ngmn
the engine of wireless innovation

5G E2E TECHNOLOGY TO SUPPORT VERTICALS URLLC REQUIREMENTS

5G E2E Technology
to Support Verticals URLLC Requirements

by NGMN Alliance

Version:	**1.6**
Date:	**10-February-2020**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P - Public**
Authorised Recipients: (for CR documents only)	

Project:	**Verticals URLLC Requirements**
Editors / Submitters:	**Dhruvin Patel, Joachim Sachs**
Contributors:	András Zahemszky (Ericsson), Simone Ferlin (Ericsson), Ana Galindo-Serrano (Orange), Berna Sayrac (Orange), Bruno Tossou (Orange), Anthony Chan (HKT), Wang Rui (China Mobile), Karin Loidl (Fraunhofer IIS), Billy Liu (HKT), Lei Cao, Liang Xia, Rui Wang, Hang Cui, Lin Zhao (China Mobile)
Supporters:	
Approved by / Date:	**NGMN Board / 31st October 2019**

Address:
ngmn e. V.
Großer Hasenpfad 30 • 60598 Frankfurt • Germany
Phone +49 69/9 07 49 98-0 • Fax +49 69/9 07 49 98-41

Abstract:

The aim of this work is to illustrate how 5G end-to-end deployment can meet vertical requirements and also highlight what are the main 5G URLLC technology enablers for use cases identified in the report [1]. Some of these new use cases impose extremely low latency and high reliability requirements on 5G system from end-to-end perspective. This report also briefly describes 5G URLLC enablers along with reference architecture for specific use case and further evaluate how 5G URLLC can enable such use cases.

Document History

Date	Version	Author	Changes
26th March, 2019	V 0.1	Joachim Sachs, Dhruvin Patel (Ericsson)	Skeleton
2nd May, 2019	V 0.2	Dhruvin Patel, Joachim Sachs (Ericsson)	Table of content update
27th May, 2019	V 0.3	Dhruvin Patel, Joachim Sachs, András Zahemszky	Section on time synchronisation and smart manufacturing architecture
14th June, 2019	V 0.4	Ana Galindo-Serrano (Orange), András Zahemszky, Simone Ferlin (Ericsson)	Latency URLLC technical enabler, non public network.
27th June, 2019	V 0.5	Ana Galindo-Serrano (Orange), András Zahemszky (Ericsson)	Section 6.1, section 7.2
27th July, 2019	V 0.6	Wang Rui (China Mobile)	Section 5 URLLC use cases and requirement summarization, section 6.5 throughout
31st July, 2019	V 0.7	Karin Loidl (Fraunhofer IIS)	Section 6.4 on positioning
5th September, 2019	V 0.8	Karin Loidl (Fraunhofer IIS)	Minor update in section 6.4,
10th September, 2019	V 0.9	Dhruvin Patel, Joachim Sachs (Ericsson), Wangrui (China Mobile)	Chapter 7 updates on the Smart Grid and Factory of future deployment architecture, Section 5 update
13th September, 2019	V 1.0	Dhruvin Patel (Ericsson)	Chapter 2, 3
14th September, 2019	V 1.1	Anthony (HKT), Karin Loidl (Fraunhofer IIS), Bruno (Orange)	Networking slicing, Edge computing, Chapter 7 wide area URLLC reference architecture
25th September, 2019	V 1.2	Berna Sayrac (Orange), Bruno Tossou (Orange)	Update on technical enablers section and evaluation of the URLLC smart grid architecture
30th September, 2019	V 1.3	Ana Galindo-Serrano (Orange), Billy Liu (HKT)	Clarification on evaluation of TDD configuration, update in URLLC deployment architecture
7th October, 2019	V 1.4	András Zahemszky (Ericsson), Ana Galindo-Serrano (Orange), Dhruvin Patel (Ericsson), Joachim Sachs (Ericsson)	Update in network slicing aspect, addition of section on Non-Public Networks, update based on the Orange comments, editoral suggestions
7th October, 2019	V 1.5	Joachim Sachs (Ericsson)	Editorial update in chapter 7**.**
10th February, 2020	V 1.6	NGMN Office	Added missing references

Contents

1 EXECUTIVE SUMMARY

5th Generation (5G) of mobile communication network is foreseen to set the stage for enabling new applications across various verticals. This will create new business opportunities for verticals and Mobile Network Operators (MNOs). Vertical requirements developed in the previous report of NGMN project "Vertical Ultra Reliable Low Latency Communication (URLLC) requirements" summarise the key URLLC use cases as developed in various Standard Development Organization's (SDO). It is observed across various verticals the need to support URLLC communication services in 5G mobile communication networks. Today, commercial deployment of 5G network has already begun. Also, there is a huge momentum from verticals to integrate 5G networks. By nature of such verticals use cases, it is necessary for MNOs to understand various 5G deployment architectures.

The present report targets to demonstrate the capabilities of the 5G URLLC technology enablers, as developed by 3GPP standardisation body. The main outcome of this report is to showcase the various public and non-public network deployment architectures for vertical use cases identified in previous report of NGMN project. Furthermore, it provides an interpretation of different deployment options with respect to deployment attributes.

2 INTRODUCTION AND MOTIVATION

Starting from 3GPP Release 15, careful consideration is taken in both 5G New Radio (NR) and Core Network (CN) design to support URLLC communication services. Today 5G network products and solutions based on the Release 15 are commercially available, and 3GPP is moving forward to enable 5G for verticals. Looking at different verticals requirements [1] it is clear that the support for URLLC type of communication services is essential. It will make 5G an integral part of next generation Information and Communication Technology (ICT) infrastructure supporting such verticals. Considering ongoing rollout of 5G networks around the world, and high interest from verticals to rapidly adapt and integrate 5G networks, there is the need to understand from end-to-end (E2E) deployment architecture perspective how 5G system can fulfil URLLC vertical requirements.

The report describes from viewpoint of vertical requirements what are the main technical enablers of URLLC communication services taking URLLC vertical requirement report [1] as baseline. Finally, as the main outcome of the report, different deployment architectures of 5G mobile communication network are investigated including Non-Public Networks (NPN) which can be supported by MNOs.

Deployment architecture options for highly relevant verticals such as smart manufacturing, smart grid, 360^{o} panoramic Virtual Reality (VR) view video broadcasting and for VR cloud gaming that can be supported by MNOs are elaborated. 5G deployment architecture for smart manufacturing focuses more on the local area deployment while remaining verticals are more aligned towards wide area. Evaluation of such deployment options against requirement and deployment attributes is further described.

The main targets of the report are illustrated in the table below

Table 1 Targets of the report

Target 1	Demonstrate the capabilities of the 5G URLLC technology enablers/features
Target 2	Illustrate various 5G deployment architecture options to support URLLC use cases

3 References

[1] NGMN "Vertical URLLC Use cases and requirement v2.5.3" 16th July 2019, https://www.ngmn.org/wp-content/uploads/Publications/2019/190919-Verticals-URLLC-Requirements-v2.5.3.pdf

[2] J. Sachs, G. Wikström, T. Dudda, R. Baldemair, K. Kittichokechai, "5G radio network design for ultra-reliable low-latency communication", IEEE Network, vol. 32, pp. 24-31, Mar./Apr. 2018. https://ieeexplore.ieee.org/document/8329620 https://ieeexplore.ieee.org/document/8329620

[3] 3GPP TS 23.501: "System Architecture for the 5G system (5GS)", 3GPP, 2019

[4] 3GPP TS 38.214, "NR; Physical layer procedures for data," *Rel. 15, tech. rep., 3GPP,* June 2018.

[5] 3GPP TS 38.2012, "Multiplexing and channel coding," *Rel. 15,* 2018.

[6] 3GPP TS 38.2013, "Physical layer procedures for control," *Rel. 15,* 2018

[7] 3GPP TS 38.211, "NR; Physical channels and modulation," *Rel. 15, tech. rep.,* 2018

[8] R1-1907402 - IMT-2020 self-evaluation - NR URLLC user plane latency http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_97/Docs/R1-1907402.zip

[9] IEEE Std 802.1AS-Rev/D7.3, August 2018: "IEEE Standard for Local and metropolitan area networks--Timing and Synchronization for Time-Sensitive Applications".

[10] TR 38.855 Study on NR positioning support, March, 2019

[11] TS 38.305 Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN

[12] ITU-R M.2410 Minimum requirement related to technical performance for IMT-2020 radio interface, November 2017

[13] 3GPP TS 38.101, "User Equipment (UE) radio transmission and reception, Release 15, 2018

[14] 3GPP TS 22.261, "Service Requirement for next generation new services and markets", Release 15, October 2019

[15] 5G-ACIA non-public networks for Industrial Scenarios, March, 2019, https://www.5acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_5G_for_Non-Public_Networks_for_Industrial_Scenarios/5G-ACIA_White_Paper_5G_for_Non-Public_Networks_for_Industrial_Scenarios.pdf

[16] J. Sachs, K. Wallstedt, F. Alriksson, G. Eneroth, "Boosting smart manufacturing with 5G wireless connectivity," Ericsson Technology Review, January 2019. https://www.ericsson.com/assets/local/publications/ericsson-technology-review/docs/2019/etr-magazine-2019-01.pdf

[17] K. Gold, K. Wallstedt, J. Vikberg, J. Sachs, "Connectivity for Industry 4.0," book chapter in Dastbaz, Cochrane, Industry 4.0 and Engineering for a Sustainable Future, Springer 2019.

[18] GSMA, An Introduction to Network Slicing, 2017, https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf

[19] 5gsmart.eu

[20] János Farkas, Balázs Varga, György Miklós, Joachim Sachs 5G-TSN integration meets networking requirements, Ericsson Technology Review, August ,2019

[21] J. Sachs, P. Popovski, A. Höglund, D. Gozalvez-Serrano, P. Fertl, "Machine-type communications", book chapter in Osseiran, Monserrat, March, 5G Mobile and Wireless Communications Technology, Cambridge University Press 2016. https://www.cambridge.org/core/books/5g-mobile-and-wireless-communications-technology/machinetype-communications/E2D92674FF003B22E0922A40C1B62C42 https://www.cambridge.org/core/books/5g-mobile-and-wireless-communications-technology/machinetype-communications/E2D92674FF003B22E0922A40C1B62C42

[22] Ericsson, Network Sharing and Slicing for Railway, https://www.itu.int/dms_pub/itu-r/oth/0a/0E/R0A0E0000BA0001PDFE.pdf, July, 2018

[23] G. Wikström et al., "Wireless Protection of Power Grids over a 5G Network," 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), Bangkok, Thailand, 2019, pp. 976-981.

[24] NGMN "5G Extreme Requirements: End-to-End Considerations" 14th August 2018, https://www.ngmn.org/wp-content/uploads/Publications/2018/180819_NGMN_5G_Ext_Req_TF_D2_2_v2.0.pdf
[25] NGMN "5G Extreme Requirements: Radio Access Network Solutions" 5th June 2018, https://www.ngmn.org/wp-content/uploads/Publications/2018/180605_NGMN_5G_Ext_Req_TF_D2_1_v2.5.pdf
[26] NGMN "5G Extreme Requirements: Operators' Views on Fundamental Trade-Offs" 28th November 2017, https://www.ngmn.org/wp-content/uploads/Publications/2017/171128_NGMN_5G_Extreme_Requirements_Task_Force_D1_v1.0.pdf

4 SCOPE

The main scope of the document is:

1. Study how 5G can meet the vertical requirements from an E2E perspective, and what are the architectural considerations and deployment models.
2. Find out a balance between highly-specialised customisation of solutions and a one-size fits all types of approach. Address as many use cases as possible with a common solution, while providing some differentiation in the offerings for different industry sectors.

The report is structured as follows:

- Section 5 summarises the URLLC use cases and their requirements.
- Section 6 provides an insight to URLLC technical enablers in 5G.
- Section 7 describes 5G URLLC deployment architecture considering four use cases.
- Section 8 describes evaluations of such deployment architecture.
- Section 9 provide summary of the work.

5 URLLC use case and requirement summary

This report takes the prioritised use cases as basis to illustrate how 5G URLLC technology enablers support vertical requirements. Part of the prioritised use cases occur within a fixed area, such as controlling the journey of automated guided vehicles, production line enhancements, positioning Indoor/on campus, augmented workers, remote control of automated cranes. These types of use cases could be realised with dedicated deployments of a 5G network. Another part of the prioritised use cases, such as 360 panoramic VR view video broadcasting, Unmanned Aerial Vehicle (UAV) traffic control, Augmented reality (AR) and Mixed Reality (MR) cloud gaming, smart energy, outdoor positioning, can occur anywhere. This type of use cases should be realised with public deployments of the 5G network.

Based on the analysis above, we divide the use cases into two classified from the perspective of network deployment in this report:

1. **Local-area deployment**: use cases that can be realised with dedicated deployment of the 5G network within coverage range of 2 to10 kilometres.
2. **Wide-area deployment**: use cases that can be realised with public deployment of the 5G network with coverage range of more than 10 kilometres.

Table 2 summarises the use cases in these two categories with relevant requirement categories as described in [1]. Considering importance and versatility, five requirement categories are considered: reliability, latency, throughput, positioning and time synchronous accuracy.

Table 2 summary of use cases

Deployment	Use cases	Requirement category

		Reliability	Latency	Throughput	Positioning	Time Synchronous Accuracy
Local deployment	Control the journey of Automated Guided Vehicles (AGV)	99.999%	5ms	Ctrl:100 kbps 1080 P video: 3~8Mbps in UL	<20cm	
	Production line enhancement – robot tooling in the factory	99.9999%	1ms robotic motion ctrl; 1~10ms machine ctrl			
	Positioning Indoor/campus				<1m	
	Augmented worker	99.9999%	10ms		<1m	
	Remote control of automated crane	99.999%	20ms	Ctrl:100kbps 1080P video: 40Mbps in UL per crane		
Wide area deployment	360 panoramic VR view video broadcasting	99.999%	<20ms	40Mbps~5Gbps		
	UAV traffic control	99.999%	<100ms	<128bps		
	AR and MR cloud gaming	99.999%	<7ms (uplink)	1Gbps		
	Smart Grid Differential protection	99.999%	<15ms	2.4Mbps		10us
	Fault Isolation and System Restoration (FISR)		<25ms	10Mbps		
	Fault location identification	99.999%	140ms	100Mbps		5us
	Fault management for distributed electricity generation in smart grids	99.999%	<30ms	1Mbps		
	Positioning as service – Rural/Outdoor				<5m	

6 5G URLLC TECHNICAL ENABLERS

ITU-R WP5D in 2012 defined three usage scenarios including URLLC concerning both machine centric and human centric communication for next generation of International Mobile Telecommunications (IMT) system commonly referred as “5G”. 3GPP, a global standardization body considering such scenarios and requirements has developed an URLLC tool box of enablers. Reference [2] provides an insight to 5G NR radio interface design for URLLC communication services. Figure 1 shows the technical enablers to activate URLLC communication services in 5G networks. Further this sections dives into description of such technical enablers.

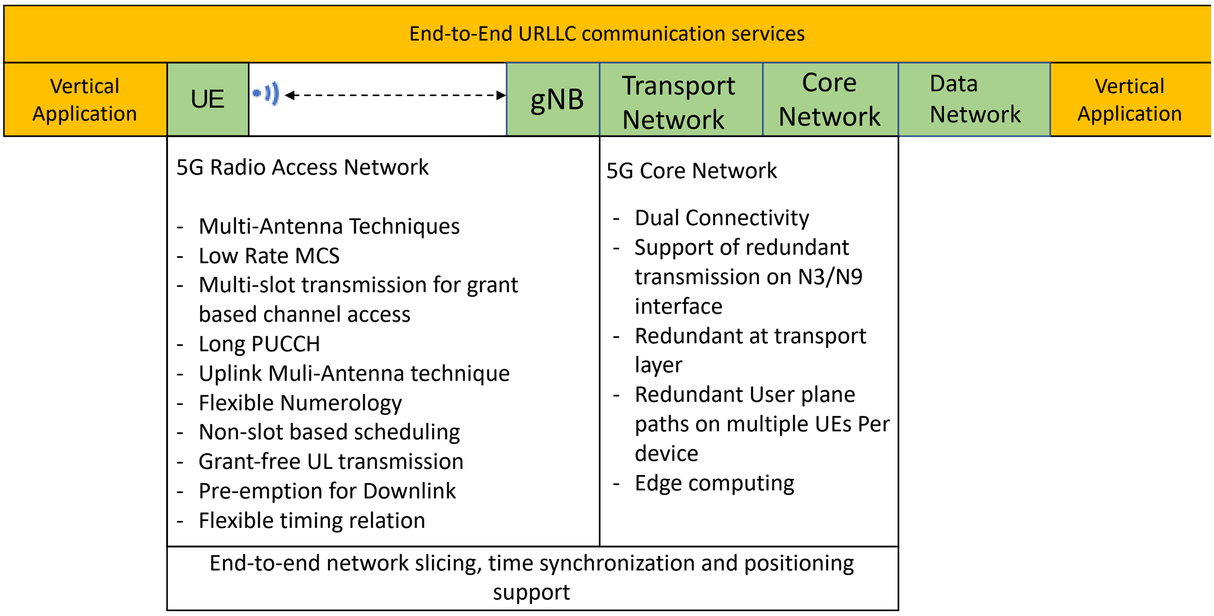


Figure 1: 5G technical enablers for URLLC communications services.

6.1 Reliability

Reliability represents the capability of transmitting a given amount of traffic within a predetermined time duration with high success probability. In what follows, features introduced in NR release 15 enabling high reliability features are detailed concerning enhancement in NR and 5G CN. Multi-slot transmission for the grant channel access, low rate Modulation and Coding Scheme (MCS) and further uplink multi-antenna techniques are key features which increase the reliability of the radio link between UE and gNB.

6.1.1 Low rate MCS.

This is a basic feature to provide the reliability target of 10^{-5} Block Error Rate (BLER) stated in Release 15 for URLLC applications. Reliability enhancement can be achieved by extending MCS to support operations at lower code rates. Figure 2 depicts channel adaption procedure.
To do so there are two Channel Quality Indicate (CQI) tables for URLLC CQI reporting:

- A first URLLC CQI table for target BLER 10^{-1}: It is the same as the existing 64QAM CQI table without any change.
- A second new CQI table for URLLC for target BLER 10^{-5}: this table will secure the reliability at first transmission if needed. The coding rate will range from as low as ~1/17 (QPSK) to the highest coding rate ¾ (assuming 64 QAM).

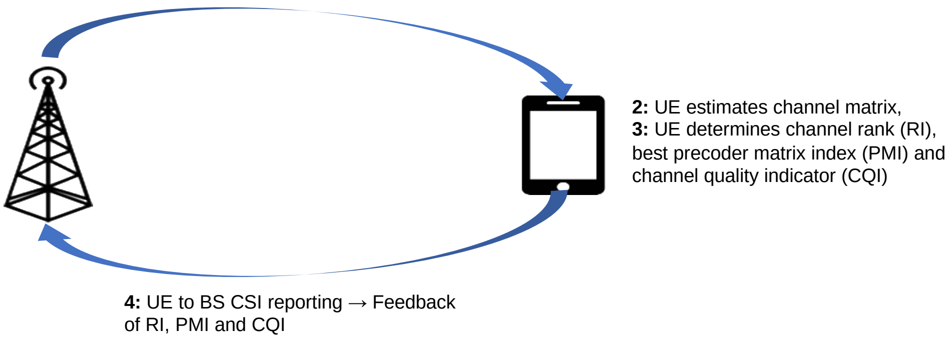


Figure 2: Channel adaptation procedure

6.1.2 Multi-slot transmission for grant based channel access

This feature consists of sending multiple times the transport block in consecutive slots without waiting for HARQ ACK/NACK messages. To do so, the UE is configured by upper layers with an Aggregation Factor > 1 in UL and DL in such a way that it will have the same symbol allocation across the Aggregation Factor number of consecutive slots.

The Aggregation Factor can be configured at 1, 2, 4 and 8 for both UL and DL and it should adapt to the latency budget i.e., the number of possible retransmissions might be limited depending on the service radio latency budget. This feature may be applicable for the high Sub-Carrier Spacing (SCS) options of NR thanks to the shorter slot duration.

The advantage of having consecutive repetitions is that it allows to cumulate the received energy for each information bit i.e., better Signal to Interference Noise Ratio (SINR), increasing the data transmission reliability. Also, it allows a reduction of control overhead. On the other hand, its drawback is that spectral efficiency can be affected. Depending upon the application requirement latency budget, Aggregation Factor can be configured in UE.

6.1.3 Long PUCCH

UL control channel Physical Uplink Control Channel (PUCCH) carries a set of control information called Uplink Control Information (UCI). UCI has different type of information such as: HARQ-ACK/NACK, Scheduling Request (SR), CSI reports. Having a long PUCCH consists in transmitting UL control channel in long duration i.e., from 4 to 14 UL symbol(s) in a slot, and it can be repeated over multiple slots. Long PUCCH can use from 1 to many UCI bits and can be transmitted in UL-only slots or UL-centric i.e., slot dominated by UL symbols with some DL symbols at beginning for the guard band. On top of improving the reliability, having long PUCCH will allow also provide a better coverage.

6.1.4 Uplink Multi-Antenna technique

Apart from providing high throughput, multi-antenna system also increases effective SINR thus increasing system robustness. In particular, two multi-antenna techniques are used in NR which increases URLLC robustness: UL Coordinated Multi-Point (CoMP) and UL pre-coding. UL-CoMP improves reception quality, URLLC communication services need to be in range of intra-site CoMP where sector antenna within gNB site are used in the reception. The feature does not require extra control signaling overhead and also can be applied without UE support. In NR three UE capabilities are defined for UL Multiple-Input Multiple-Output (MIMO) transmission: full coherent, partial coherence and non-coherent. NR supports codebook based or non-codebook based PUSCH transmission.

6.1.5 5G Core Network features to support high reliability

From release 15 onwards, new features and enhancements are introduced in 5G system architecture ensuring high reliability and availability for URLLC communication services [3]. Relevant solution concerning URLLC communication requirements for supporting high reliability by redundant transmission in the user plane are described below:

Solution-1: Dual connectivity based end-to-end redundant user plane paths
Figure 3 shows the dual connectivity based E2Eredundant user plane paths. Below is the procedure to enable the feature:

1. The user subscription indicates if a user is allowed to have redundant PDU sessions.
2. UE initiates two redundant PDU sessions.
3. Session Management Function (SMF) determines whether to handle them redundantly and selects different User Plane Functions (UPFs).
4. SMF allocates and provides to NG-RAN a Redundancy Sequence Number (RSN) for each PDU Session.

Different RSN values trigger NG-RAN to establish dual connectivity so that the sessions have E2Eredundant paths.

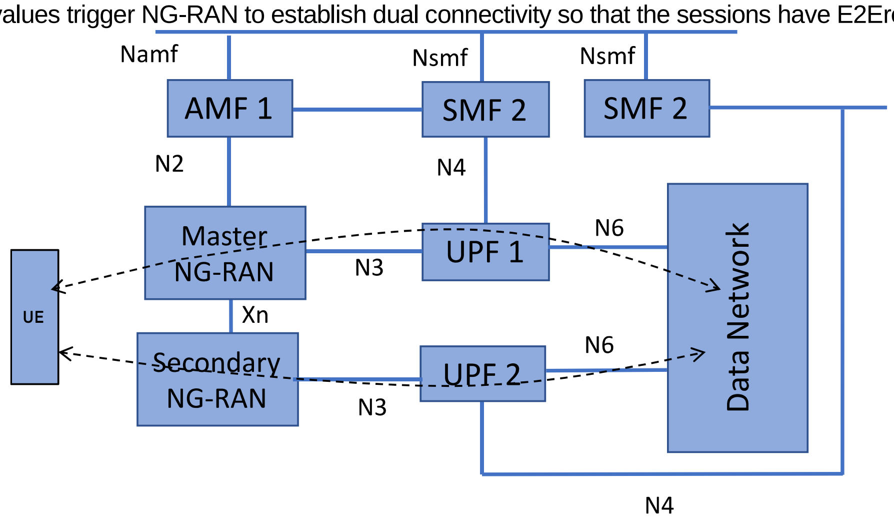


Figure 3: Dual Connectivity based E2E Redundant User Plane Paths

Solution-2: Support of redundant transmission on N3/N9 interfaces
In order to ensure high reliability and availability for backhaul deployment, network redundant transmission is performed on N3/N9 interfaces. For a given 5G deployment, two independent N3 tunnel are realised between the NG-RAN and UPFs. Single data PDU transmission is done over two independent N3 tunnels. To ensure such disjoint transport layer paths, SMF should be able to provide different routing information that need to be mapped to disjoint transport layer paths in network deployment configuration. For downlink, for each packet received by UPF from Data Network (DN), the same GPRS Tunneling Protocol (GTP-U) sequence number is assigned for two redundant transmissions; a similar action is performed by NG-RAN for uplink data transmissions. Figure 4 below describes the redundant transmission on N3/N9.

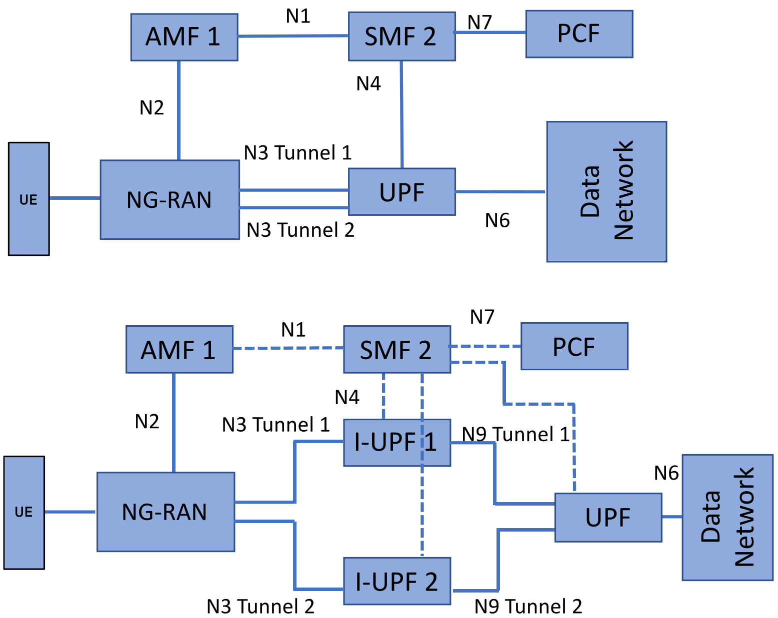


Figure 4: Redundant transmission performed on N3/N9 interfaces.

Solution-3: Support of redundant transmission at transport layer

Compared to solution 2, without enabling duplicating GTP-U sequence number over two disjoint paths, in this solution SMF selects UPF and NG-RAN which inherently support redundant transmission. This particular solution does not have impact on 3GPP protocols. Figure 5 describes the redundant transmission at the transport layer.

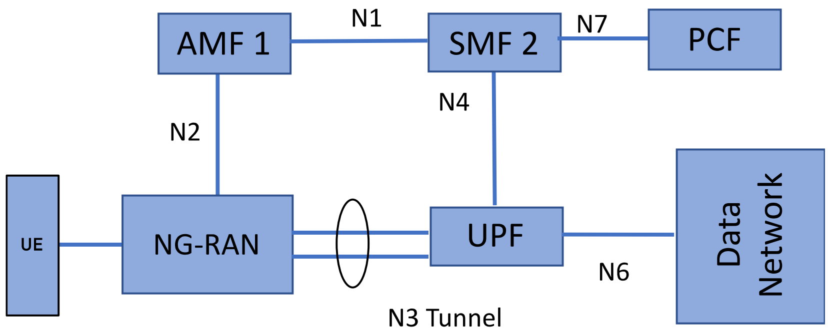


Figure 5: Redundant transmission at the transport layer.

Solution-4: Support of redundant user plane paths based on multiple UEs per device

In order to ensure high reliability and availability for communication services, redundant user plane paths are introduced. Figure 6 shows the redundancy framework, where redundant Radio Access Network (RAN) and CN deployment is considered for deployment. Also, this framework can facilitate 5G-Time Sensitive Networking (TSN) integration, where TSN features such as Frame Replication and Elimination for Reliability (FRER) can utilize multiple redundant user plane paths to achieve high E2Eavailability. It is also ensured in the architecture that UE PDU sessions utilize independent RAN and CN network resources. Two UEs can be implemented in the device and the traffic is sent towards the two UEs simultaneously. Figure 6 describes the redundant transmission based on multiple UEs per device.

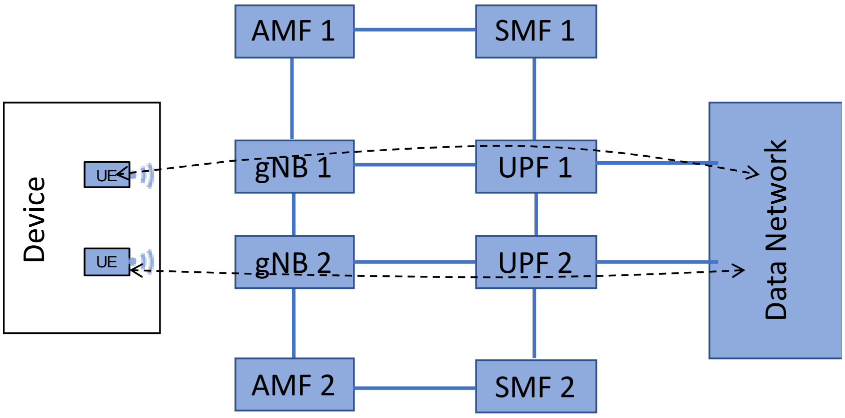


Figure 6: Redundant transmission based on multiple UEs per device.

6.2 Latency

5G system from Release-15 has introduced a toolbox of low latency features including Edge Computing (EC), flexible numerology, mini-slots, short PUCCH, faster processing time, fast HARQ, UL transmission and edge computing.

6.2.1 Edge computing enabling low latency

Edge Computing (EC) refers to data storage and processing at the edge of a network instead of centralized cloud data centres. EC is a method of optimizing cloud computing systems by pushing computing power and storage capacity to the Edge of a network (i.e. in EC servers), near the source of data, reducing the application traffic and enhancing user privacy of applications involving private data. Furthermore, the performance of latency-sensitive applications is improved by EC because of the short geographic distance to the computing function [3].

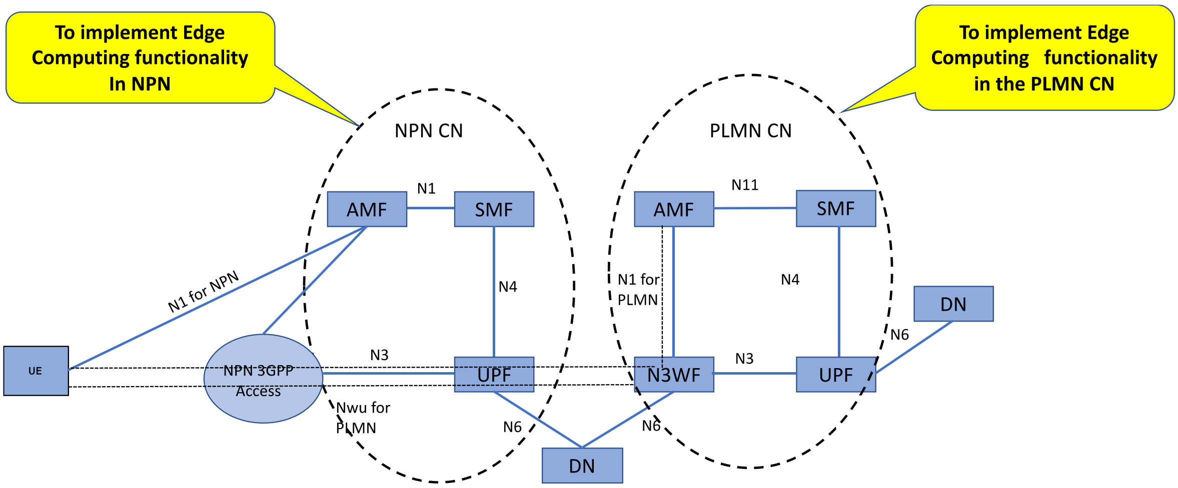


Figure 7: Access to PLMN services via Stand-alone Non-Public Network.

The support of EC is specified in 3GPP SA2:
EC enables vertical applications to be hosted close to UE's access points. This allows to provide efficient service delivery through reduced E2Elatency and load on transport network. UPF in close proximity to UE is selected by 5G CN where traffic is steered from UPF to local DN via N6 interface.

There are many ways by which EC can be supported by selection of UPF to route the user traffic to the local DN, for example considering non-public network deployment, 5G CN will select UPF in geographical proximity or may be in premises of the URLLC communication service application server.

Figure 7 shows which part of 5G system architecture can be realised to have EC functionality for non-public and public network deployment. To realise Non-Public Network (NPN) for local area deployments, Access and Mobility Management Function (AMF), SMF and UPF can be implemented as an EC node. Chapter 6 provides further details on the possible NPN deployments. Concerning wide area URLLC use cases the PLMN CN can be implemented as EC node.

6.2.2 Flexible numerology

NR is based on Orthogonal Frequency Division Multiplexing (OFDM) and uses as reference the numerology used in LTE. A numerology is defined by a SCS and a Cyclic Prefix (CP) overhead. In the case of LTE the SCS was fixed to 15 kHz resulting in a OFDM Symbol (OS) duration of 66,67 µs. In NR the numerology can be flexible, this means that the SCS can be of 15, 30, 60 and 120 kHz [4]. On the one hand, increasing the SCS means reducing the OS duration i.e., for SCS 30 kHz, the corresponding OS duration is 33,33 µs, for 60 kHz SCS the OS duration is 16,67 µs and for 120 kHz the OS duration last only 8,33 µs. On the other hand, as presented in Figure 8, reducing OS duration by increasing the SCS requires higher channel bandwidth.

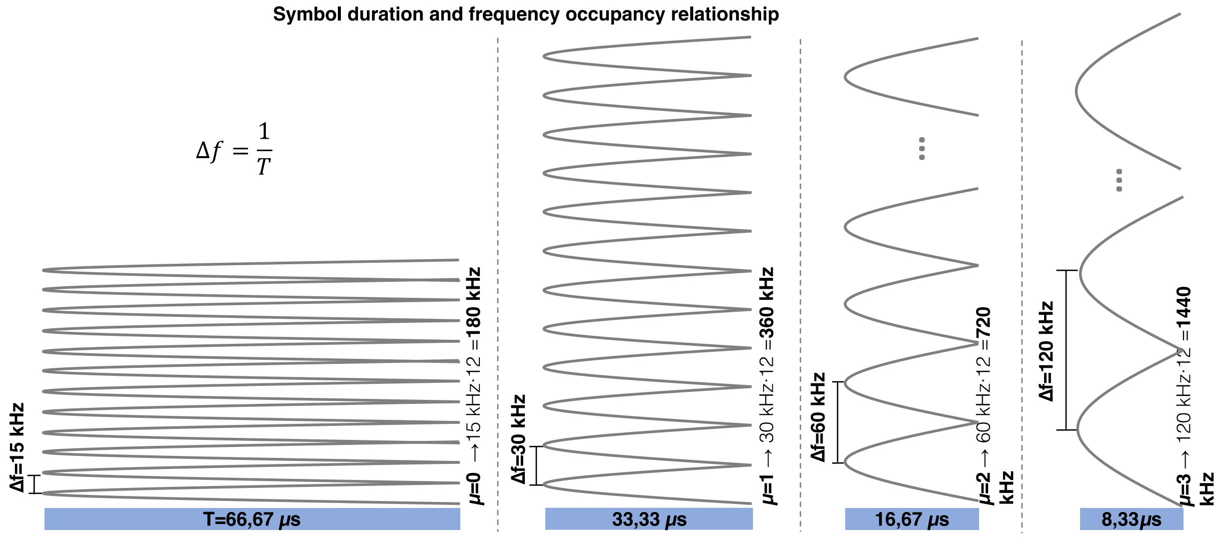


Figure 8: Symbol duration and frequency occupancy relationship for different SCS.

Figure 9 presents the LTE and NR time reference intervals. The reference frame duration is of 10 ms containing 10 subframes of 1 ms each. In LTE the scheduling is done on a subframe basis. Reducing the SCS also means that in 1 ms it will be possible to accommodate more sets of OS, or slots. For example, as presented in Figure 9, when considering a slot of 14 OS with 15 kHz SCS, one subframe will accommodate one slot, for 30 kHz SCS, one subframe will contain two slots, for 60 kHz SCS one subframe will enclose four slots and for 120 kHz SCS one subframe will contain eight slots. The main advantage of having more slots per subframe is to have higher scheduling opportunities and therefore higher retransmission possibilities, crucial to increase the reliability.

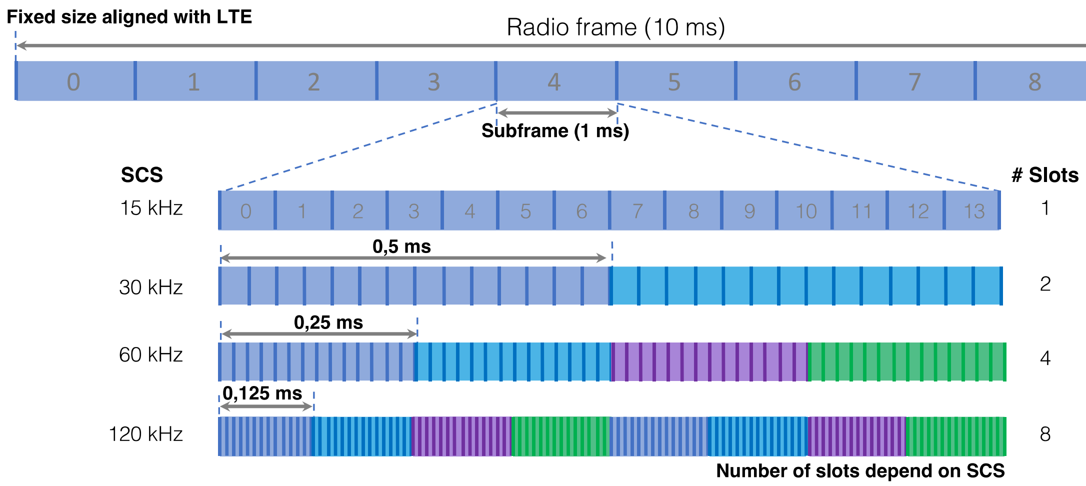


Figure 9: Number of slots per subframe for different SCS.

6.2.3 Non-slot based scheduling

The duration of a transmission in the radio link is known as the Transmission Time Interval (TTI) duration which depends on the number of OS per slot. In LTE it is fixed to 14 and in the case of NR, the number of OS per slot can be configured in a more flexible way. To do so, 3GPP has introduced the concept of mini-slot which consists in transmitting two, four, or seven symbols instead of fourteen [4]. A mini-slot can start at any OFDM symbol without waiting for the start of a slot boundary. Thus, the mini-slot feature supports transmissions shorter than the regular slot duration and also allows transmissions to start immediately. The scheduling can be done in a slot- or non-slot based manner:

- Slot based (Mapping type A): The UE receives the scheduling information at the beginning of the slot which contains the time allocation of Physical Downlink Shared Channel (PDSCH)/ Physical Uplink Shared Channel (PUSCH), i.e., the starting symbol and the length in terms of contiguous symbols, as presented in Figure 10. The resource allocation in time-domain for mini-slot in UL and DL will be given by:
  - K $\rightarrow$ identifies the slot.
  - S $\rightarrow$ identifies the starting symbol relative to the start of the slot.
  - L $\rightarrow$ identifies the number of consecutive symbols counting from S.

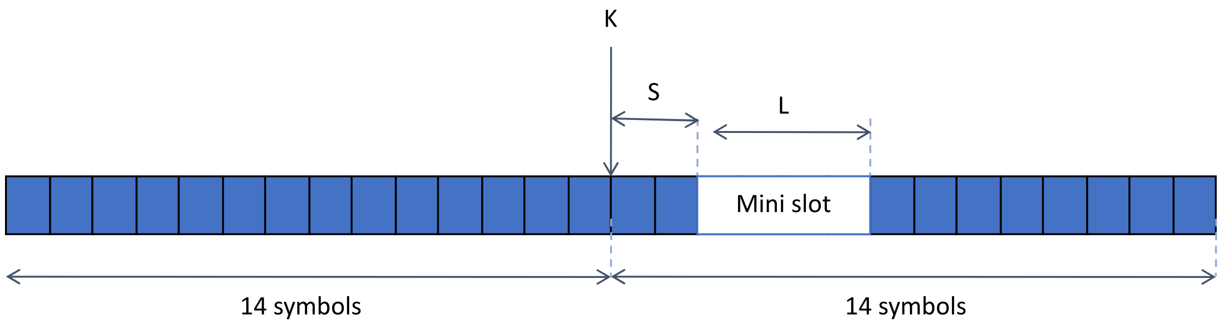


Figure 10: Resource allocation for mini-slot mapping type A.

- Non-slot based (Mapping type B or mini-slot): The UE will be configured (by the Radio Resource Control (RRC) layer) with a given search space which determined the set of resources that the UE has to monitor. The UE is configured with a monitoring periodicity of minimum one symbol that will start at the first symbol of the search space. When the control information is received the subsequent symbols given by the size of the mini-slot (2, 4 or 7 OS) will carry the data. Figure 11 presents an example of non-slot based scheduling with monitoring periodicity of 4 symbols and mini-slot of 2 OS. This feature provides fast transmission opportunities at the cost of higher control overhead and energy consumption at the UE side.

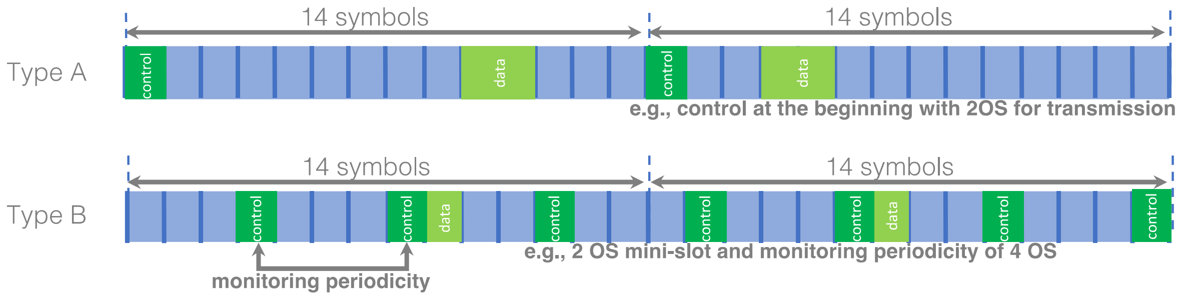


Figure 11: Resource allocation for mini-slot mapping type B.

6.2.4 Grant-free UL transmission

In mobile networks, the gNodeB is able to efficiently carry out downlink transmissions since it has complete control of the medium access. However, up to now, the uplink transmissions needed an extra signalling overhead to indicate to the UEs the resources to be used to perform the communication, which also leads to undesirable communication delays. The LTE dynamic scheduling scheme is illustrated in Figure 12 (a). When a UE has data to transmit it i) sends a Scheduling Request (SR) to the eNodeB, ii) the eNodeB allocates radio resources to transmit the requested traffic and notifies them to the UE using a Scheduling Grant (SG) and iii) the sends its data in the assigned resources. Given the default TTI size of 1 ms and the SR periodicity of 10 ms, the average uplink latency in LTE is of 11,5 ms. In 5G Release 15 grant-free uplink access was introduced in order to allow the UE to have a faster access to the channel. Grant-free is faster since the SR and SG phases are removed, as presented in Figure 12 (b). Based on Semi-Persistent Scheduling (SPS), the gNodeB can configure the UE to have pre-allocated periodic radio resources available for transmissions. More precisely, the gNodeB provides configured UL transmission opportunities to the UEs. *N* Transmission Occasions (TO) are configured within a period *P* for repetition and retransmission, and each TO is used for transmission of one repetition of a Transmit Block (TB). The start of repetition of a TB within each periodicity is flexible. Therefore, a device can directly transmit when it has data to send, reducing the average radio access delay for uplink data by more than half, the equivalent of one feedback round-trip time. Furthermore, the explicit signaling overhead of SR and SG is eliminated. On the other hand, the reserved resources are tied up for the UE channel access, but by assigning overlapping grants to multiple UEs the resource waste is reduced, and at lower loads the impact on reliability due to collisions can be manageable.

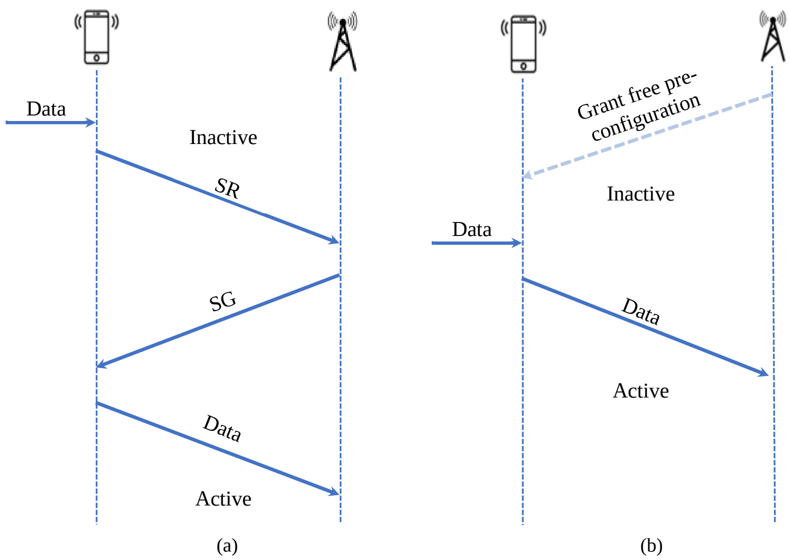


Figure 12: Grant-based (a) vs. Grant-free (b) uplink channel access.

6.2.5 Pre-emption for DL

A resource and latency-efficient scheduling solution is to multiplex data with different TTI lengths (i.e., mini-slots and slots) particularly in the case of resource limitations. This type of multiplexing is also referred to as preemption. In NR, a mini-slot carrying high-priority or delay-sensitive data in DL can preempt an already ongoing slot-based transmission on the first available OFDM symbol without waiting until the next free transmission resource. This operation enables ultra-low latency for mini-slot-based transmission, especially in the scenario where a long slot-based transmission has been scheduled. This feature allows URLLC services to pre-empt resources already allocated to eMBB services. This functionality is activated per UE, therefore, the UE is configured by higher layers with the parameter *Preemp-DL=ON* to monitor the DL control channel containing the Pre-emption Indication (PI).

When the gNodeB decides to pre-empt radio resources allocated to some ongoing eMBB transmission, the punctured resources are communicated to the eMBB UE based on a PI carried in next slot Downlink Control Information (DCI), as presented in Figure 13. Those resources should not be taken into account for the decoding/combining of retransmissions.

The UE behavior in between the reception of the PI and the ACK/ NACK transmission is up to implementation i.e., the UE can re-try to decode the preempted packet or can ask for retransmission of the incorrectly decoded code blocks only [6] [7]

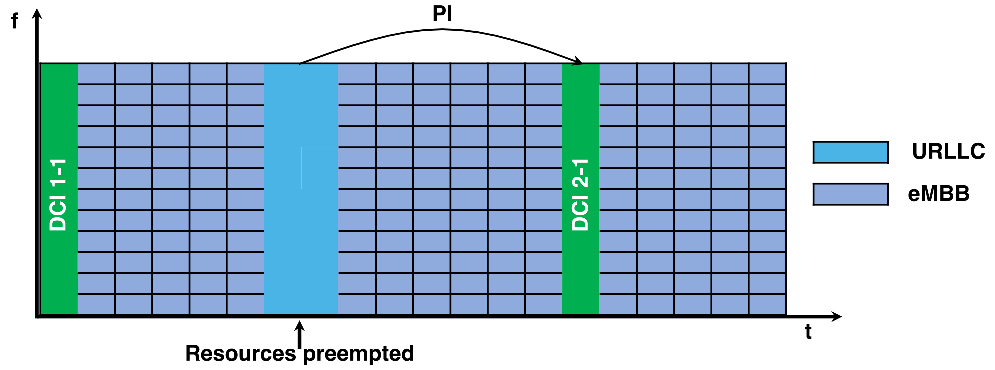


Figure 13: Preemption mechanism for DL.

6.2.6 Flexible timing relation

In radio communications when a packet is transmitted an acknowledgment response is required. In NR the timing relation between the reception of data and the transmission of the acknowledgment has been optimized with respect to LTE. Figure 14 represents the timing relations in uplink and downlink defined by 3GPP and given by:

- *K1* → Delay in TTI between downlink data (carried by PDSCH) reception and corresponding ACK/NACK transmission on uplink.
- *K2* → Delay in TTI between uplink grant reception in downlink and uplink data (carried by PUSCH) transmission.
- *K3* → Delay in TTI between uplink NACK reception and corresponding retransmission of data (PDSCH) in downlink.
- *K4* → Delay in TTI between uplink data reception and corresponding ACK/NACK.

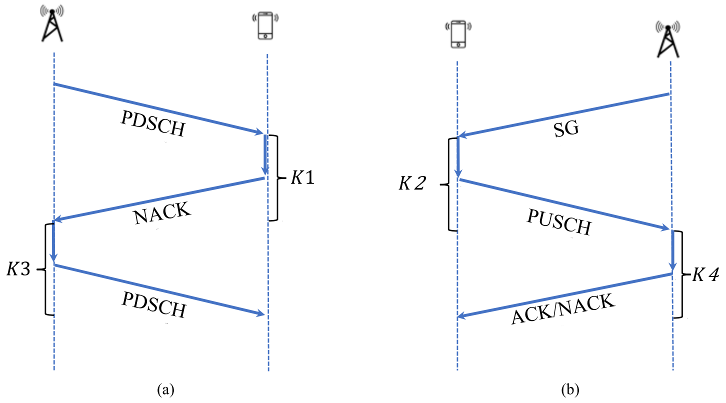


Figure 14 FDD timing relations in (a) DL and (b) UL.

As presented in Figure 14 for the case of LTE in Frequency Duplexing Division (FDD), the ACK/NACK for downlink data PDSCH or the uplink data corresponding to the uplink grant transmitted to the UE in subframe *n* is sent in subframe *n + 4*. In the case of NACK or PUSCH transmitted in *n + 4,* the eNB retransmits the downlink data or the uplink ACK/NACK in subframe *n + 8*. So, in LTE there is a fixed timing relation for transmission and acknowledgement. On the other hand, in NR, a flexible timing relation has been introduced, since the acknowledgment can be transmitted in sub-frame *n + 1, 2, 3* or *4*, as presented in Figure 15. *K1* and *K2* will be configured depending on the UE processing capability and network load.

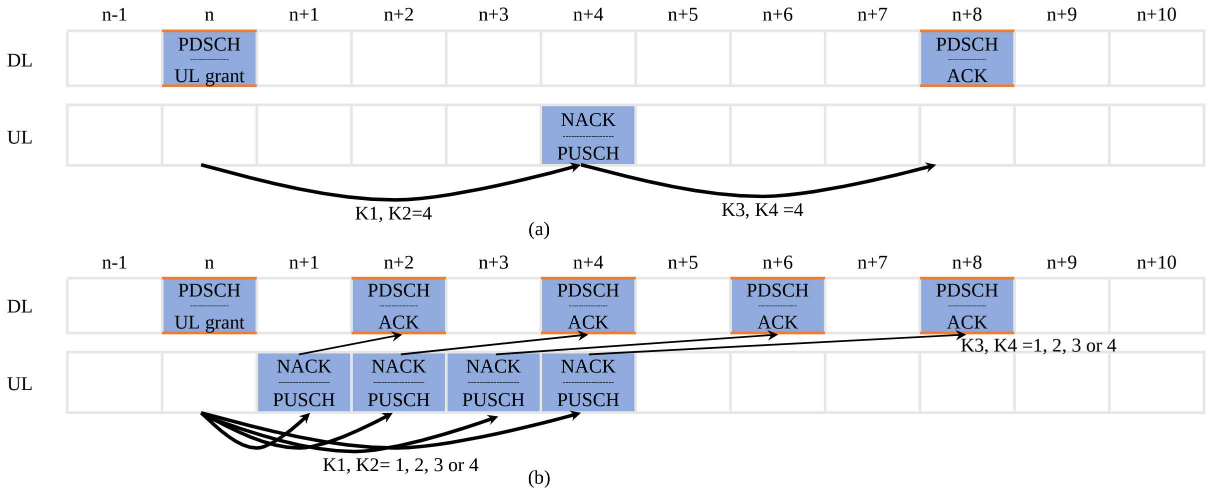


Figure 15: FDD timing relations for (a) LTE; (b) NR.

For the case of LTE in Time Duplexing Division (TDD), it has 7 possible fixed configurations for DL/UL ratios with maximum two special sub-frames per frame. DL data (PDSCH) and corresponding acknowledgment in UL is given by *K1* $\geqslant$ 4. An example of LTE TDD configuration type 2 is shown in Figure 16. UL grant in DL and corresponding UL data (PUSCH) transmission given by *K2* $\geqslant$ 4. On the other hand, NR time relations are flexible so that *K1*, *K2*, *K3* and *K4* can be dynamically adapted by the network depending on the UE processing time capability, timing advance, DL/UL ratio and switching points. But in NR it is also possible to have a fix configuration as it is in the case of LTE. Figure 17 presents NR TDD configuration type 2, where *K1* is not any longer limited to be higher than 4 subframes.

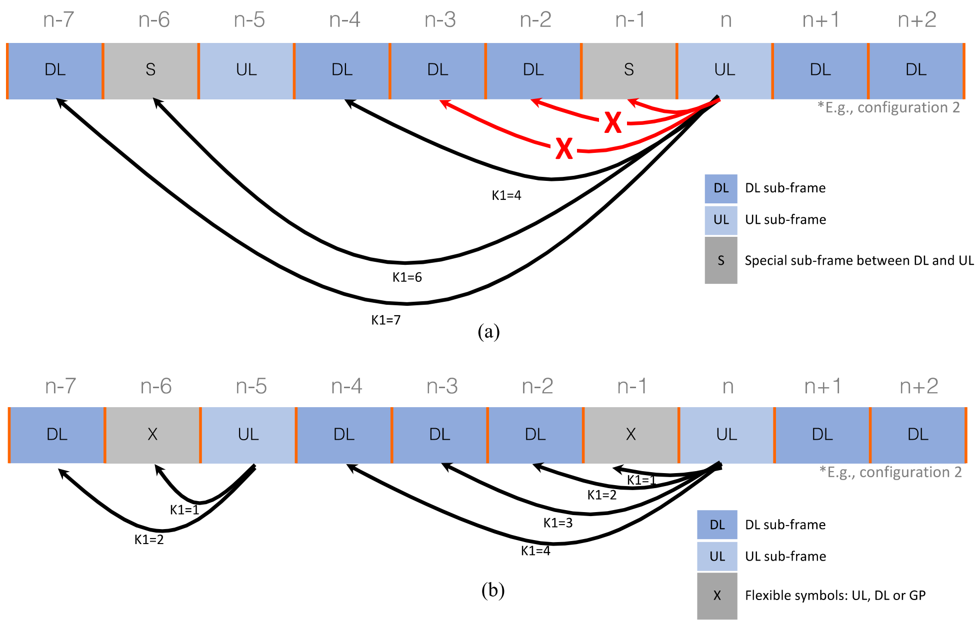


Figure 16: TDD timing duplexing, LTE (a) and NR (b).

Nevertheless, in NR TDD, there is full flexibility as each slot can be independently configured using the Slot Format Indicator (SFI) carried by group common PDCCH. The SFI carries an index to the SFI table which contains 57 possible configurations, going from all symbols in downlink or uplink, or a mix of DL/UL/X symbols, going up to two switching points. This configuration is called dynamic slot configuration and some slot configuration examples are presented in Figure 17 [6] [7].

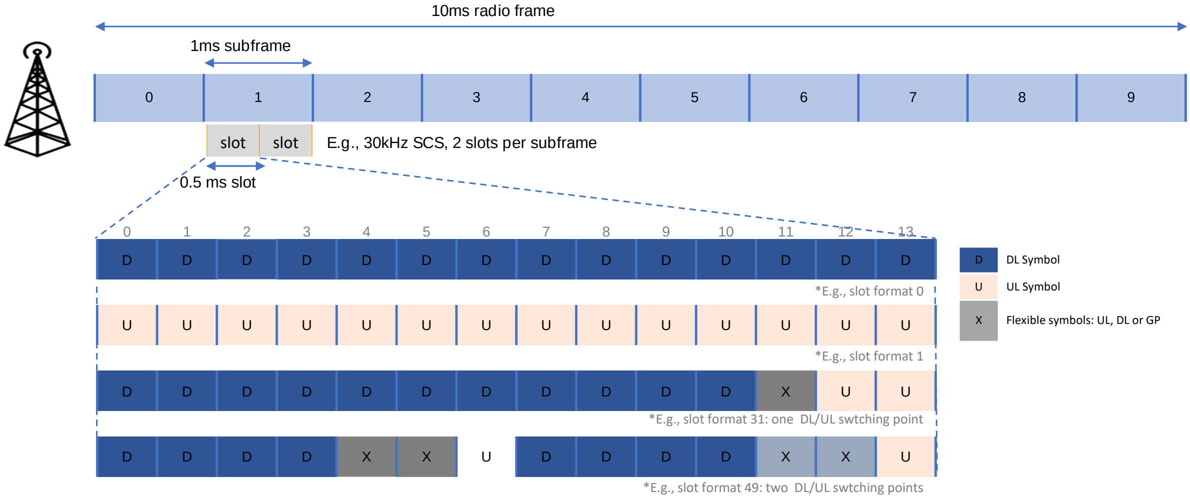


Figure 17: dynamic slot configurations.

In NR TDD there exists also the semi static configuration which consists of DL-unknown-UL configuration with periodicity of 0.5ms, 0.625ms, 1ms, 1.25ms, 2ms, 2.5ms, 5ms, and 10ms. This will depend on slot duration given by the SCS. As presented in figure 18, DL-unknown-UL periodicity consists of X full contiguous DL slots (from left to right), Y full UL slots (from right to left) and Unknown slots in between DL and UL slots which can be overwritten by a per slot using the SFI carried by group common PDCCH and/or per symbol using the DCI carried by PDCCH.

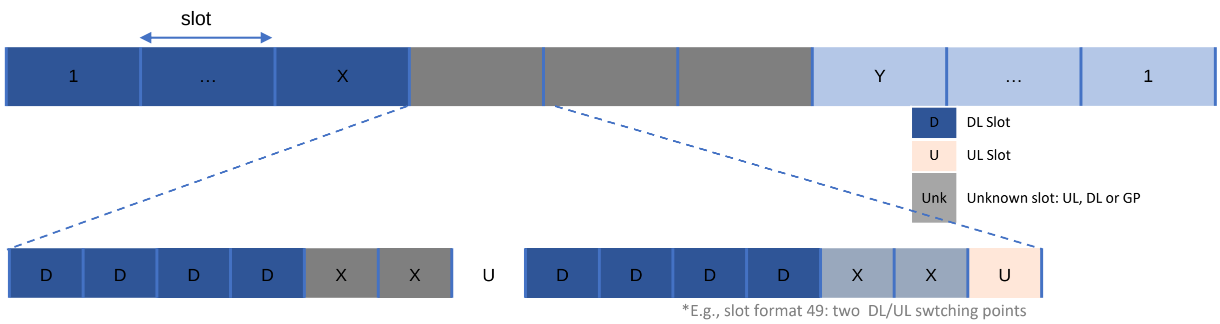


Figure 18: Semi static slot configuration.

Finally, this flexible timing relation introduces the concept of flexible HARQ which allows adapting ACK/NACK transmissions depending on the latency budget. However, reduced HARQ timing will depend on the UE capability; therefore it will not be implementable for all UEs.

Latency Estimation

The overall latency depend upon the various latency components such as processing time at the UE and gNB, Alignment time, transmission and feedback time. The interesting aspect is to observe the worst case latency in User Plane (UP). Figure 19, Figure 20 and Figure 21 show the worst latency in DL, considering both SR and configure grant uplink channel access. Here, the worst case arrival time in a slot is considered which will introduce an alignment delay. The evaluation is done assuming FDD configuration.

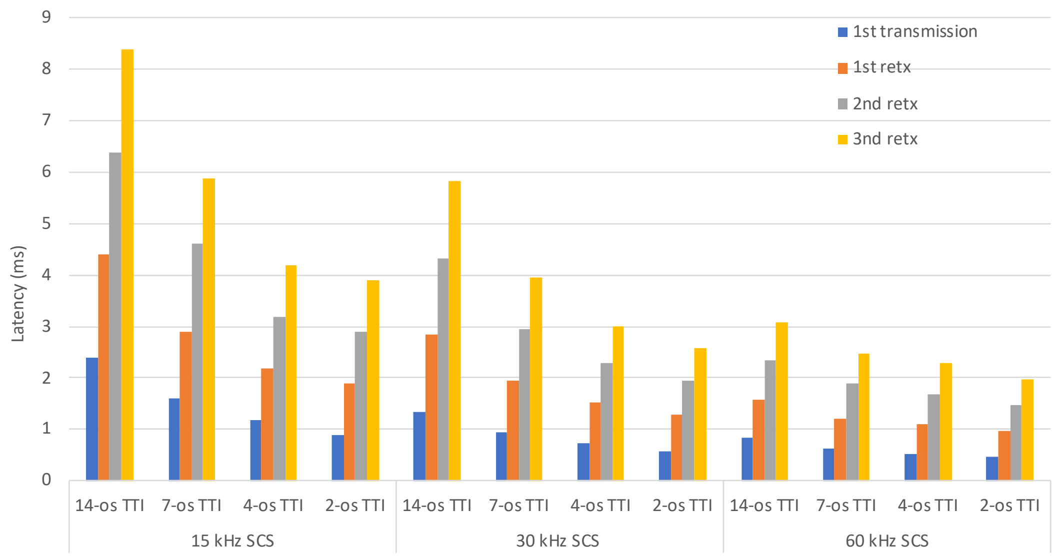


Figure 19 Worst Case downlink latency in FDD with Release 15 capabilities Rel-15 UE capability #2 and 4 PDCCH monitoring occasions

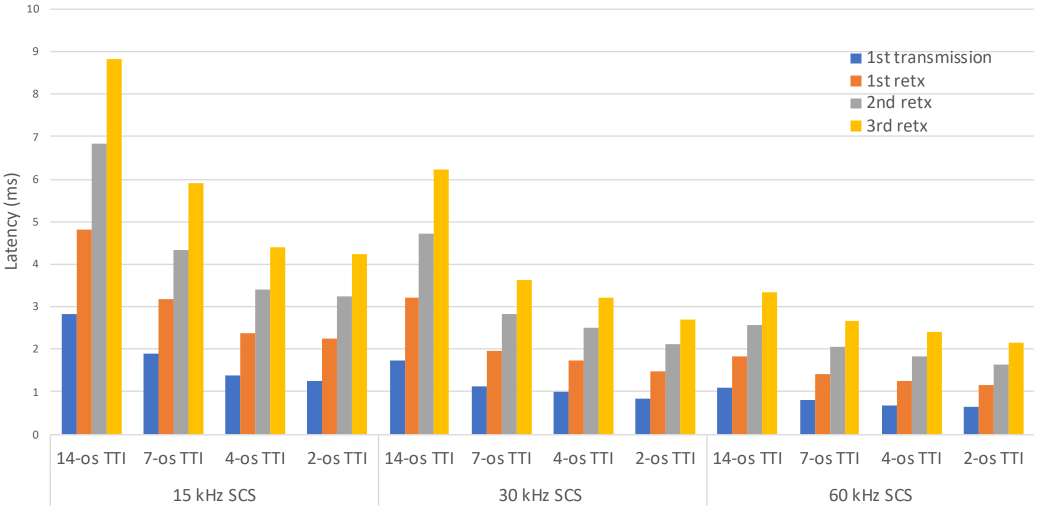


Figure 20 Uplink latency in FDD considering SR with Release 15 capabilities #1 and 4 PDCCH monitoring occasions

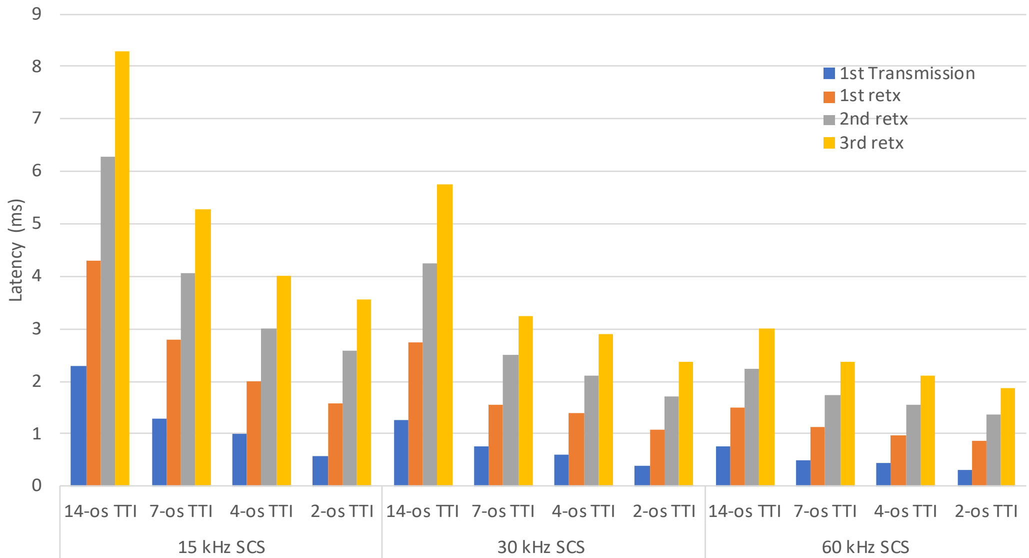


Figure 21 Cofigured grant uplink latency with Release 15 capabilities #1 and 4 PDCCH monitoring occasions

Worst case latency performance for TDD is shown in Figure 22, where the considered slot configuration is [D,D,D,D,D,D,D,G,U,U,U,U,U,U] for DL latency evaluation and [D,D,D,D,D,D,G,U,U,U,U,U,U,U], for UL latency evaluation where D is a symbol in downlink, U a symbol in uplink and G is a guard symbol. This is one of the possible configurations that may be used thanks to the dynamic TDD configuration introduced by NR. Nevertheless, this type of configurations will be more likely in millimeter waves than in midbands. In midbands it may be needed to have a TDD frame configuration alignment between operators in the same region in order to avoid interference. It is clearly observed for FDD, both DL and UL transmission with 30KHz SCS fulfills requirement of 1ms URLLC requirements. Further assumption on simulation parameters can be found in the 3GPP contribution[8].

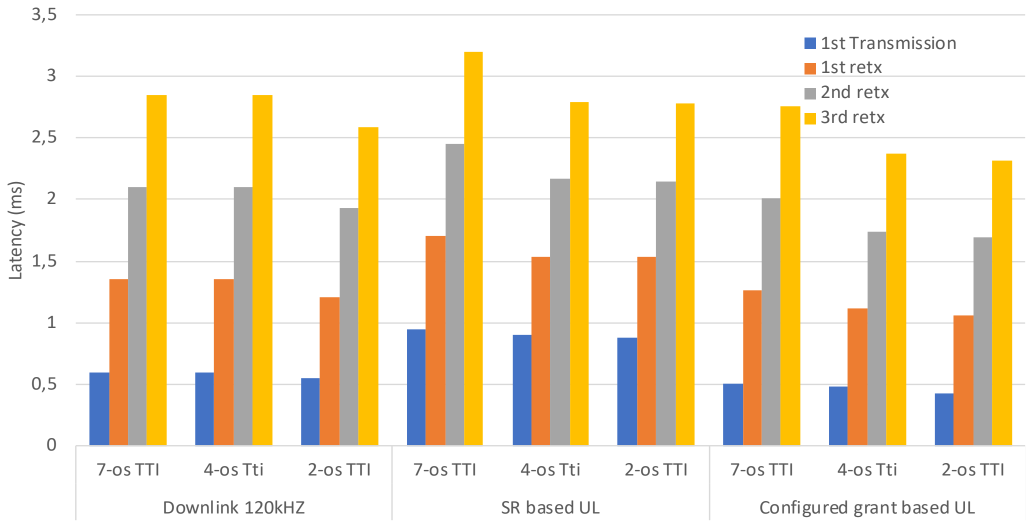


Figure 22 Worst case downlink and uplink (SR based and Configured grant based UL) latency TDD latency with Release 15 capability #1 and 4 PDCCH monitoring occasions

6.3 Time synchronization

Time synchronization is one of the key requirements for URLLC applications in vertical domains such as industrial automation and smart grid. For instance, in smart manufacturing, time synchronisation is utilised for industrial control, time aware scheduling in the TSN packet network. For smart grid applications, time synchronisation between nodes is required for time stamping of the Phasor Measurement Unit (PMU) measurements.

In the 5G system time synchronization will also be supported with the integration of TSN in Release 16. TSN Translators (TT) are introduced at the edge of the 5G system which support TSN IEEE 802.1 operations, DS-TT at UE side and NW-TTs at UPF side. UE, gNB, UPF, NW-TT and DS-TT are synchronised with the 5G system internal clock. The E2E5G system can be considered as an 802.1AS "time-aware system".

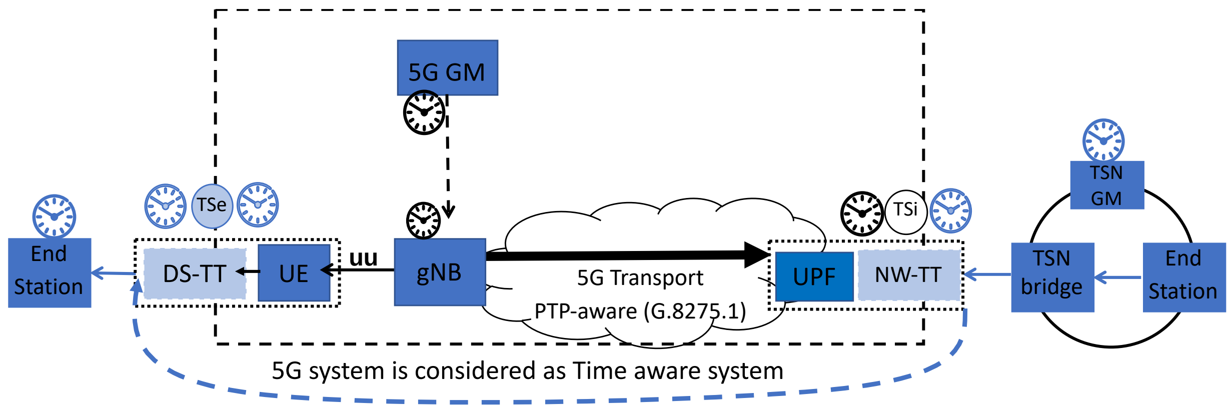


Figure 23 5G system model as IEEE 802.1AS compliant time aware system for supporting TSN time synchronisation [3]

As depicted in the Figure 23, two time synchronisation systems are independent including 5G System (5GS) synchronisation and TSN domain synchronisation. NG RAN synchronisation mechanism is utilised for the 5G system [3]. TSN system time synchronisation is specified by IEEE 802.1AS [9]. To enable E2Etime synchronisation over the 5G system, the 5GS calculates the measured residence time for generalised Precision Time Protocol (gPTP) packets between the ingress and egress TSN Translators and adds this residence time into the correction field of the time synchronisation packets of the TSN domain. TTs operations include timestamping, gPTP support, and the Best Master Clock Algorithm (BMCA).

6.3.1 5G system support for TSN time reference distribution

Time synchronisation for TSN as specified by IEEE P802.1AS utilise generalised Precision Time Protocol (gPTP). gPTP supported networking nodes creates a time aware network, which specifies the operation of time-aware systems on packet network [9]. To support seamless TSN integration, in downlink on arrival of gPTP frames, NW-TT entity (i.e. edge of the 5G system) performs ingress timestamping for each gPTP (sync message). Further, gPTP messages are transmitted over user plane on a specific PDU session. On receiving gPTP message at UE side, DS-TT creates an egress timestamping for the gPTP messages for external TSN working domains. The difference between ingress and egress timestamps is the overall residence time spent by the gPTP message in the 5G system, which is forwarded to the end station of the TSN network, for resolving time difference enabling time synchronisation. The timestamping on DS-TT and NW-TT side is based on the 5G system clock.

6.3.2 5G New Radio time reference delivery method

As part of enhancement of NR to support Industrial Internet of Things (IIoT), new capabilities over the radio interface are developed enabling accurate time reference delivery form a gNB to one or more UEs. Considering the 5G system internal clock being distributed throughout 5G network as shown in Figure 23, the gNB acquires a reference time values, and it modifies the acquired time reference to a reference point in the system frame structure occurring at the BS Antenna Reference Point (ARP). Further, this value indicating the corresponding reference point (System Frame Number) is embedded in the System Information Block (SIB) or RRC message which is further transmitted to the UE. The message also contains an uncertainty value regarding the reference time application to the reference points. This reference time received by the UE is further used to perform egress timestamping for calculating residence time of gPTP messages in the 5G system.

6.3.3 Support for multiple TSN working domains

Various URLLC applications operating over the same networks can have conflicting timescale requirements. TSN supports multiple time domains on the same physical network by enabling multiple gPTP domains; each supports a separate industrial control application for instance. Each TSN working clock domain has its own gPTP messages transmission including a domain number and has its own grand master reference clocks. This domain number indicates the time domain of the gPTP message. The 5G system supports multiple TSN working clock domains, at the edge of the 5G system NW-TT performs timestamping for the gPTP event message of all domains. UE on receiving gPTP messages, forwards all the message to DS-TT which performs egress timestamping for the gPTP event message for every external TSN working domain.

6.4 Positioning

Positioning in 5G offers huge benefits for use cases like Industry 4.0. These benefits are grounded on improvements in accuracy and availability due to a larger bandwidth of higher frequencies and dense deployments. To actually calculate the position of an object one can choose between several measurement principles. The most common ones are timing based estimation using Time Difference Of Arrival (TDOA) or Round Trip Time (RTT) methods, Direction based estimation (DoA) based on AoA (Angle of Arrival) or AoD (Angle of Departure) measurements and Received Signal Strength based Estimation (RSSI). LTE supports TDOA techniques in Uplink (UTDOA) and Downlink (OTDOA) as well as Enhanced Cell ID (E-CID) which mainly includes RTT, AoA and RSSI measurements. These approaches will be also supported by NR and extended to include new measurements like AoD estimation with beam sweeping and extend other procedures like multi-RTT to enable efficient RTT measurements with neighboring cells.

6.4.1 Positioning methods in a nutshell

Positioning methods for mobile networks can be categorized into Radio Access Technology (RAT)-dependent and RAT-independent techniques. RAT-dependent techniques rely on measurements carried out by the Radio Access Network (RAN) on its own using the 3GPP air interface, e.g. E-CID, OTDOA, UTDOA estimation or AoA measured at the base station in case the necessary antenna configuration is supported. RAT-independent techniques are methods that are 3GPP external, like satellite systems GNSS/A-GNSS/RTK, WLAN, Bluetooth, inertial sensors and barometric sensors. RAT-independent techniques may be integrated into positioning schemes depending on the capabilities of the UE. The main characteristics of the RAT-dependent technologies are summarized in the following.

Timing based estimation (TOA)
Principle: The concept of timing based estimation requires measuring the propagation time from the transmitter to the receiver, to be specific between the antennas of the respective devices. The propagation time of the radio wave is directly proportional to the distance that it travels. By multiplying it with the speed of light the distance between, e.g. a mobile terminal (UE) and a base station (gNB) is obtained. In order to measure the propagation time directly from the TOA measurement, the transmission time relative to the receiver clock must be known. This requires that clocks at the transmitters and receivers should be synchronized.

Observed time difference of arrival OTDOA (DL-TDOA)
Principle: The UE measures the TOA from the received DL-PRS transmitted from multiple synchronized gNBs. The UE then reports the Reference Signal Time Difference (RSTD) to the Location Management Function (LMF). At least three RSTDs measurements are needed to determine an unambiguous UE position. DL-TDOA is an efficient method if many UEs in an area must be localized or when UL signal cannot be received from enough neighboring gNBs/Transmission Reception Point (TRPs). Due to processing gain of the PRS and the high transmit power in DL, UEs with challenging channel conditioning can still be detected. On the other hand, DL-TDOA requires "always on broadcast signals" from all gNBs/TRPs in the positioning area.

Uplink time difference of arrival UTDOA (UL-TDOA)
Principle: UE transmits UL-PRS signal (SRS, for example). The UL-PRS signal is received by many receiving units. Each gNB reports to the LMF the TOA relative to a common reference time. UTDOA has a very low processing complexity and power consumption in UE and the position update rate can be adjusted individually for each UE (e.g. depending on the speed of the device). Higher performance schemes (e.g. industrial applications with high accuracy indoor positioning) can be supported. But required resources used for UL-PRS may become high if many UEs are located.

Direction based estimation (DoA)
Principle: Direction estimation in uplink or downlink is per definition network centric as it is based on multiple antenna gNB capability to position a mobile user. For uplink, the gNB measures the AoA of the reference signal transmitted from an UE. The gNB forwards the AoA measurements to the LMF, which can be combined with RTT measurements of the serving cell or with AoA measurements from the neighboring cells to determine the UE position. For downlink, Angle of Departure (AoD) is an approach which uses the knowledge of multiple beam directions at the gNB antenna side together with the UE measurements (ex. RSRP) of the different beams.

Round trip Time (RTT)
Principle: In LTE, the RTT procedure can be performed using the timing advance procedure for E-CID. This implies the use of time difference measurement to compute the distance between the two ranging devices. Three distance measurements are needed between the UE and different gNBs in order to apply unambiguous RTT-based positioning. However in LTE, the time difference measurements are restricted to the serving eNB and cannot be performed simultaneously for multiple eNB. In NR, the procedure will highly likely be introduced to support Round-trip time (RTT) with one or more neighbouring gNBs for NR DL and UL positioning.

Compared to TDOA, RTT has the advantage that no tight synchronization is required between involved positioning infrastructure. This comes at the expense of extra signaling when compared to TDOA where one-way transmission is sufficient to find the device position.

6.4.2 Features with impact on positioning performance

Each positioning method has its own pros and cons. In the list below you can find a short overview of the features that have an actual impact on the positioning performance.

Signal Bandwidth - A wider bandwidth offers higher TOA resolution and helps to resolve multipath inaccuracies in certain scenarios. Improves positioning accuracy especially for time-based systems.

Network Synchronization – for ambitious time-based positioning systems the synchronization between gNBs has to be in the order of nanoseconds or even a fraction of a nanosecond.

Multipath Propagation – the separation or identification of the multipath and the Line of Sight (LOS) path plays a central role for the performance of positioning systems

Antenna Imperfections – many AoA estimations take the antenna geometry into account. This may cause deterioration in performance if the location of the antenna elements is not precisely known.

Movement – principles that exploit measurements over a long period of time are sensitive to movement e.g. AoAchange.

Deployment density – increases LOS probability as the distance between the multiple TRPs and the target UE decreases. Provides more redundancy in terms of the TRPs used for determining a UE position.

6.4.3 Positioning in NR

The technical solutions proposed for 5G NR positioning are covering different architectures and measurement principles, also covering sensor fusion, etc. In the first step, the target performance is defined as horizontal positioning accuracy of < 3 m indoors. The 3GPP 5G radio-technology provides enhanced location capabilities which will be specified in Release 16, focusing on advanced capabilities like signal design, antenna configuration or network coordination [10][11]. The implementations of highly precise positioning functions are expected from Release 17 onwards (Standard available 2022 approximately). Figure **24** shows expected positioning performance of NR.

There is also an extensive range of parameters to be taken into account (e.g. carrier frequencies, bandwidth and antenna configuration). NR enables a positioning reference signal to be configured with a 100 MHz bandwidth in FR1 and 400 MHz bandwidth in FR2, which provides a better time resolution for TOA estimation and hint also towards higher performance targets. Also the possibility to combine multiple methods (hybrid positioning/sensor fusion) bears huge potential to achieve higher performance values. The fact that indoor dense deployments are foreseen for 5G supports also the idea that high performance positioning is realistic. Also the recent advances in massive antenna systems (massive MIMO) can provide additional information to enable more accurate localization by exploiting spatial and angular domains of the propagation channel in combination with time measurements.

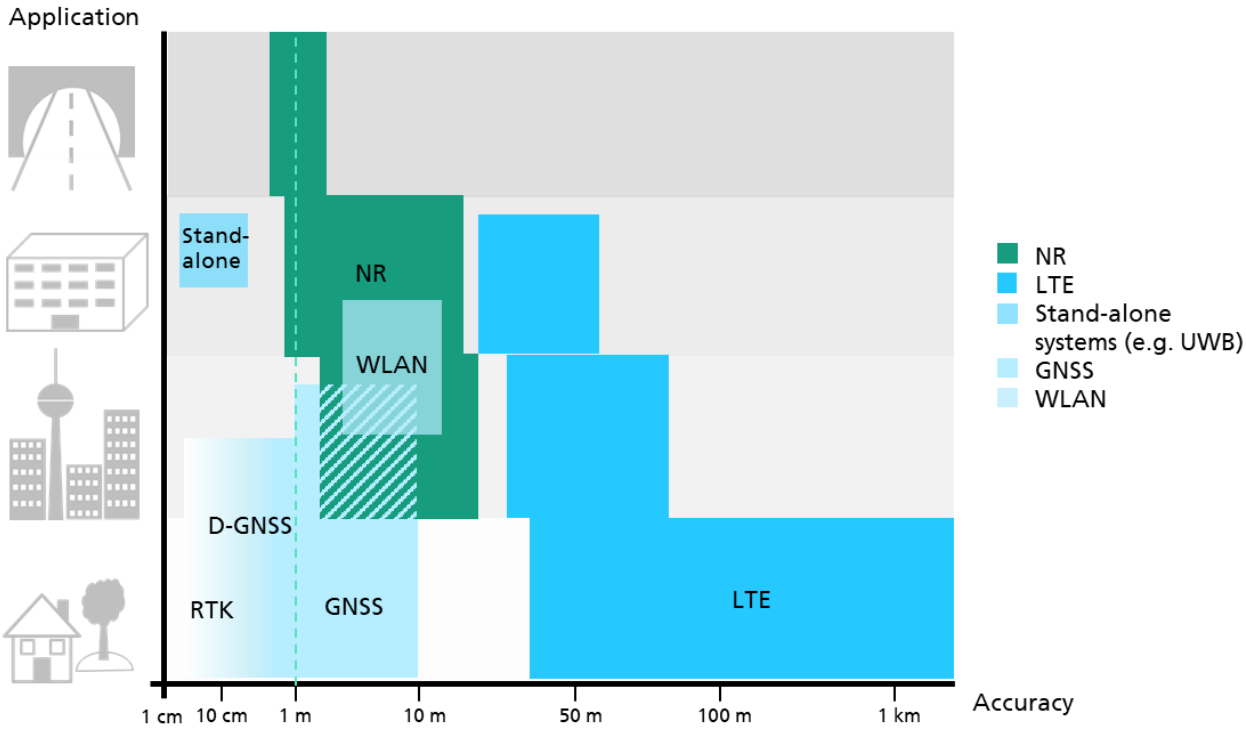


Figure 24 Expected Positioning Performance of 5G NR

For the support of high performance positioning some fundamental adjustments to mobile network design and layout are necessary like optimized antenna deployment for positioning or very dense deployments. A prominent example is the possibility of dense indoor deployments in support of industrial indoor use cases. Apart from density, the actual selected spatial location of distributed antennas matters for positioning. Therefore, increased positioning performance for specialized commercial use cases will be only available in selected areas. The high performance positioning with 5G NR for campus applications perfectly complements with the well-established Global Navigation Satellite Systems (GNSS) for outdoor scenarios.

6.4.4 NR positioning support for GNSS

In open sky scenarios standalone GNSS solutions can deliver positions with an accuracy of well below 10 meters. The new wide band Galileo signals (E5AltBOC with 54 MHz bandwidth) reach precision of around 10 cm. Assisted GNSS (A-GNSS**)** supports the receiver with information about orbital parameters and the availability of satellites and systems. The initialisation time is therefore reduced from about 30 s to 1-2 s. Differential GNSS (DGNSS**)** is correction data for the range measurements to each satellite which raises the accuracy to submeter level and mitigates errors caused by the atmosphere. The Real Time Kinematic (RTK) method substracts the user measurement from measurements of a nearby base station in order to compensate all errors coming from the system (satellite clock, satellite position, …), and atmosphere**.** With the resulting error free measurements the receiver can very precisely calculate the distance vector to the base station, of which the position is well known. Additionally the carrier phase ambiguities can be solved after 30 s for single frequency measurements and below 10 s for dual frequency. Therefore for RTK, carrier and code measurements of a base station and its coordinates have to be transferred to the receiver.

Standard RTK provides cm-level positioning accuracy in an area of 10 km around the base station. New correction methods use state space representation (SSR) of the correction data. In SSR the error is split up into the different error sources. Data vectors which represent these error sources are transferred. The advantage of the latter is that SSR can be used in wide areas. The area is only dependant on the amount of data that has to be communicated for the corrections.

6.5 Throughput

Throughput is one of the most important requirements in 5G, and also for some of the URLLC applications in vertical domains such as AR/VR, real time video (1080P) for human control. For instance, user experienced throughput needs 1 Gbps for AR and MR cloud gaming and 40Mbps in uplink for remote control.

In report [12], the following performance metrics related to throughput are defined.

- **Peak spectral efficiency**
  The maximum data rate under ideal conditions normalized by channel bandwidth (in bit/s/Hz), where the maximum data rate is the received data bits assuming error-free conditions assignable to a single mobile station, when all assignable radio resources for the corresponding link direction are utilized (i.e. excluding radio resources that are used for physical layer synchronization, reference signals or pilots, guard bands and guard times).
- **Peak data rate**
  The maximum achievable data rate under ideal conditions (in bit/s), which is the received data bits assuming error-free conditions assignable to a single mobile station, when all assignable radio resources for the corresponding link direction are utilized (i.e. excluding radio resources that are used for physical layer synchronization, reference signals or pilots, guard bands and guard times).
- **Average spectral efficiency**
  The aggregate throughput of all users (the number of correctly received bits, i.e. the number of bits contained in the SDUs delivered to Layer 3, over a certain period of time) divided by the channel bandwidth of a specific band divided by the number of TRxPs and is measured in bit/s/Hz/TRxP.
- **User experienced data rate**
  The 5% point of the Cumulative Distribution Function (CDF) of the user throughput. User throughput (during active time) is defined as the number of correctly received bits, i.e. the number of bits contained in the Service Data Units (SDUs) delivered to Layer 3, over a certain period of time.
- **Area traffic capacity**
  The total traffic throughput served per geographic area (in Mbit/s/m2). The throughput is the number of correctly received bits, i.e. the number of bits contained in the SDUs delivered to Layer 3, over a certain period of time.

Throughput is enabled by:
Massive MIMO
Massive MIMO allows more antennas to support multiple data streams and beamforming. This helps to improve the throughput and spectral efficiency. A closed loop Demodulation Reference Signal (DMRS) based spatial multiplexing is supported for Physical Downlink Shared Channel (PDSCH), and up to 8 orthogonal DL DMRS ports per UE are supported for SU-MIMO, that means 8 data streams could be transmitted simultaneously. Beamforming based on massive MIMO could also enhance received power and reduce interference in of multi-user systems, which will improve average spectral efficience and user experienced data rate.

Large bandwidth
Large bandwidth means more radio resource could be utilized to improve throughput. We can consider typical bandwidth in 5G system upto 100 MHz in FR1 and 400 MHz in FR2. The maximum transmission bandwidth configuration N_{RB} for each UE channel bandwidth and subcarrier spacing is specified in Table 3 [12].

Table 3 Maximum transmission bandwidth configuration NRB for FR1

SCS (kHz)	5MHz	10MHz	15MHz	20 MHz	25 MHz	30 MHz	40 MHz	50MHz	60 MHz	80 MHz	90 MHz	100 MHz
	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}
15	25	52	79	106	133	160	216	270	N/A	N/A	N/A	N/A

30	11	24	38	51	65	78	106	133	162	217	245	273
60	N/A	11	18	24	31	38	51	65	79	107	121	135

Table 4 Maximum transmission bandwidth configuration NRB for FR2

SCS (kHz)	50 MHz	100 MHz	200 MHz	400 MHz
	N_{RB}	N_{RB}	N_{RB}	N_{RB}
60	66	132	264	N.A
120	32	66	132	264

High order modulation

High order modulation could increase spectral efficiency. For NR, both PDSCH and PUSCH support 256QAM. With above techniques, Table 5 provides the downlink/uplink peak data rate evaluation results for the specific Component Carrier (CC) bandwidth [13]. From the table, it is observed that more than 1Gpbs peak data rate for both downlink and uplink could be achieved by NR. However user experienced data rate is also highly dependent on deployment scenario and base station density. For URLLC use case, considering that it is more challenged to satisfy requirement on reliability, latency and throughput simultaneously, the density of base station should be carefully considered for URLLC scenario.

Table 5 NR peak data rate

Duplexing	SCS [kHz]		Per CC BW (MHz)	DL Peak data rate per CC (Gbit/s)	UL Peak data rate per CC (Gbit/s)
FDD	FR1	15	50	2.31~2.41	1.12~1.18
		30	100	4.67~4.89	2.28~2.39
		60	100	4.62~4.82	2.27~2.38
TDD (DDDSU)	FR1	15	50	1.81	1.12~1.18
		30	100	3.68	2.28~2.39
		60	100	3.62	2.27~2.38
	FR2 (N_{layer}=6)	60	200	5.33	
		120	400	10.7	
TDD (DSUUD, S slot= 11DL:2GP:2UL)	FR1	15	50	1.32	
		30	100	2.69	1.06
		60	100	2.64	1.05
	FR2 (N_{layer}=6)	60	200	3.86	1.91
		120	400	7.81	3.85
TDD (DSUUD, S slot= 6DL:2GP:6UL)	FR1	15	50	1.13	
		30	100	2.30	1.05
		60	100	2.26	1.04
	FR2 (N_{layer}=8)	60	200	4.38	2.02
		120	400	8.76	4.04

6.6 Network Slicing

A network slice can be understood as an E2Elogical network that runs on single shared physical infrastructure. The technology enabling network slicing is transparent to business customers [18]. Considering network slices for the URLLC communication services, there are different mechanism to enable network slice as shown in Figure 25.

For Public Land Mobile Network (PLMN) based network slicing method, RAN sharing is applied which allows base station to announce multiple PLMN IDs. Network should assure the traffic belonging to different PLMN IDs is routed properly. UE select PLMNs based on normal procedure of network selection.

For Access Points Names (APN) based network slicing mechanism, UE can be configured with multiple APNs. Compared to first method, only one PLMN ID is announced by RAN, however user plane traffic is routed to respective CN based on the identified APN.

From Release-13, Dedicated Core Network (DECOR) was introduced which allowed selection of network slice by providing a capability of having a Dedicated Core Network) within one PLMN-ID for as different type of end-devices. To enable identification of which UE belongs to corresponding DCN a new subscription parameter is introduced called “UE Usage Type”. It is stored in HSS along with other subscription information of UE.

Enhanced DECOR (eDECOR) further reduced signalling between UE and DCN by addition of assistance information in UE. 5G slicing (also shown in Figure 26) and the corresponding selection mechanism allow UE to connect to multiple NW slices at the same time. Network Slicing Selection Assistance Information (NSSAI) is introduced to assist slice selection. The NSSAI consists of list of Single (S-NSSAI). Each S-NNSAI contains a Slice Service Type (SST). SST refers to expected Network Slicing behaviour in terms of features and services [3].

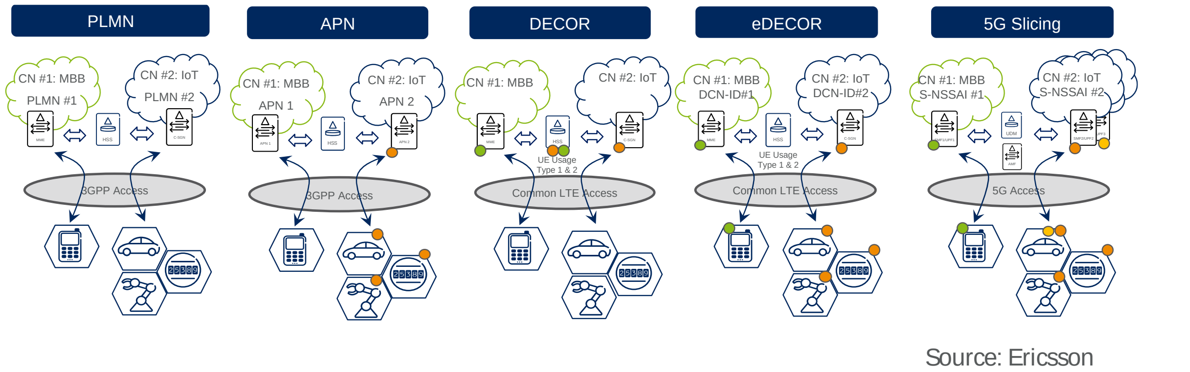


Figure 25: Mechanism to enable Network Slicing [22]

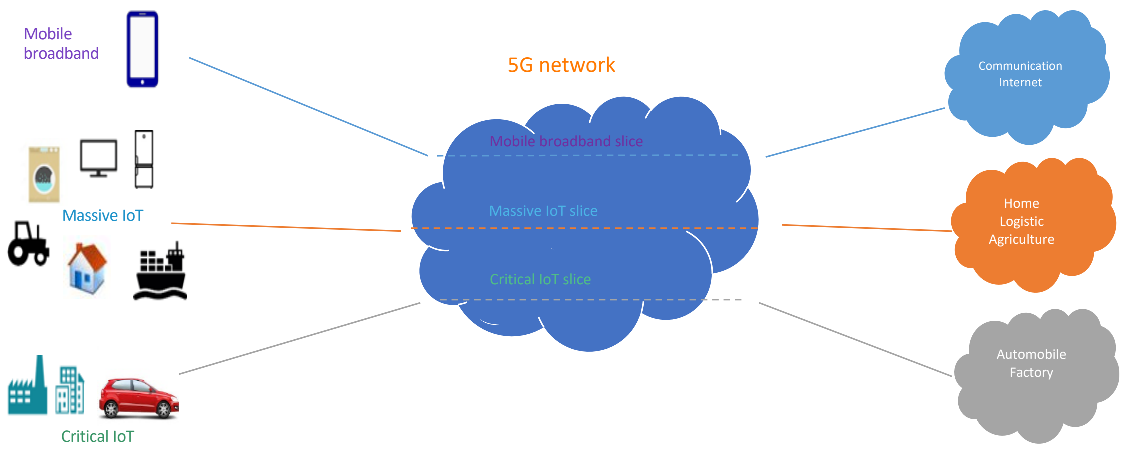


Figure 26: 5G network Slices

Opportunity for Verticals with NPN and URLLC Network Slice

With 5G and Network Slicing functionality there will be an opportunity for verticals using NPN to provide an efficient URLLC service by deploying Edge Computing functionality and high reliability with redundant transmission in user plane as described in sections above 6.2.1 and 6.1.5

6.7 Non-public networks

According to the definition in [14], "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, utilising both virtual and physical elements". 3GPP distinguishes two type of non-public network (NPN) deployments: Stand-alone Non-Public Network (SNPN), and Public network integrated NPN. The former is operated by an NPN operator and does not rely on network functions provided by PLMN, while the latter is deployed with the support of a PLMN. 3GPP in Rel-16 introduced new functionality to support NPNs, in order to fulfill requirements listed in [14]. Architectural aspects are documented in [3].

An SNPN is identified by the combination of a PLMN ID and a Network identifier (NID). The NG-RAN node providing access to SNPNs broadcast the following information: one or more PLMN ids, list of NIDs (per PLMN id) identifying the SNPNs the NG-RAN provides access to, and another optional information (human-readable network name, and information to prevent UEs not supporting SNPNs accessing the cell. It is possible for a UE that has registered to an SNPN to perform another registration with a PLMN, via the SNPN user plane, to be able to access PLMN services via stand-alone non-public networks. Also, a UE that has registered with a PLMN may perform another registration via the PLMN user plane, to be able to access SNPN services via PLMN.

Public network integrated NPNs can be enabled 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 used to prevent UEs which are not allowed to access the NPN via the associated cells, from selecting and accessing the associated cells.

7 5G URLLC REFERENCE ARCHITECTURE

5G system architecture comes in two flavors non-standalone and standalone. Considering standalone architecture both user plane and control plane signalling is handled by gNB compared to non-standalone mode. From requirement of URLLC communication services it is observed that standalone reference architecture is more suitable to support deployment for URLLC communication services.

Further this section describes deployment architectures based on the public and Non-Public Network (NPN) services. Two main use cases covering wide area and local area deployment are investigated based on the various (NPN) deployment options. Finally, evaluation of the such deployment options concerning URLLC use case against various deployment attributes is elaborated.

7.1 Public and non-public network services

Mobile network services are commonly realised over a network infrastructure offered to the general public, however, the 3GPP 5G standards also allow ways to deploy a network to be used as Non-Public Network (NPN). In contrast to a Public Network (PN), a NPN is a network deployed to provide communication services for a clearly defined group of users inside an organisation or group of organisations [15]. A 5G NPN can be deployed within organisation premises e.g. campus or factory. Generally, a NPN can be deployed in conjunction with PNs, where the PN itself is not limited to serve the NPN(s), but they also operate as a regular public network to its subscribers. This form of NPN deployment is also known as integrated *NPN*, whereas standalone NPNs without interaction with the PN are also known as an isolated NPN. These deployment options, on a very high-level, permit NPNs to be completely isolated networks or partially integrated with PNs. While the isolated deployment is straightforward, i.e. a NPN has all necessary components such as a 5G Radio Access Network (RAN) and CN running locally, the integrated deployment has many attributes to be considered such as frequency allocation, e.g. in cases the RAN is to be shared, operation and management, security, trust and isolation, device connectivity when it comes to service coverage, reachability and continuity (e.g. roaming), QoS, economic feasibility, etc. We summarise these attributes in the following categories:

Table 6: NPN attributes for deployment choice

Deployment aspects	**Comment**
Device connectivity, i.e. coverage, reachability, service continuity (e.g. roaming)	A NPN may have extended coverage and have ways to provide service continuity, if devices are allowed to automatically connect to a PN, i.e select a PN ID. This may be the case when the coverage of the NPN needs to be extended beyond a local campus area, e.g. transport and logistics between buildings. Whether devices should be configured to be able to select the PN ID when out of coverage of the NPN depends on verticals customers' requirements.
Isolation	Service and/or network isolation levels vary depending on if the NPN control- or data-plane data should not leave the NPN, e.g. local device authentication and registration or local user plane. Also, any other network resource and/or service is only available locally in an standalone NPN.
Quality of Service (QoS)	QoS in terms of latency and jitter can vary depending on the deployment. For example, if the NPN's user plane is sharing a common infrastructure together with the PN, tight latency may be harder control for the NPN.
Operation and Management	There are differences in Operation and Management (O&M) between a standalone NPN and a NPN that is intergrated with a PN. For a standalone NPN the operator of the network has full control and responsibility of O&M and access to all related management data. For an integrated NPN, the PN and the NPN share at least some part of a common infrastructure. The NPN may further be integrated and connected with other private infrastructure that has its own O&M procedures and ownership, e.g. when a 5G NPN is included as a wireless component into the industrial system of a factory. For PN integrated NPN, control of O&M aspects should be controlled by operators considering security of PN. For the isolated NPN, the O&M is within the responsibility of the NPN operator.
Security and Trust	Depending on the trust level between NPN and PN, or legal frameworks in place such as GDPR, access to data can be restricted to the NPN. When it comes to access to the network, USIM and non-USIM methods are available and can be implemented as part of the NPN or fully outsourced to the PN. In the latter case, the NPN users will be also PN subscribers.

7.1.1 Non-public network deployment options

The 5G-enabled PN also can provide a wide range of communication services including enhanced mobile broadband (eMBB), massive machine type communication (mMTC) and URLLC. The URLLC use cases as discussed in chapter 5 can be realised over a PN or NPN type of deployment, which highly depend on different factors as mentioned in Section 7.1. Further detail aspect on local connectivity and wide area connectivity solution using 5G system for the industries is elaborated in [17]. The URLLC use cases and their requirements will likely drive the deployment choice attributes from Table **6**. Between fully isolated and integrated NPN deployment options, four possibilities are identified and summarised in Table **7** based on [15].

Depending upon the Radio Access Network (RAN) deployment various low latency enablers as describe in the section 6.1 can be enabled, considering MOCN and MORAN deployment proper radio planning is required to facilitate such features in real world deployment. Further, to enable time synchronisaiton over 5G system, integrated 5G-TSN functionalities must be part of CN and RAN. In particular, NW-TT would be part of UPF, which depending upon the NPN would be part of local breakout or shared user plane function, and UE must support DS-TT functionality for accurate time reference delivery to URLLC services. Postioning function to support high accuracy and low latency positioning services can be deployed above core control plane [17].

Table 7 NPN deployment options

Deployment option No.	NPN deployment options	Characteristic/Details	Key observations
NPN 1	NPN hosted by the PN	Both NPN traffic and PLMN traffic are off-premises but treated differently through virtualization of network functions, APNs, slicing IDs, etc.	NPN data traffic and control traffic are off-premises NPN devices are PLMN subscribers Global connectivity enabled by operator roaming arrangements
NPN 2	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. Segregation can be realized via usage of different APNs, or slicing IDs, etc.	NPN data traffic remains local and NPN control traffic is off-premises NPN devices are PLMN subscribers Seamless roaming is possible for NPN devices aiming to connect to a NPN service
NPN 3	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 segregrated, also data flows remains local. It can be realized by: • Multi-Operator Core Network (MOCN), where two or more entities are sharing eNodeB/ gNodeB and spectrum • Multi-Operator RAN (MORAN), where two or more entities are sharing eNodeB/gNodeB with non-shared spectrum	NPN data traffic and NPN control traffic remain local Dedicated spectrum is required for deployment considering MORAN Spectrum can be shared when MOCN is considered for deployment
NPN 4	Standalone	All NPN functionalities are on-premises. NPN is a fully separate physical network from the PN with dedicated NPN ID. Howerver, dual subscription with NPN and PLMN is possible. Access to PLMN services can be realized via an optional firewall connection and roaming agreement.	Requires deployment of all 3GPP functions and URLLC technical enablers on premises. NPN data and control traffic will remain local Dedicated or leased spectrum for deployment is required

7.1.2 Public Network options

Below, we briefly mention some of the main technical enablers for the PN deployments mentioned throughout Section 7.1.

MORAN/MOCN: MORAN and MOCN are two different network sharing techniques, more specifically, they are two RAN sharing techniques. With MOCN, two or more entities can share both RAN equipment and spectrum. Briefly, MORAN is a variant of MOCN, where the spectrum is not shared between the entities and own spectrum holdings are required.

DNN (equivalent to APN): One of the 5G CN main tasks is to connect the RAN to Data Networks. Here, the DNN is the service that identifies a certain Data Network. DNN is equivalent to the access point name (APN) used in the Evolved Packet System. Further, the DNN may be used as a parameter in the SMF and in the UPF selection procedure. When setting up a PDU session, the DNN is either sent by the UE itself when requesting a PDU session establishment, or it is derived from the UE's subscription parameters. With the usage of DNNs and separate Data Networks, it is possible to support URLLC use cases by isolating URLLC traffic from other, less latency restrictive traffic. A UE may be able to access public services and non-public services in the same time by setting up two parallel PDU sessions towards two different Data Networks.

Network slicing: The GSMA definition of network slicing is the following [18] "From a mobile operator's point of view, a network slice is an independent E2Elogical network that runs on a shared physical infrastructure, capable of providing a negotiated service quality. The technology enabling network slicing is transparent to business customers. A network slice could span across multiple parts of the network (e.g. terminal, access network, CN and transport network) and could also be deployed across multiple operators. A network slice comprises dedicated and/or shared resources, e.g. in terms of processing power, storage, and bandwidth and has isolation from the other network slices." Using slicing, it is possible to provide logical networks with different characteristics towards different use cases and/or different customers.

5G QoS Framework: In the 5GS, the QoS model is based on QoS Flows. Each PDU Session may consist of one or multiple QoS Flows. Every packet is marked by a QoS Flow Identifier (QFI), indicating the QoS flow the packet belongs to. Every packet belonging to the same QoS Flow will receive the same traffic forwarding treatment. A QoS Flow is associated with a QoS profile, that contains QoS parameters, where 5QI (5G QoS Identifier) and Allocation and Retention Priority (ARP) are mandatory for all flows. Flows may be either guaranteed bitrate (GBR) or non-guaranteed bitrate (non-GBR) flows. In the GBR case, maximum and guaranteed flow bitrate are additional parameters. The 5QI is a scalar that is referencing QoS characteristics, and standardized values have one-to-one mapping to a combination of QoS characteristics: Resource Type, Priority Level, Packet Delay Budget, Packet Error Rate, Averaging Window (when applicable), Maximum Data Burst Volume (when applicable). To better support URLLC services, 3GPP defined the Delay-critical GBR resource type.

7.2 5G URLLC local reference deployment architecture for Factory of Future

By introducing 5G in the current ongoing digital transformation of smart manufacturing often referred to as Factories of the Future, higher production efficiency and high customization can be achieved. As such, the 5G network deployment can be configured in different ways, depending upon the URLLC use cases that can be realized by the aforementioned technical enablers. Figure 27 provides a reference architecture for the deployment of 5G for smart manufacturing including many components from section 7.1 and with some of the technical enablers and targeted URLLC use cases illustrated as "Industrial Devices". Future smart manufacturing is envisoned to migrate from a hierarchical design with a diversity of incompatible fieldbus technologies to a fully connected factory design based on IEEE 802.1 standard ethernet based communication. TSN – as extension to standard fixed ethernet, provides the required level of deterministic, URLLC for critical production processes [20]. At the same time, 3GPP is working towards evolution of 5G to support TSN integration. Integrated 5G and TSN can provide deterministic performance in terms of requirements from end industrial devices to industrial controllers. Reference [20] further discusses integration aspects of 5G and TSN for industrial automation. 5G-SMART a European Union project funded under

Horizon 2020 work programme is investigating different 5G network architectures for smart manufacturing use case, and in addition, new technical features such as time synchronisaiton and positioning are investigated [18].

In the Figure 27 below, a URLLC reference deployment architecture is shown, including end industrial devices such as TSN end devices, which are connected to a 5G network via a UE with Device Side - TSN Translator functionality (DS-TT) and connect via the 5G network to an Industrial controller that is connected to the UPF with a Network TSN Translator functionality. The 5G system has to provide support to enable configuration and management of the industrial networks. Relevant TSN features for smart manufacturing such as FRER, Time Synchronisation can be supported by the 5G system.

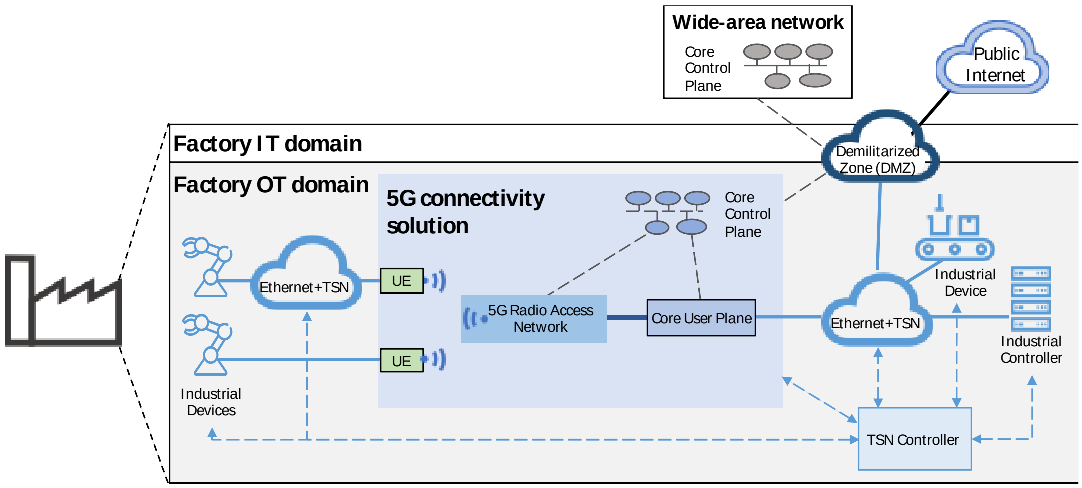
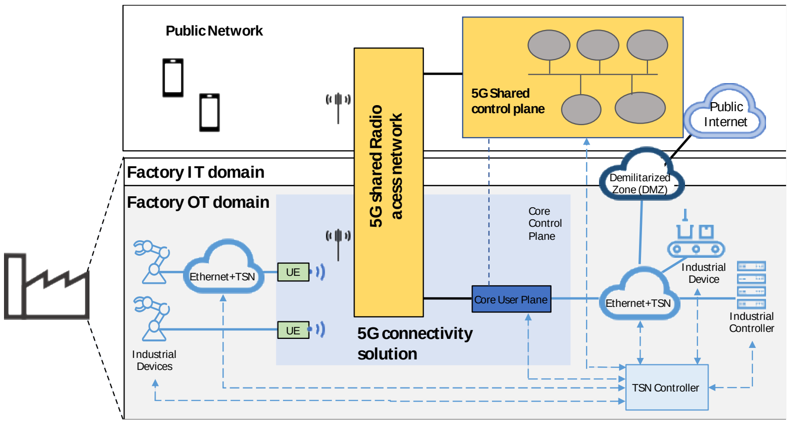
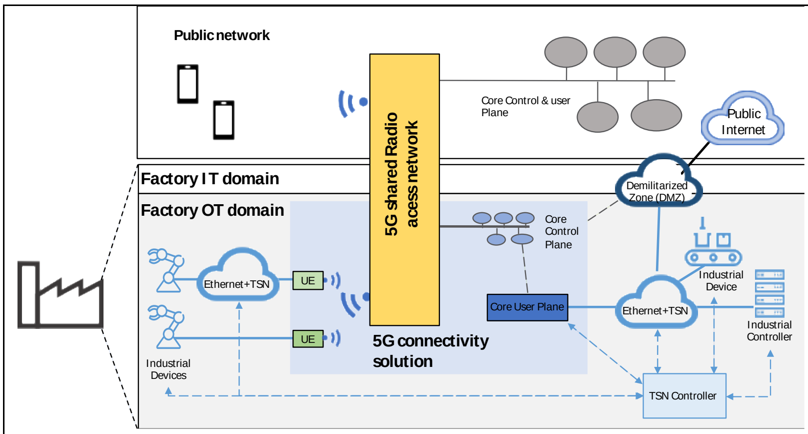


Figure 27 5G URLLC reference architecture (based on the report[16])

7.2.1 Deployment architecture consideration

Different PN and NPN configurations can be applied to realise the deployment of the reference architecture for a factory of the future. Table 8, further visualizes and describes which key features are applied to address the requirements. Based on factors including but not limiting to geographical location, spectrum availability, deployment attributes requirements, a suitable deployment can be configured.

5G URLLC architecture for smart manufacturing	Key features related to the requirement
 NPN 2 from Table **7** (Shared RAN and control plane)	- Cloud capability on-premise and local user data breakout to support deterministic latency requirement - Segregation realised 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 PN and the private OT management system is required - Local survivability may not be achieved if connection to the PN core network is lost - Concerning data privacy, user data does not leave the factory premises however the users' subscription data, as well as connectivity related session and mobility information is present in the operator's premises.

 NPN 3 from Table **7** (Shared RAN)	- Cloud capability on-premise and local user data breakout, enabling deterministic low latency - Share RAN with the PN enabled by MOCN or MORAN. Proper local radio planning and dimensioning is required to fullfill URLLC requirements, and in case of MORAN licensed spectrum is required with dedicated radio equipment (e.g. antenna). - Coordination of RAN deployment and operation between the PN (mainly off-premise) and NPN (on premise) possible - 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 - 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.

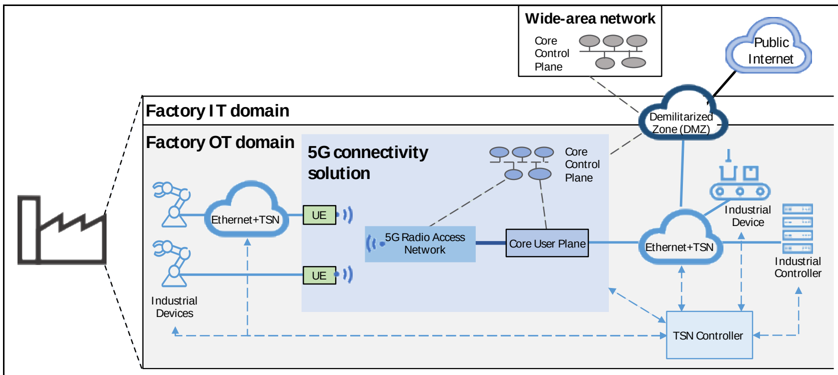
 NPN 4 from Table 7 (Standalone private network)	- Cloud capability on-premise and local user data communication, enabling deterministic low latency - Dedicated RAN deployed on premise - Separate coordination of RAN deployments and operation of the PN (off-premise) and NPN (on premise) is needed for coexistence - 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 - 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.

Table 8 5G URLLC solution architecture options for Factory of the future use case

7.3 5G URLLC wide-area reference deployment architecture for the Smart Grid

With ongoing increase of the renewable energy sources, leading to a bidirectional flow of energy, transformation in energy networks can be observed. In particular, smart distributed power distribution automation and fault management for distributed electricity generation in the smart grid are considered to be important 5G use case.
5G deployment is considered to be a cost effective alternative solution compared to optical fibre buildout to enable power grid services such as Differential Protection (DP) and FISR system [1] [23]. Considering the nature of such use cases with its geographic spread, a wide-area 5G system deployment is assumed as reference deployment as depicted in Figure 28 . The main observations for the given deployment are also described along with the deployment architecture. Further one can configure wide area deployment in two main options as shown in Table 9.

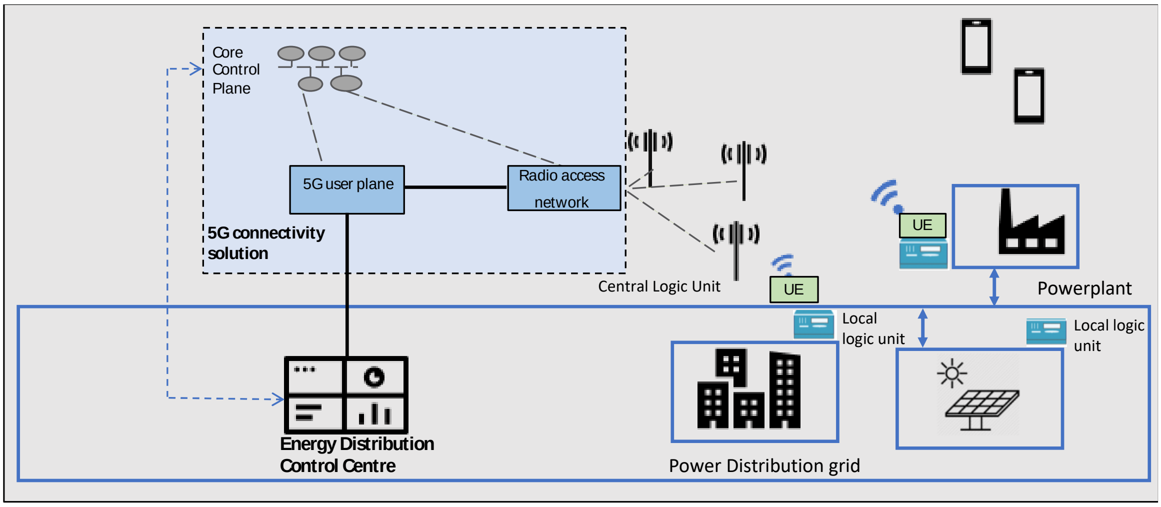
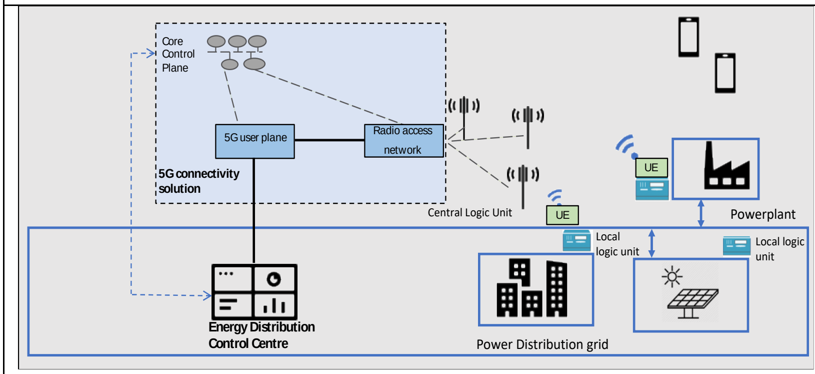


Figure 28 5G URLLC deployment architecture for Smart Grid

7.3.1 Deployment architecture consideration

The energy distribution network usually covers a wide area. For smart distribution automation in the energy distribution network, the zones of coordinated automation are regional and typically require an E2Ecommunication range of around 10 km [1]. The connection density is around 10~100 smart grid devices per square kilometre. For the FISR grid use case, the typical E2Ecommunication range can reach up to 100 kilometer. Considering factors such as geographical area of energy distribution network and connection density of end devices, standalone network deployment seems to be infeasible from the resource and cost perspective.

<table>
<tr><th>5G URLLC deployment architecture for the Smart Grid</th><th>Key features related to the requirement</th></tr>
<tr><td>


NPN 1 from Table 7 (NPN hosted by PN)
</td><td>

- Smart Grid related data traffic and public network traffic is treated differently by connecting it to different network entities, like for example, different virtualised network functions in a virtualisation network realization based on a distributed telco cloud infrastructure.
- Radio network planning is a part of public network deployment which need to consider requirement of the Smart Grid use cases
- Network slicing for dedicated URLLC services (for e.g. Differential Protection, FISR system) can be deployed across different parts of the network including shared CN and shared radio access network to fullfill stringent requirement on E2Elatency.
- Interaction between 5G system and the energy distribution centre is required concerning operation and management functions
- High network availability of the 5G network is required due to the critical nature of the use cases and the potential impact that a failure of the energy distribution may incur. High availability can be provided by e.g. geo-redundancy which can be realized based on virtual network functions in a distributed telco cloud.
- Relating to data privacy, the PN and NPN networks both may share the same physical but logically separated database
- Accurate time reference delivery of absolute time is needed to enable time synchronisation among end device like e.g. measurement units
</td></tr>
</table>

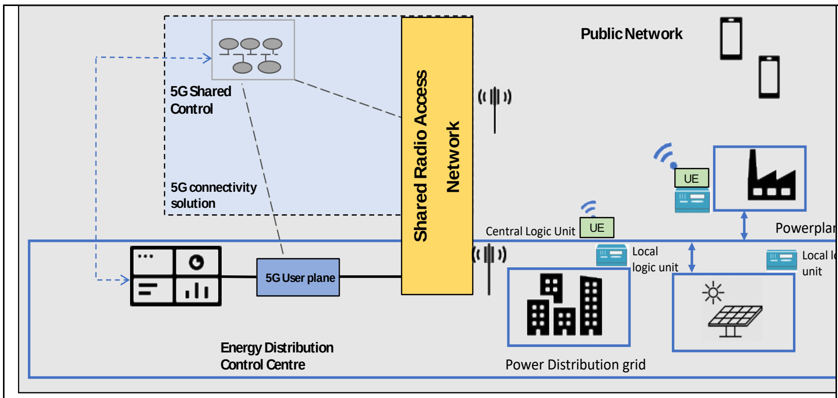
 NPN 2 from Table **7** (Shared RAN and control plane)	- Smart grid applications (Differential Protection, FISR system) can be deployed with local breakout, e.g. with an distributed cloud implementation. The network function instances can be geographically dispersed and can be realised in proximity of energy operator premises. This allows for short E2Ecommunication ranges and latencies between Phasor Measurement Units (PMU) and the energy distribution control centre (in the order of ˜2km) - Different DNN can be configured for the different services or different slice instances may be used - Smart Grid services can be assigned to high priority QoS flows that have priority over other traffic. Application functions would be required for converting the desired requirements from smart grid application to QoS flow characteristics - Accurate time reference delivery of absolute time is needed to enable time synchronisation among end device like e.g. measurement units

Table 9 URLLC deployment architecture options for smart grid

7.3.2 5G URLLC wide-area reference deployment architecture for 360º panoramic VR view video broadcasting

360º panoramic VR view services will be the next generation of video broadcasting services bring in revolutionary visual experience to the end users. To enable the immersive experience of users, quality of video should at least reach 8K while the motion-to-photon (MTP) response time should be less than 20ms. A 5G network is the key enabler for such an bandwidth and latency demanding application. In the meantime, 5G can also enable the mobility of video uploading from the camera side in event-based broadcasting. Since the video will be broadcasting to massive group of end users geographically separated, wide area 5G system deployment is required. The deployment architecture is shown in Figure 29. Table 10 provides details on feasible deployment options.

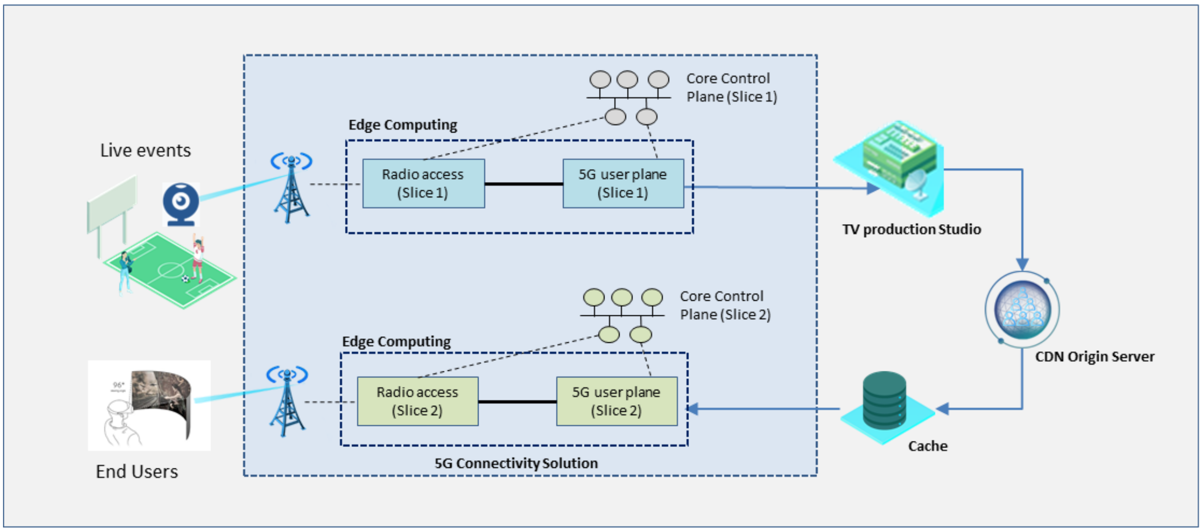
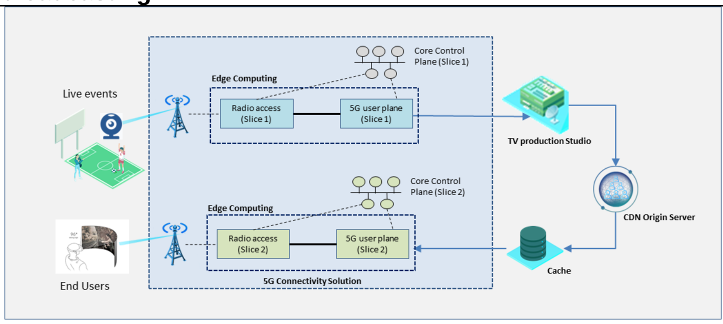
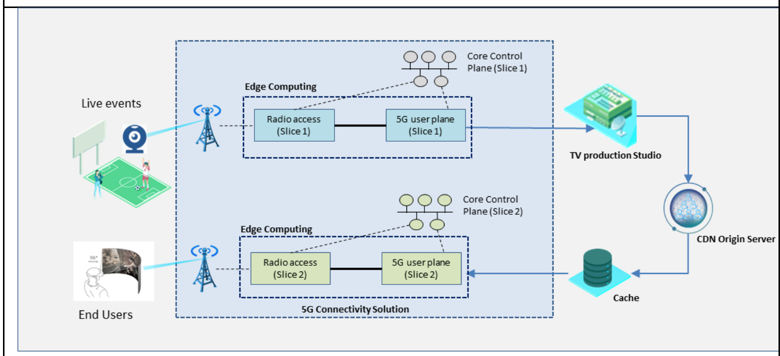


Figure 29 5G URLLC deployment architecture for 360 panoramic VR view video broadcasting

Table 10 5G URLLC deployment architecture for the 360 VR video broadcasting

5G URLLC deployment architecture for the 360 VR video broadcasting	**Key features related to the requirement**
 PN from Table **7**	- For live video broadcasting service, the deployment option should be PN only due to the massive group of end users. - Video traffic is divided into two paths, upstream from the 360 cameras and downstream to the end users. Dedicated network slice for upstream and downstream is proposed to ensure service performance with isolation of dedicated resources. This is in particular beneficial, if the location of video capturing is overlapping with the location of broadcasting. - Edge Computing should be deployed for the nation wide network to reduce the latency. - The cache server of CDN should be connected to the nearest UPF to optimize the performance. - To support higher resolution of video (e.g. 8K, 12K, etc.), it is not feasible to transmit full 360 video data. Instead, transmission of the *field of view* (FOV) will probably be the direction to go. This involves interaction between devices, cache server and CDN interworking.
 NPN hosted by PN (NPN 1 wth shared control plane in Table **7**)	- 360º view VR video broadcasting on the other hand can be deployed as Campus based application like infrastructure monitoring. In this case, NPN deployment can be a choice. - NPN can be hosted in PN, Network slicing and edge computing can be realised to make sure the radio access and user plane remains non-sharing but the control plane can still be shared with PN network. - Dedicated network slices are assigned for the Campus. - Different DNNcan be configured for the specialized application. - Like for a PN deployment option, upstream and downstream should be separated into two network slices to isolate the resources in radio access network as possible.

7.3.3 5G URLLC wide-area reference deployment architecture for VR Cloud Gaming

Cloud based VR based gaming is another bandwidth (300Mbps for 4K image) and latency (< 5ms) demanding application. In particular, the latency as gaming is a bi-directional interactive type of application. One of the challenges in VR game industry is the cost of a VR head-mounted display (HMD) is still too high and the size is too large for the mass market. Quality of image is limited by the computing power of HMD as well. A 5G network with its low latency and high bandwidth can enable the compute intensive rendering work migrating to cloud platform. This can dramatically help reducing the size and cost of the VR HMD. The deployment architecture is shown in Figure 30. Table 11 provide description of feasible deployment options for use case.

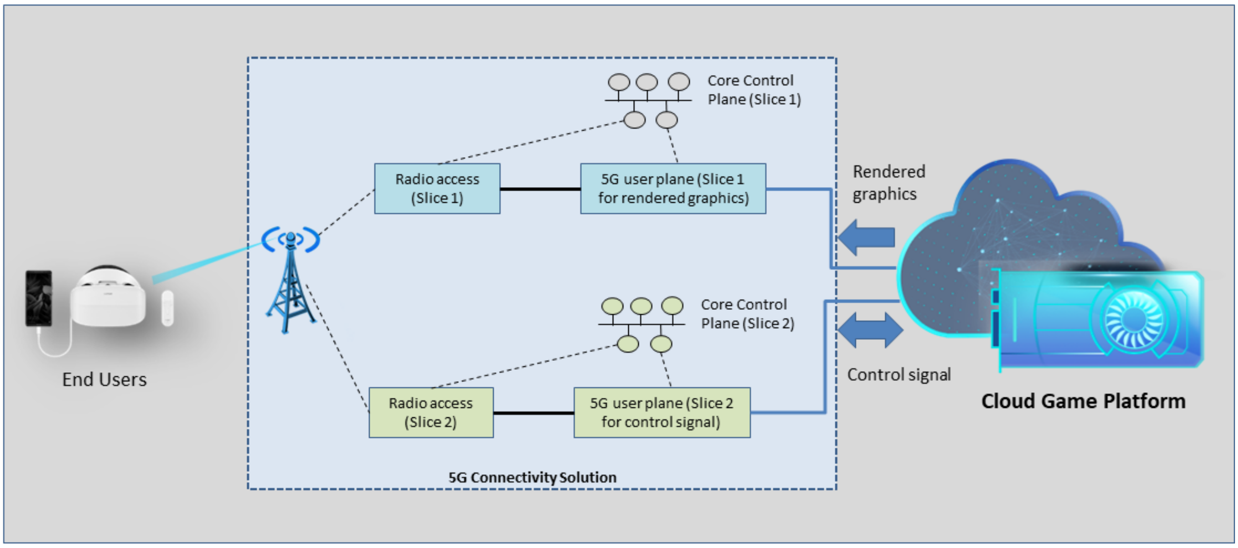


Figure 30 5G URLLC deployment architecture for VR Cloud Gaming

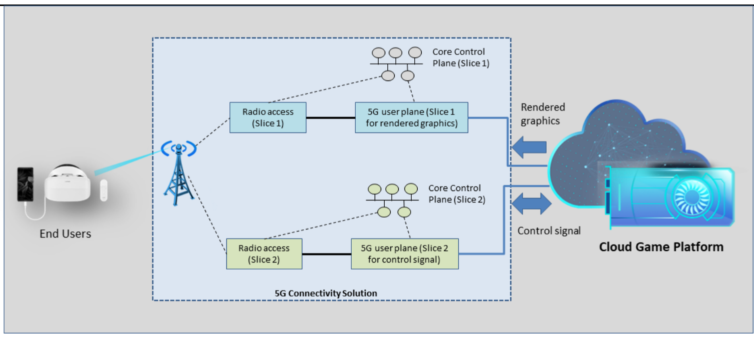
 Public Network deployment	- The deployment option should be PN only due to the massive group of end users. - The high-resolution graphics from the cloud platform should be isolated from the control signal flows between the client devices and the cloud platform due to the unique nature of traffic. So, at least two network slices with different architecture design are required. - The primary design objective for slice of images is to carry large amount of data. In other words, packet replication should not be applied while the slice should be eMBB one. - For control signal, the reliability is critical, packet replication should be applied to secure the delivery of control signal.

Table 11 5G URLLLC solution architecture for VR Cloud Gaming

7.4 5G URLLC architecture evaluation

The deployment attributes as described in Table 6 can be considered as basis for the architectural evaluation. The evaluation can be done on the basis of the degree of compliance. *High compliance* indicate that the URLLC communication service is fully supported with existing deployment architecture, *medium compliance* indicate that the URLLC communication service is supported under certain condition or with some adaption and *low compliance* indicate that the service is either not supported or only supported with significant adaptions. This evaluation is done considering reference [15] on non public network industrial scenarios. Table 12 shows potential mapping of the use case versus possible deployment 5G architecture.

Table 13 further provides insight to degree of compliance of the each usecase deployment architecture options, by considering deployments as described above in Table 7.

Table 12 Evaluation of the URLLC use case deployment architecture

Use Cases	**PN**	**NPN**			
		NPN 1 - NPN hosted by the PN	NPN 2 - Shared RAN and control plane	NPN 3 – shared RAN	NPN 4 Standalone
Factory of the future	X	X	X	X	X
Smart Grid	X	X	X	-	-
VR	X	X	X	-	-

Table 13 Evaluation of the URLLC use case deployment architecture

Use Cases	**Deployment attributes**	**NPN**			
		NPN 1 (NPN hosted by PN)	NPN 2 (Shared RAN and Control plane)	NPN 3 (Shared RAN)	NPN 4 (Standalone non-public network)
Factory of future	Global connecitivty	High	High	Low/High [1]	Low/High 1
	Service continuity	High	High	Low /Medium [2]	Low/Medium 2
	Latency & Availability	High/Medium/ Low [3]	High/Medium [4]	High/Medium 4	High [5]
	Acess to monitoring data and O&M functions	High/Medium [6]	High/Medium 6	High/Medium 6	High/Medium 6
	Data privacy through isolation	Medium	Medium	High	High
	Control and management privacy through isolation	High/Medium/ Low [7]	High/Medium [8]	High	High
	Flexibility in choice of security mechanisms	Low	Low	Low	High/Low [9]
	Global availability of security mechanism	High	High	High	Medium/High [10]

[1] Low would be when device can connect to NPN only and high would be when device has an optional public network subscription [15]

[2] Low is when device cannot connect to public network and is only connected to NPN and medium is optional connection to public network and public network subscription is available [15]

[3] As NPN data and network functions are external, based on URLLC service level requirement, degree of compliance can vary [15]

[4] High if traffic isolation is produced by dedicated resource allocation in shared RAN or dedicated RAN is used to enable physical isolation and traffic isolation for the user plane traffic, it can be medium, if QoS requirement is not taken in consideration for deployment of shared RAN [15]

[5] High as deployment architecture allows physical isolation of the network resources [15]

[6] High considering NPN operator has full access to its monitoring and O&M functions, medium when NPN operator has delayed access to network function and information[15]

[7] Depending upon public network implementation degree of compliancecan vary [15]

[8] High when control and management plane are exclusively utilized for NPN, medium if control and management functions are operated by single operator for both network [15]

[9] To accessing only NPN using USIM based methods, high degree of compliance may be realised, low when device has also public network services. Further it may be also necessary to allow lawful interception depending upon the regulatory imperatives [15]

[10] Medium when accessing only NPN, high when public network security mechanism that is supposed to be able to access public network [15]

Smart Grid	Global connecitivty	High	High	-	-
	Service continuity	High	High	-	-
	Latency & Availability	Depends on the placement of the UP functionalities in the public network	High	-	-
	Acess to monitoring data and O&M functions	Depends on the agreement between the public MNO and the grid operator	Depends on the agreement betwwen the public MNO and the grid operator	-	-
	Data privacy through isolation	High through logical isolation provided by E2E slicing in the public network	High through physical isolation of the UP in the CN and logical isolation provided by RAN sharing	-	-
	Control and management privacy through isolation	High through logical isolation provided by E2E slicing in the public network	High through logical isolation provided by slicing on the CP	-	-
	Flexibility in choice of security mechanisms	Low	Low	-	-
	Global availability of security mechanism	High	High	-	-
VR	Global connecitivty	High	High	-	-
	Service continuity	High	High	-	-
	Latency & Availability	High	High	-	-
	Acess to monitoring data and O&M functions	High/Medium	High/Medium	-	-
	Data privacy through isolation	Medium	Medium	-	-

	Control and management privacy through isolation	Low	Low	-	-
	Flexibility in choice of security mechanisms	Low	Low	-	-
	Global availability of security mechanism Remarks - Security control on VR applications still under evolution	Medium	Medium	-	-

8 SUMMARY

Today 5G networks are reality, mostly supporting mobile broadband type of communication services. As we move forward 5G for URLLC communication service is a next major wave of deployment foreseen by many players in the mobile communication ecosystem. As the nature of such URLLC use cases considering both requirements and deployment perspective is different from traditional mobile broadband networks, it is necessary to understand possible deployment options for such verticals.

The report starts with classifying verticals use cases into two major deployment possibilities; wide area and local area deployment. Further, different 5G technical enablers are elaborated which demonstrate URLLC capabilities satisfying requirement of vertical use cases.

Further, various Non-Public Network (NPN) and Public Network (PN) deployment options are investigated based on the 5G standalone architecture. Taking such NPN and PN options as base line, the report furthers develops various deployment architectures for wide area and local area vertical use cases. Possible deployment architecture options for four major vertical use cases are demonstrated.

Finally, the report evaluates provided deployment architecture against deployment attributes which are major 5G deployment requirements. The evaluation is done based on the degree of compliance of such attributes by each deployment options.

It is clear from this study; there are possibilities of public and non-public network deployments for each of such vertical use cases. Concerning NPN deployment, various configurations are possible in combination with wide area mobile network deployment. Apart from key technical enablers in radio access networks, edge computing and network slicing also play major roles in realizing such Non-Public Networks (NPN) deployment.

9 LIST OF ACRONYMS

3GPP	3rd Generation Partnership Project
5GS	5G System
A-GNSS	Assisted Global Navigation Satellite System
ACK	Acknowledgement
AGV	Automated Guided Vehicles
AL	Access Link
AMF	Access and Mobility Management Function
AOA	Angle of Arrival
AOD	Angle of Departure
AR	Augmented Reality
ARP	Antenna Reference Point
AS	Application Server
BLER	Block Error Rate
CC	Component Carrier
CDF	Cumulative Distributed Function
CDN	Content Delivery Network
CN	Core Network
CP	Cyclic Prefix
CQI	Channel Quality Indicator
DCI	Downlink Control Information
DL	Downlink
DN	Data Network
E2E	End-to-End
FDD	Frequency Duplexing Division
FISR	Fault Isolation and System Restoration
FRER	Frame Replication and Elimination for Reliability
GNSS	Global Navigation Satellite System
GTP-U	GPRS Tunnelling Protocol-U
HARQ	Hybrid Automatic Repeat Request
IMT	International Mobile Telecommunications
IP	Internet Protocol
LTE	Long Term Evolution
MCS	Modulation and Coding Scheme
MOCN	Multi Operator Core Network
MR	Mixed Reality
MgNB	Master gNB
MORAN	Multi Operator Radio Access Network
NACK	Non-Acknowledgement
NG-RAN	NG Radio Access Network
NPN	Non-public network
NR	New Radio
O&M	Operation and Management
OFDM	Orthogonal Frequency Division Multiplexing
OS	OFDM Symbol
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PI	Pre-emption Indication
PUCCH	Physical Uplink Control Channel

PUSCH	Physical Uplink Shared Channel
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
QUIC	Quick UDP Internet Connections
RAN	Radio Access Network
RLF	Radio Link Failure
RRC	Radio Resource Control
RSN	Redundancy Sequence Number
RTK	Real-Time Kinematic
SCS	Sub Carrier Spacing
SDO	Standard Developing Organization
SDU	Service Data Units
SCS	Subcarrier Spacing
SFI	Slot Format Indicator
SG	Scheduling Grant
SIB	System Information Block
SINR	Signal to Interference Noise Ratio
SNPN	Stand-alone Non-public network
SPS	Semi-Persistent Scheduling
SR	Scheduling Request
SMF	Session Management Function
SgNB	Secondary gNB
TB	Transport Block
TCP	Transport Control Protocol
TDD	Time Duplexing Division
TDOA	Time Difference of Arrival
TO	Transmission Occasions
TSN	Time Sensitive Networking
TTI	Transmission Time Interval
UCI	Uplink Control Information
UDP	User Data Protocol
UE	User Equipment
UL	Uplink
UPF	User Plane Function
URLLC	Ultra-Reliable and Low Latency Communication
eNB	enhanced NodeB
gNB	next generation NodeB