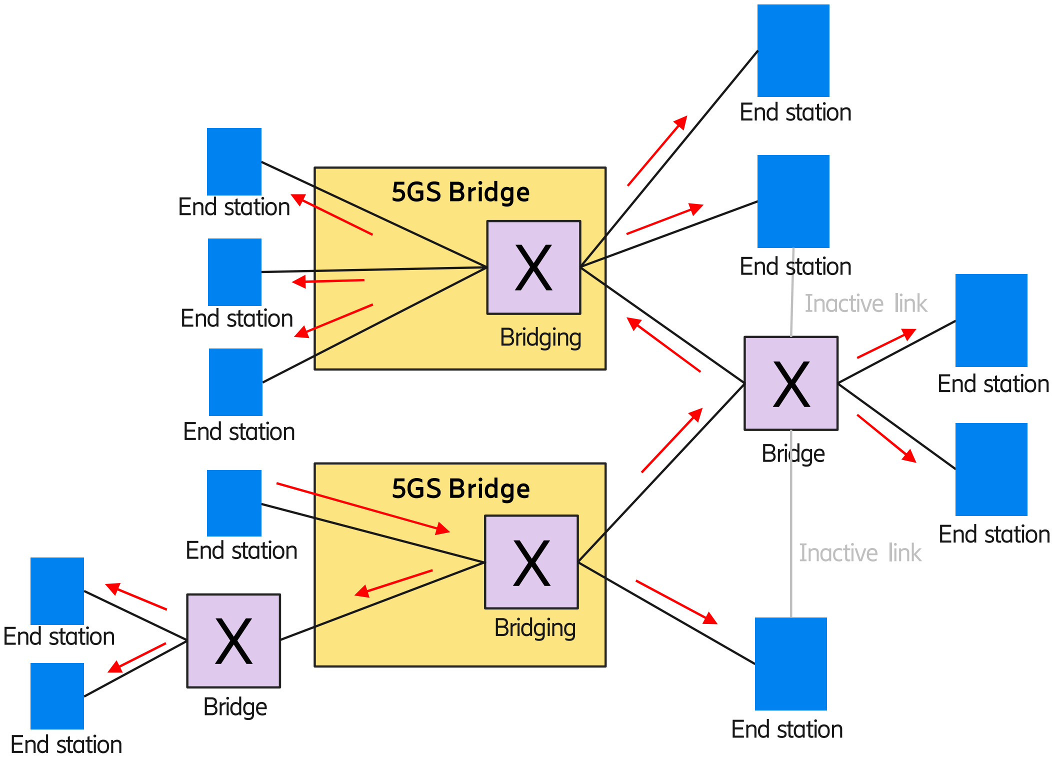


Figure 38 Flooding mechanism in integrated 5G network

Flooding is the default Ethernet behaviour which guarantees the simple plug&play nature of Ethernet networks and it is significantly different from how IP networks behave. The key differences between Ethernet and IP regarding the default behaviour are highlighted in the table below.

Ethernet	IP
Default: Flood	Default: Drop
Broadcast: flood Multicast: flood Unknown unicast: flood	Broadcast : (not really used) Multicast : drop (default) Unknown unicast: drop (default)
Send out frame on all active port except incoming	drop unknown traffic
Ethernet supports broadcast/multicast by default	Multicast traffic dropped by default
Ethernet is plug&play	IP require configuration

The flooding mechanism makes sure that all end stations are always reachable as Ethernet destinations, and in this way Ethernet networks are plug&play – no preconfiguration is needed for the network to operate. The MAC learning mechanism can help to limit flooding, i.e., for known unicast and multicast destination addresses it is possible to send out the frame only on the interface(s) where the destination end station is reachable. It is also possible to use central configuration to establish forwarding, i.e., to populate the Filtering Database. The very basic plug&play nature of Ethernet must be maintained, as it is commonly used e.g., to enable bootstrapping and to support management traffic.

List of abbreviations

3GPP	3rd Generation Partnership Project
5GS	5G System
5QI	5G Quality of Service Identifier
AAA	Authentication, Authorization and Accounting
AF	Application Function
API	Application Programmable Interface
BC	Boundary clock
C2C	Controller to Controller
C2D	Controller to Device
CaaS	Container as a service
CAG	Closed Access Group
CNC	Centralized Network Configuration
CNCF	Cloud Native Computing Foundation
CUC	Central User Configuration
D2Cmp	Device to Compute
DMZ	Demilitarized Zone
DNN	Data Network Names
DS-TT	Device Side TSN Translator
DTLS	Datagram Transport Layer Security
EAP	Extensible Authentication Protocol
E2E	End to End
eMBB	enhanced Mobile Broadband
FDB	Filtering database
FR	Foundational Requirements
GM	Grand Master
I-LAN	Industrial LAN
IaaS	Infrastructure as a Service
IACS	Industrial Automation and Control System
IEC	International Electrotechnical Committee
IoT	Internet of things
KPI	Key Performance Indicator
LAN	Local Area Network
LLDP	Link Layer Discovery Protocol
MAC	Medium Access Control
MIoT	Massive Internet of Things

mMTC	Massive machine type of communication
MNO	Mobile Network Operator
MOCN	Mobile Operator Core Network
MORAN	Multi-Operator RAN
NF	Network Function
NG-RAN	Next Generation Radio Access Network
NID	Network Identifier
NMS	Network Management System
NPN	Non-public network
NSSAI	Network Slice Selection Assistance Information
NW-TT	Network Side TSN Translator
O&M	Operation and Management
OPC-UA	Open process Control unified architecture
OT	Operational technology
OTDOA	Observed Time Difference Of Arrival
PaaS	Platform as a service
PCP	Priority code point
PDU	Packet Data Unit
PFSP	Per-Stream Filtering and Policing
PKI	Public key infrastructure
PLC	Programmable Logic Controller
PLMN	Public Land Mobile Network
PNI-NPN	Public network integrated non public network
PRTC	Primary reference time clock
QoS	Quality of Service
RAN	Radio Access Network
ROS	Robot operating system
SA	Standalone architecture
SaaS	Software as a service
SEAL	Service enabler architecture
SL	Security Level
SLA	Service Level Agreement
SMF	Session Management Function
SNPN	Standalone Non-Public Network
TLS	Transport Layer Security
TSC	Time Sensitive Communication
TSCAI	Time Sensitive Communication Assistance Information
TSN	Time Sensitive Networking
UE	User Equipment
UPF	User Plane Function

URLLC	Ultra-Reliable Low Latency Communication
VAL	Vertical Application Layer
VLAN	Virtual Local Area Network
VM	Virtual Machine
VN	Virtual Network
VPN	Virtual Private Network

Table 7: List of abbreviations