

The background of the slide is a photograph of a modern highway at night. The road is dark asphalt with a white line marking, curving into the distance. On the left side of the road is a white concrete guardrail. In the background, a city skyline is visible under a dark blue night sky. The skyline includes several tall buildings, with the most prominent one being the Burj Khalifa on the far left. To the right of the road, there are several large, illuminated arches that look like part of a bridge or a decorative structure, glowing with a warm light. The overall scene is a blend of modern infrastructure and urban development.

NOKIA

5G evolution

Learn what is behind it and how it
paves the way toward 5G-Advanced

Turning technology into innovation and growth

Nokia has played a key role in driving the first set of globally interoperable 5G standards to completion with 3GPP Release 15. We continue to realize the holistic 5G vision within 3GPP Releases 16, 17 and now with 5G-Advanced starting with Release 18.

We drive evolution for operator horizontal network expansion and enhancements for vertical industry sectors. Nokia shows exemplary leadership in the next stage of 5G standardization e.g. with Release 16 and 17 enablers for 5G Private Networks and Time Sensitive Communications (TSC) – a critical element in taking 5G to support deterministic wireless networks for applications that many industries need. In Release 18 we are e.g. leading the study for XR in the RAN domain.



Table of contents

Introduction	The evolution of 5G: What's behind it and how does it pave the way toward 5G-Advanced?
--------------	--

Rel-15	Rock solid mobility innovations from 5G to 5G-Advanced
--------	--

Rel-16	5G Non-Public Networks
	The evolution of 5G
	New Radio positioning technologies
	5G time service

Rel-17	5G reduced capability devices
	5G empowering Uncrewed Aerial Systems

Abbreviations

Additional resources

Introduction

The evolution of 5G

What's behind it and how does it pave the way toward 5G-Advanced?

By Peter Merz - Head of Nokia Standards



For more than 30 years, Nokia has defined many of the fundamental technologies used in virtually all mobile devices and networks. These have increasingly included consumer products, IoT-connected industrial equipment and vehicles. Nokia has been instrumental in enabling entire solutions and ecosystems built on its GSM, 3G/UMTS, 4G/LTE, 5G and now 5G-Advanced standards leadership. Contributing to the development of industry standards, such as 5G, accelerates innovation and digitalization. It also drives interoperability and expands customer and consumer choices.

We in Nokia Standards base our leadership approach around three main beliefs:

- Having a clear vision for the future of communications.
- Adopting Nokia-pioneered technology innovations into standards.
- Utilizing top-flight experts from Nokia Standards to guide and shape the relevant standards organizations and industry associations.

Nokia's standardization leadership position in 5G and 5G-Advanced is shaping how 5G will have real-world impact, changing the way we interact with each other, through the mass deployment of augmented, mixed/merged and extended reality (XR), for example.

At Nokia, we organize, optimize and customize various industry verticals, such as manufacturing, automotive, logistics, railways and more. We control and automate the environment around us, through automated driving and remote control. As a trusted vendor and partner, we put security at the center of all what we do.



Nokia played a key role in driving the first set of globally interoperable 5G standards to completion with 3GPP Release 15. We continued to realize the holistic 5G vision within 3GPP Releases 16 and 17. And now we are doing the same with 5G-Advanced, starting with Release 18, and expect to see the first set of products hit the market in the middle of the decade.

We spearhead the evolution for operator horizontal network expansion and enhancements for vertical industry sectors.

In this publication, we will detail this 5G evolution of the mobile networks in various ways.

The first section looks at Release 15 and further releases of ***Rock solid mobility innovations from 5G to 5G-Advanced***. High performing mobility solutions are at the heart of each mobile communications system. Today's 5G deployments largely rely on 3GPP Release 15 (the first 5G release) mobility mechanisms with traditional network-controlled and user equipment assisted handovers. Mobility innovations continue to be introduced in the different 3GPP releases and are one of the essential components of 5G-Advanced. 3GPP has already defined a set of functionalities and features that will be a part of the 5G-Advanced Release 18 package.

The next three sections focus on Release 16 and beyond. In ***5G Non-Public Networks*** we look at the two main deployment models: Standalone Non-Public Network and Public Network Integrated Non-Public Network. The basic idea is to enable seamless deployments of private networks, either as standalone networks or by integration into a public mobile network, while leveraging the strengths of mobile networks. We highlight the motivation for non-public networks, describe different deployment models and explain the features that are introduced.

In ***The evolution of 5G new Radio positioning technologies*** we show how positioning services are becoming important components of current and future cellular networks. From providing emergency services with device locations to tracking automated forklifts on a factory floor, the use cases for 5G NR technology continue to be enhanced through the release of 3GPP specifications. This paper looks at a wide diversity of anticipated use cases and scenarios.



Next comes **5G time service**. For nearly 100 years, “time-of-day” telephony or dedicated radio station services have helped humans set their clocks. Soon, applications and devices will use 5G networks to automatically set and keep their time to microsecond accuracy enabling new private, public and industrial use cases and experiences. In this paper, we explore some of the many possible use cases that can benefit from 5G time synchronization services.

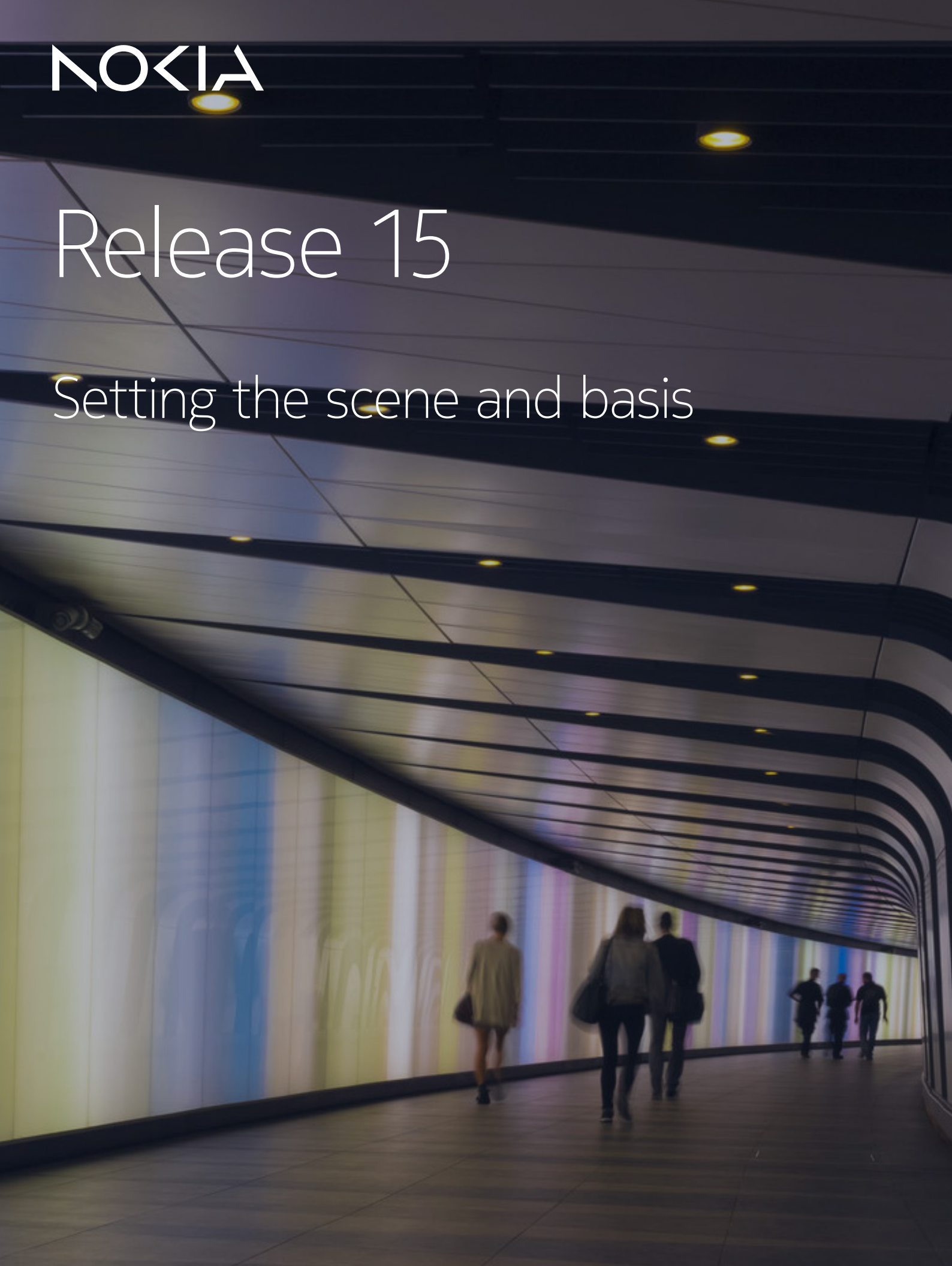
The final two sections relate to Release 17. In **5G reduced capability devices**, we provide an overview of the new device types and describe the new complexity reduction and power saving features being introduced to support mid-range IoT use cases. Our analysis shows that a significant reduction in complexity can be achieved. In addition, this paper presents network coverage analysis to show the impact from the introduction of reduced capacity devices.

Finally, **5G empowering Uncrewed Aerial Systems** introduces the concept of identifying user equipment used with aerial vehicles such as drones or uncrewed aerial vehicles. In this white paper, we highlight the motivation for supporting uncrewed aerial systems, describe different deployment models and explain the features that were introduced by 3GPP Release 17.

We hope these technical write ups provide an in-depth perspective on all these major themes as we march toward 5G-Advanced.

Stay tuned!

Peter

A photograph of a modern, curved tunnel with a checkered floor. The ceiling is dark with recessed lights. The walls are curved and feature a series of vertical, colorful light bands in shades of blue, yellow, and purple. Several people are walking away from the camera towards the end of the tunnel.

NOKIA

Release 15

Setting the scene and basis

Rock solid mobility innovations from 5G to 5G-Advanced

Contents

Introduction	3
Mobility framework in 5G	6
Main mobility metrics	8
Mobility robustness	8
Mobility interruption	8
Mobility signaling overhead	9
Mobility configuration and decisions	9
Mobility enhancements	10
Conditional handover for improved robustness	10
Mobility solutions for reduced handover interruption time	11
Fast failure recovery mechanisms	14
Slim mobility for fast cell switching	16
Machine learning for mobility	18
DCCA enhancements	18
Early measurement reporting	19
Secondary cell activation time improvements	21
Conditional SN addition and change for fast access	21
Activation of secondary cell group	23
High-speed train scenarios	24
Summary of features	25
Conclusion and outlook	26
References	27

Introduction

Well performing mobility solutions are at the heart of each mobile communications systems, with the first patent on handover for cellular dating back to 1972 by A.E. Joel from Bell Labs. Seamless handover between cells is essential for users to experience good quality of service while on the move. As an example, good handover performance is a prerequisite for users to move and engage freely with their physical environments without worrying about losing connectivity as they transit from indoors to outdoors and between neighborhoods or on the road. This requires highly agile and robust handover solutions with virtually zero interruptions. Today's 5G deployments largely rely on 3GPP Release 15 (the first 5G release) mobility mechanisms with traditional network (NW) controlled and user equipment (UE) assisted handovers, showing good mobility performance in the field. Mobility innovations continue to be introduced in the different 3GPP releases, including migration from the traditional NW controlled and UE assisted RRC-based handovers to more agile mobility solutions with higher degree of UE involvement and towards truly zero interruption time. Moreover, the innovations expand to new use cases such as e.g., high speed train (HST) scenarios, industrial IoT (IIoT) applications, and lately also to eXtended Reality (XR) as one of the prominent uses cases for 5G-Advanced that are in focus for 3GPP Release 18.

Mobility is one of the essential components of 5G-Advanced. 3GPP has already defined a set of functionalities and features that will be a part of the 5G-Advanced Release 18 package as illustrated in Fig. 1. These functionalities can be grouped into four areas: providing new levels of experience, network extension into new areas, mobile network expansion beyond connectivity, and providing operational support excellence. Mobility enhancements in Release 18 will be an important part of the 'Experience enhancements' block of features, with the goal of reducing interruption time and improving mobility robustness.

Figure 1: Overview of 5G-Advanced functionalities and features.

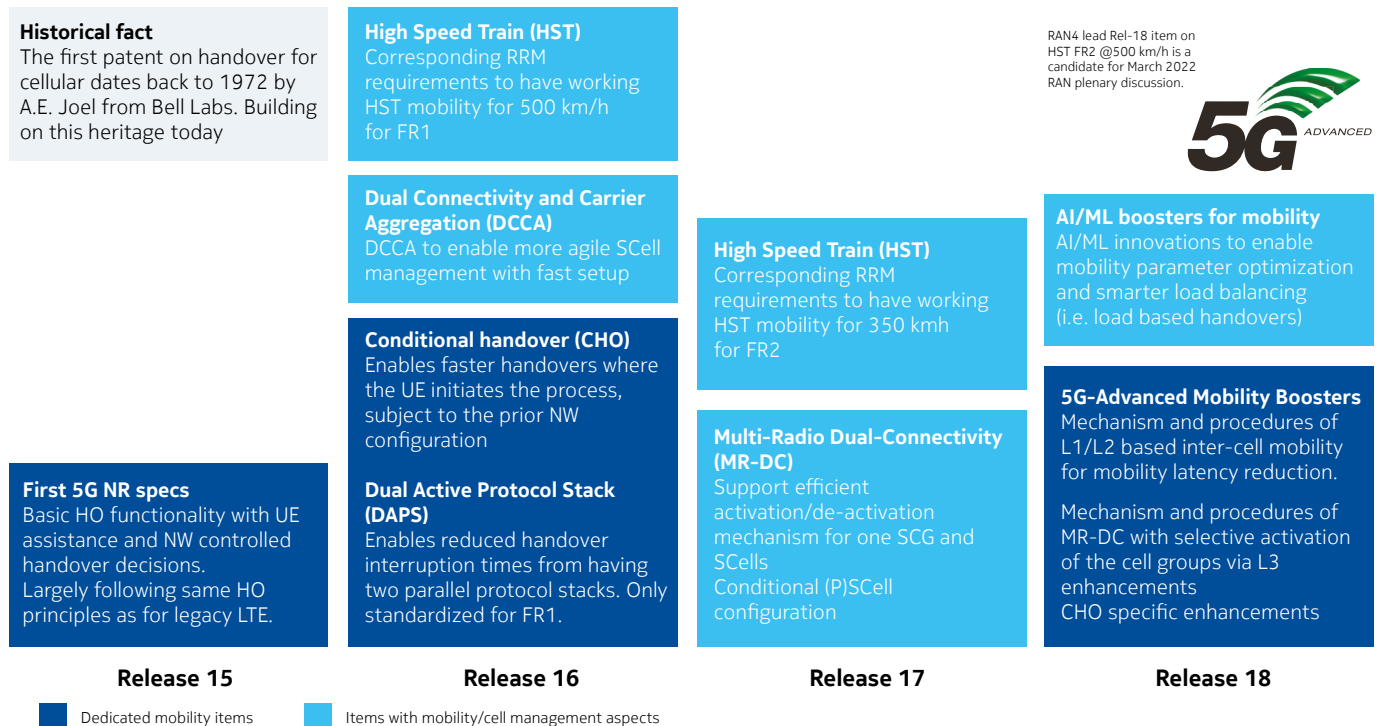


Fig. 2 shows a high-level schematic of mobility and dual connectivity (DC)/Carrier Aggregation (CA) related mechanisms that are introduced in the different 5G legacy releases towards 5G-Advanced in Release 18. Innovations such as Conditional Handover (CHO) and dual active protocol stack (DAPS) are introduced in Release 16. More efficient operation of carrier aggregation (CA), dual connectivity (DC), and the combination of those denoted as DCCA, as well as Multi-Radio Access Technology DC (MR-DC) are introduced through Releases 16 and 17. The enhancements for DCCA and MR-DC essentially enable more agile and efficient usage of such multi-carrier aggregation techniques to unleash their full potential by ensuring fast and timely configuration and activation of carriers in line with the traffic demands. MR-DC also provides a solution for non-standalone (NSA) deployment cases where a UE is connected to both Long Term Evolution (LTE) and 5G New Radio (NR) networks.

As pictured in Fig. 2, the first 5G-Advanced release (Release 18) is set to include additional mobility performance boosters making the handovers truly transparent for the end-user application with zero interruption time. It is desirable to achieve this with only a single transceiver in the devices including enhancements for inter-cell beam management. For enhancing mobility in millimeter wave frequency bands, the target is also to study improvements for the secondary cell setup latencies in carrier aggregation and dual connectivity to faster take advantage of the higher data rates from access to more bandwidth when needed. Essentially the enhancements will contribute to improved user experience as outlined in the 5G-Advanced white paper [1]. Finally, 5G-Advanced is also set to inject more artificial intelligence (AI) and machine learning (ML) technologies in the radio access network (RAN) and core network (CN). Among others, 5G-Advanced is expected to include innovations to facilitate smarter mobility parameter optimizations and decisions for triggering handovers.

In the rest of this white paper, we will present the main mobility aspects and provide an overview of the primary mobility enhancements introduced in 5G towards 5G-Advanced, as well as corresponding recommendations and best-practices. We address aspects related to both Frequency Range 1 (FR1) that covers up to 7.125 GHz and Frequency Range 2 (FR2) that covers mmWave frequencies up to 52.6 GHz. In addition, we describe the main enhancements that are introduced for DCCA where aggregation of FR1 and FR2 carriers is used and provide an overview about high-speed train scenarios.

Figure 2: Overview of mobility and DCCA related mechanisms in the different 5G and 5G-Advanced 3GPP releases.



Mobility framework in 5G

In NR, there are three Radio Resource Control (RRC) modes for the UE [2]:

- RRC Idle,
- RRC Inactive and
- RRC connected.

RRC Inactive is a new state that has been introduced for NR to reduce the latency in resuming the RRC connection after some inactivity. To this end, both control and user plane connections between 5G core (5GC) and RAN are maintained and the UE context is stored in the RAN and UE to enable fast resumption [2]. The basic mobility principles for RRC Idle and Inactive, defined in NR Rel. 15, follow the principles of UE controlled and network assisted procedures as also used for LTE, but are enhanced for considering the multi-beam operations and cell measurements [3]. Herein, RRC Idle/Inactive mobility is controlled by the UE where it decides on cell (re-)selection based on predefined conditions that are evaluated based on radio measurements and network configuration parameters. In contrast to RRC Idle UEs that are paged only by 5G core network for mobile terminated data, RRC Inactive UEs can be also paged by RAN in so-called RAN-based notification area (RNA).

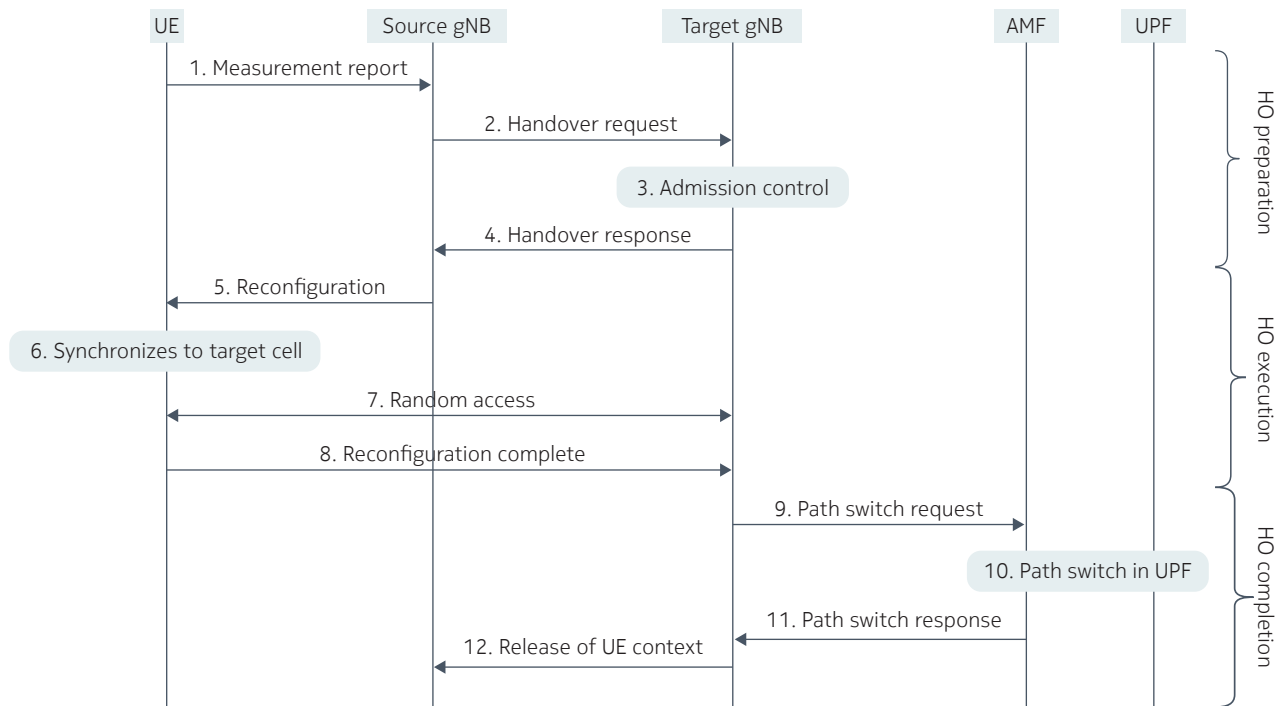
The main mobility enhancements that have been introduced in NR are rather for RRC Connected mode UEs. With the introduction of transmit beamforming in NR, 3GPP has distinguished from the beginning of Rel. 15 between inter-cell mobility of RRC Connected mode UEs, also known as handovers in case of Primary Cell (PCell) change [4], and Medium Access Control (MAC) based intra-cell beam management [5]. The beam management procedures have been defined to maintain a serving beam between the UE and the network. For beam management, the UE reports Layer 1 (L1) beam measurements of the serving cell and the network decides, based on the received measurements, on the beam(s) for the communication. The decision for beam switching, i.e., changing serving beam, is network-controlled by MAC protocol without involvement of upper Radio Resource Control (RRC) layer.

For the PCell change, which is one of the focus topics in this sequel, the UE switches the serving cell and monitors non-UE dedicated channels, e.g., broadcast and paging channels, from another cell. The decision for initiating the preparation of the PCell change is network-controlled and is given by RRC layer based on cell quality measurements that are reported by the UE. Depending on the network configuration, UE may apply additional L3 filtering to the L1 measurements in order to reduce the fluctuations in the measurements (caused by fast fading and measurement errors) and increase in turn the reliability of the cell change decisions. The decision for cell change can be either taken by the network or by the UE depending on the configured mobility solution.

UE mobility is supported in NR for high variety of scenarios including single connectivity and Multi-RAT Dual-Connectivity (MR-DC) [6], intra-/inter-frequency and intra-/inter-RAT cell changes. Moreover, mobility is supported for different types of architectures including standalone with monolithic gNBs and split architecture with Centralized Unit (CU)-Distributed Unit (DU) where the gNB consists of two separate entities, gNB-CU and gNB-DU, that are connected to each other by standardized F1 interface [7]. gNB-CU consists of two separate entities, gNB-CU-CP and gNB-CU-UP that are connected to each other by standardized E1 interface [8].

On high level, the baseline handover (HO) framework in NR Rel. 15 for PCell change consists of three main phases as shown in Fig. 3 [2].

Figure 3. Baseline handover framework in NR Rel. 15.



- Handover preparation: Based on measurements that are reported by the UE, the source gNB controlling the source cell can decide to trigger the preparation of a target cell in a target gNB. Upon receiving the handover request from a source gNB, the target gNB performs admission control, reserves radio resources that are needed to serve the UE, and generates a handover configuration to be applied by the UE at the time of handover execution.
- Handover execution: The UE applies the handover configuration of a candidate target cell by reconfiguring its protocol layers and synchronizes to the target cell by performing random access. Upon successful random access, the UE confirms to the network the completion of the reconfiguration.
- Handover completion: In the last phase of the handover, the target gNB triggers Access and Mobility Management Function (AMF) to switch the downlink user plane path of the User Plane Function (UPF) from the source gNB to the target gNB which will be serving the UE. The UE context is finally released from the source node.

The mobility enhancements that are specified for NR, up to Rel. 18, impact mainly the handover preparation and execution phases. The handover completion phase has been kept so far intact as data forwarding of the user plane data from source to target node during handover execution enables service continuity.

Main mobility metrics

In this section, we provide an overview about the main mobility metrics that have been considered by 3GPP as main targets for enhancements starting from Rel. 16 up to Rel. 18.

Mobility robustness

Robustness of the mobility procedures against failures is one of the key performance indicators. The baseline handover mechanism of NR Rel. 15 is robust and reliable solution achieving a low probability of mobility failures that is enough for many services and UEs. Nevertheless, as any other radio related procedure, the handover signaling is prone to failures which can have detrimental impact for some UE services requiring high reliability. There are two types of mobility failures depending on whether the failure happens before or during the handover execution. The failure during the handover execution phase where the UE fails to synchronize and perform random access to the target cell is called handover failure. On the other hand, the failure that occurs before the handover execution phase while the UE is still connected to the source cell is mainly referred to as Radio Link Failure (RLF).

To recover from the failures, UE performs RRC Re-establishment procedure to reconnect to a new target cell. The re-establishment procedure is costly as the UE needs to perform cell selection and read the system information of the target cell before performing a contention-based random access. Moreover, the target cell receiving the re-establishment request needs to fetch the UE context from the previous serving gNB and reconfigure the UE before it can start serving the UE. Most importantly, UE will experience interruption time during and after the mobility failure occurs until the UE completes the RRC Re-establishment procedure successfully. Thus, it is critical for the network to minimize and avoid such failures.

For UEs in DC mode, the primary cell of the secondary cell group (PSCell) may also fail. The failure of PSCell is less critical in terms of overall connectivity as the UE may still have a functioning connection with PCell of Master Cell Group (MCG). However, the failure of PSCell can be still harmful for services that are provided by SCG offering in some cases higher system bandwidth and user throughput than MCG. From this perspective, it is desirable also to minimize PSCell failures as much as the PCell failures.

Mobility interruption

Any mechanism for cell change may introduce a certain interruption in the on-going service, which is called mobility interruption. The duration between the point in time when the UE stops transmission/reception to/from the source gNB and the time when target gNB resumes the transmission/reception to/from the UE is called the mobility interruption time. For instance, in the baseline handover mechanism that is presented in Fig. 2, the interruption starts from the reception of the RRC Reconfiguration (step 5) until the transmission of the RRC Reconfiguration Complete to the target gNB (step 8). The mobility interruption time includes the processing of the RRC message, L2/3 reconfigurations, the random access procedure and the transmission of the RRC message confirming the reconfiguration to the target gNB. The mobility interruption time varies depending on the scenario, e.g., intra-frequency or inter-frequency handover, and can reach up to 80 ms [9].

Mobility interruption time is critical for services requiring high reliability or small outage. For example, AR/ XR services may be hindered by the interruption time introduced by the handovers [10][11]. The same applies for URLLC or vehicular communication services such as “Automated Intersection Crossing” or “Cooperative Maneuvers”. Due to its high relevance, 3GPP has started to address mobility interruption time early in 5G with Rel. 16 and will continue to work on this in 5G-Advanced.

Mobility signaling overhead

Another mobility aspect is the signaling overhead that is associated with the change of cell, Secondary Cell (SCell) or cell group, i.e., Secondary Cell Group (SCG) or Master Cell Group (MCG) in NR-Dual Connectivity (NR-DC) deployments, and the high delay in triggering the cell change.

Up to Rel. 17, the UE releases all stored configurations for candidate target PCells or Primary Cells of Secondary Cell Group (PSCells) upon performing a successful handover or PSCell change, respectively. Triggering a subsequent handover or PSCell change requires the network to re-initiate the cell change preparation procedure from scratch even though the UE has stored configuration for the same candidate target cell before the switch to new cell group. This does not only increase the signaling overhead over the radio and network interfaces, but it also delays the execution of the handover or PSCell change which can lead to some outage or even failures. These aspects are especially relevant in high carrier frequencies where there is high number of cell group switches and UE received signal is more vulnerable to rapid degradation caused by blockage from user body, hand or other obstacles.

Mobility configuration and decisions

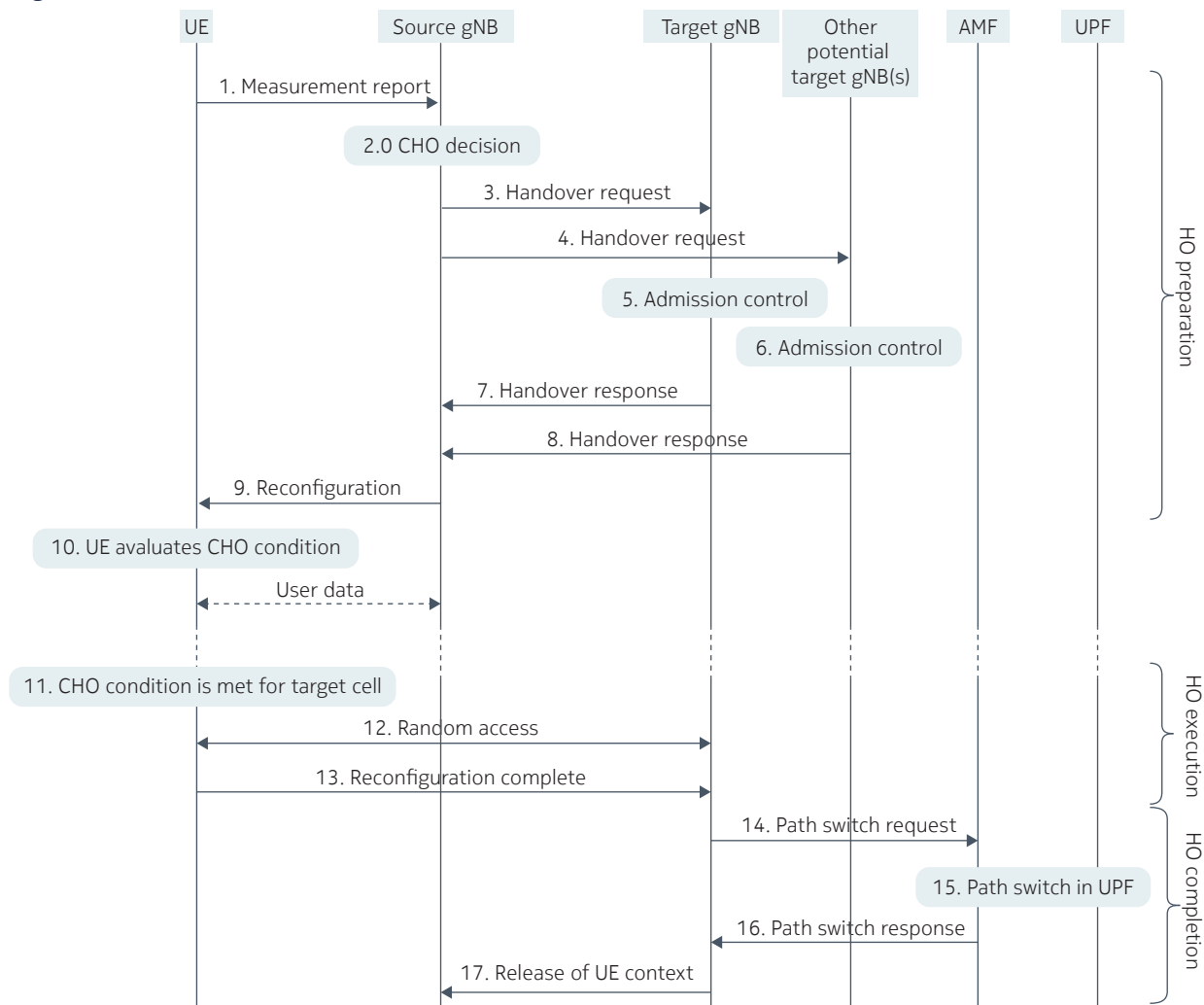
The performance of the mobility procedures highly depends on the proper configuration of the parameters controlling the handover and the decisions made by the source and target nodes. For instance, the handover parameters controlling the triggering of the measurement report in step 1 of Fig. 3 needs to be properly configured such that the UE can still receive the RRC Reconfiguration message from the source node in step 5 and at the same time be able to perform a successful random access to the target node in step 7. In addition, the source node should make in step 2 the right decision on the target cell to be prepared in the target node. To ease the handover parameter configuration for the network operators, Nokia has been developing Self Organizing Network (SON) solutions that automatically optimize the mobility parameters based on performance metrics that are collected periodically from the network. In 5G-Advanced, the optimization of the mobility parameters and decisions will benefit further from the cutting-edge AI/ML techniques that can leverage the dependencies among different handover parameters, cell borders and network nodes.

Mobility enhancements

Conditional handover for improved robustness

Most mobility failures are caused by the failure of receiving the handover command (step 5 of Fig. 3) on time from the source cell. To address this issue, CHO has been introduced in NR Rel. 16 where the handover preparation is de-coupled from the execution phase as shown in Fig. 4. Herein, the source node triggers the preparation of one or more target cells early when the UE still has a very good radio link to the source node. The source node provides the configurations of the prepared target cells to the UE in step 9 along with a CHO execution condition that is associated with each prepared target cell.

Figure 4. Conditional Handover framework in NR Rel. 16.



After receiving the configurations of the target cells, the UE evaluates the CHO execution conditions to trigger the handover execution when the condition is met as shown in step 11. The remaining steps related to handover execution and completion are similar to baseline handover of NR Rel. 15 shown in Fig. 3, with some modifications to support early and late data forwarding from source node to the target node of the handover procedure [2].

Due to its improved mobility robustness performance, CHO has been adopted by many 5G verticals such as Non-Terrestrial Networks (NTN), Integrated Access and Backhaul (IAB) networks and NR-Unlicensed (NR-U) [12].

To improve the robustness of PSCell change in MR-DC scenarios, Conditional PSCell Change (CPC) has been introduced for intra-SN (PSCells from same node) and inter-SN (PSCells from different nodes) scenarios in Rel. 16 and 17, respectively. CPC follows the same conditional reconfiguration paradigm of CHO where the UE is prepared early with one or multiple candidate target PSCells and execute the change to a prepared target PSCell when its corresponding condition is met. This would improve the reliability of the PSCell change and reduce the interruption time for the services that are provided by SN.

Further enhancements for CHO in MR-DC scenarios will be studied in 5G-Advanced Rel. 18 [13]. Herein, the radio link quality of the target PSCell will be considered by the UE for initiating the random access to target SN when CHO execution condition is met for the corresponding target PCell. This inter-working between CHO and conditional PSCell addition/change will improve the robustness of both PCell and PSCell change in inter-MN handover scenarios.

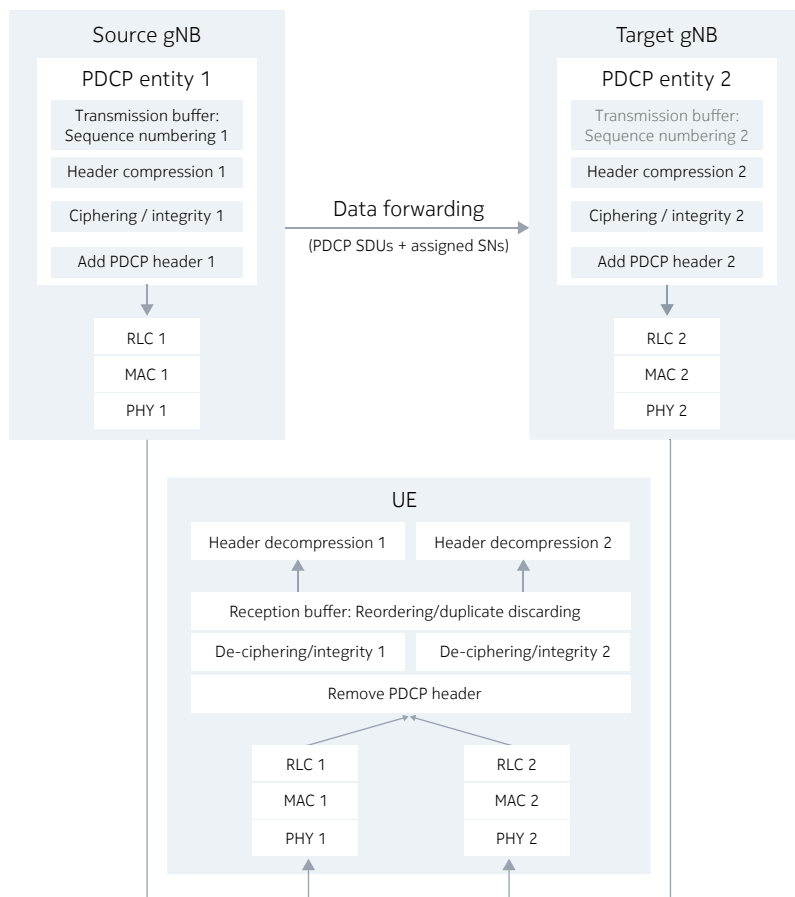
Mobility solutions for reduced handover interruption time

Dual Active Protocol Stack (DAPS) handover has been introduced by 3GPP in Rel. 16 as the first NR solution to reduce interruption time during the handover. The main difference between DAPS and baseline handover of Rel.15 is the handover execution phase which has been enhanced to minimize interruption time. Mainly, after receiving the handover command the UE maintains the connection with the source gNB and continues to receive user plane packets while performing access to the target gNB. This avoids that the UE experiences user plane outage while waiting for the target cell radio link to be ready. After completing the random access, the UE starts to receive user plane packets simultaneously from both the source and the target cells until the source cell is released by the target cell. The UE switches the uplink user plane transmissions to target cell when random access is completed. Path switch is then triggered during the handover completion phase by the target node in the same way as in baseline handover.

Fig. 6 illustrates the dual protocol stack maintained at the UE during DAPS handover to receive downlink packets simultaneously from both source and target cells. As can be noticed, the UE has dedicated physical (PHY), MAC and Radio Link Control (RLC) layers for source and target cells. Moreover, the Packet Data Convergence Protocol (PDCP) has been enhanced to support separate (de)-ciphering and header (de)-compression layers with common functionality to re-order and discard duplicate packets received from source and target cells [14].

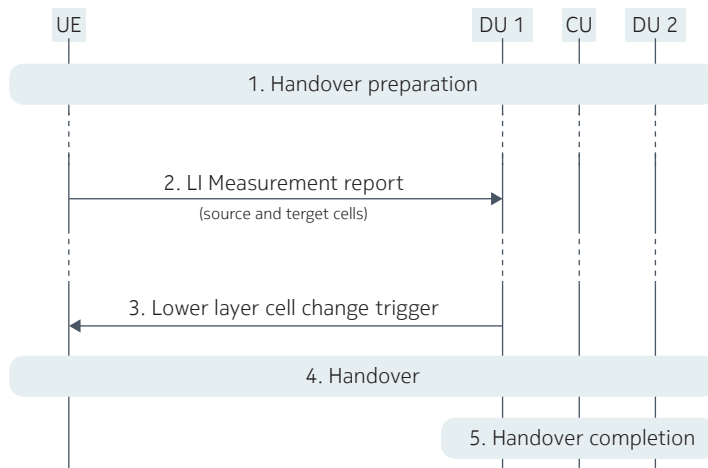
DAPS handover can reduce the handover interruption down to 2 ms [15]. As for the UL, there is an additional end-to-end delay, i.e., in addition to 2 ms, as the target gNB can send the received packets to UPF only after the target gNB contacts the source node and receives information about the first UL packet that needs to be forwarded.

Figure 6. Downlink operation of dual active protocol stack (DAPS) handover.



One limitation of DAPS handover is that it does not support the scenario where source and target cells operate in higher frequencies, thus leaving a gap for FR2 connectivity. The main reason for this is that 3GPP could not conclude in Rel. 16 that multi-panel UEs would be able to communicate simultaneously with two gNBs operating in FR2. Additionally, the implementation of two simultaneous active protocol stacks might be for the time being challenging for some UE vendors. For these reasons, 5G-Advanced will study L1/2 inter-cell mobility in Rel. 18 as an alternative solution to reduce the interruption time during handover for both FR1 and FR2. L1/2 inter-cell mobility is expected to work for UEs with single active protocol stack and to operate in CU-DU split architecture where the CU hosts the RRC and PDCP layer and the DU controls the lower PHY, MAC and RLC layers. Moreover, only intra-CU handovers will be considered at the first stage for L1/2 inter-cell mobility to reduce the interactions with other CUs/gNBs. Fig. 7 shows the high-level signaling for L1/2 centric mobility solution in 5G-Advanced.

Figure 7. High-level Signaling for L1/2 centric mobility.



The CU may initiate in step 1 the preparation of multiple target cells in serving DU 1 or another DU 2 and provide their corresponding configurations to the UE. Based on L1 beam measurements for source and target cells that are reported to the network in step 2, the serving DU 1 may trigger the handover to another cell using lower layer signaling as shown in step 3. When executing the handover in step 4, the UE does not need to spend much time on reconfiguring the higher PDCP and RRC layers and may skip the random-access procedure in case of synchronized cells which reduce substantially the handover interruption time.

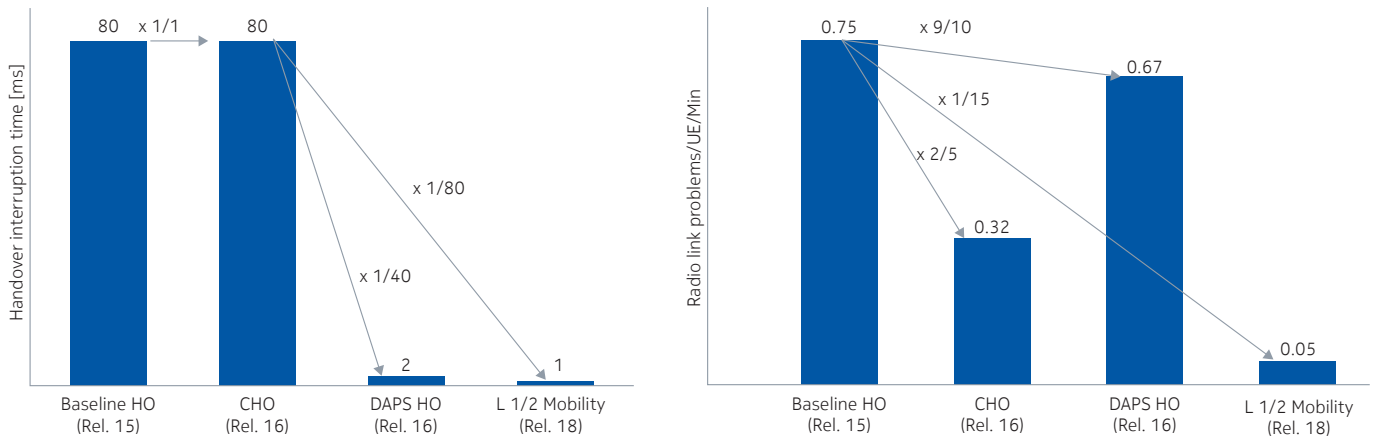
The performance of the main mobility enhancement procedures in 5G and 5G-advanced is illustrated in Fig. 8 with respect to handover interruption time expressed in ms and the normalized number of radio link problems (RLPs), i.e., the number of times the radio link quality of the serving cell falls below the acceptable level for reliable radio communication. The results for the RLPs are shown for 3GPP compliant Urban Macro (UMa) [16] with 7 sites, each with 3-sector applying analog beamforming, i.e., one serving beam at a time. The results are generated for a UE speed of 60 km/h.

According to the figure, the handover interruption time of CHO (Rel. 16) is similar to that of baseline handover of Rel. 15 which can be up to 80 ms as stated in [9]. DAPS HO of Rel. 16 reduces the handover interruption by a factor of $\times 1/40$ compared to baseline HO/CHO. The upcoming L1/2 inter-cell mobility solution is expected to reduce the interruption even more by a factor of $1/80$. As such, it should be possible to reduce the handover interruption time down to 1 ms in L1/2 centric mobility [9] which would be useful for services requiring ultra-high reliability.

In addition, CHO reduces the number of radio link problems by a factor of $\times 2/5$ compared to BHO. This is because the UE receives the configurations of the candidate target cells early which allows the UE to autonomously trigger the cell change before the radio link of the serving cell fails. The number of RLPs did not improve much for DAPS HO which was aiming at reducing the handover interruption time and not improving mobility robustness. As for L1 centric inter-cell mobility, the number of radio link problems are reduced substantially by a factor of $\times 1/15$ compared to baseline handover of NR Rel. 15. This is because the cell change is decided based on L1 beam measurements that do not suffer from the delay caused by L3 filtering and Time-To-Trigger (TTT) applied for cell quality measurements. This allows the network to react fast to the degradation in the radio link quality of the serving cell and trigger the handover fast with minimal interruption when needed. Fast triggering of the handover may lead to increase in ping-pongs (consecutive forth and back handovers). However, these ping-pongs are not that critical for L1/2 centric

mobility given that the signaling between the source and target cells are handled in the same DU entity (in case of intra-DU cell change) and the interruption time during L1 cell change is small and not costly as in baseline handover or CHO.

Figure 8: Performance comparison of baseline handover, conditional handover (CHO), DAPS HO and L1/2 centric mobility.



Fast failure recovery mechanisms

Despite mobility failures are rare events in the network, fast recovery mechanisms are specified in 5G to mitigate the service interruption time caused by failures and the additional signaling overhead and delay of the re-establishment procedure.

In order to reduce the UE outage time before failure detection, 5G has adopted T312 timer which allows the UE to initiate the re-establishment procedure earlier, and consequently, recover the radio link faster [4]. The UE starts a timer T310 when the physical layer detects that the radio link of the serving cell is degrading. In case the UE sends a measurement report to the serving cell, i.e., identifying a target cell for handover, the UE can start another T312 timer if timer T310 was already running. If the UE does not receive from the serving cell a command to trigger the handover before T312 timer expires, UE declares RLF and initiates the re-establishment procedure. As the value of T312 is configured to be much less than that of T310, the UE can perform a faster failure recovery as it does not need to wait anymore for T310 timer to expire for initiating the re-establishment. Note that shortening the value of T310 timer is not an option as the UE may start to declare unnecessary failures that are caused by temporary signal degradations and which can be recovered without re-establishment procedure that is costly in terms of additional signaling and service interruption time.

CHO recovery is another feature that has been introduced by 3GPP in Rel. 16 to reduce the interruption time caused by failures for UE that is configured with conditional reconfigurations for multiple target cells. Despite CHO minimizes the probability of mobility failures, it is still possible that UE detects a failure due to misconfiguration of the mobility parameters or handover execution to wrong cell etc. Instead of performing re-establishment, a UE supporting CHO recovery feature can make use of the stored conditional reconfiguration for prepared target cells to recover from the failure. After the failure is declared, UE initiates the cell selection and if selected cell is one of the prepared target cells, UE executes the handover towards that cell using the configuration that was provided previously by the network. Otherwise, the UE initiates the re-establishment procedure to the selected cell if it does not have a

corresponding conditional configuration. Hence, using CHO recovery, the UE can leverage the stored conditional configurations for target cells to initiate a CHO execution instead of costly re-establishment procedure, reducing thereby the interruption time after the failure.

Figure 9. Performance of CHO recovery [12].

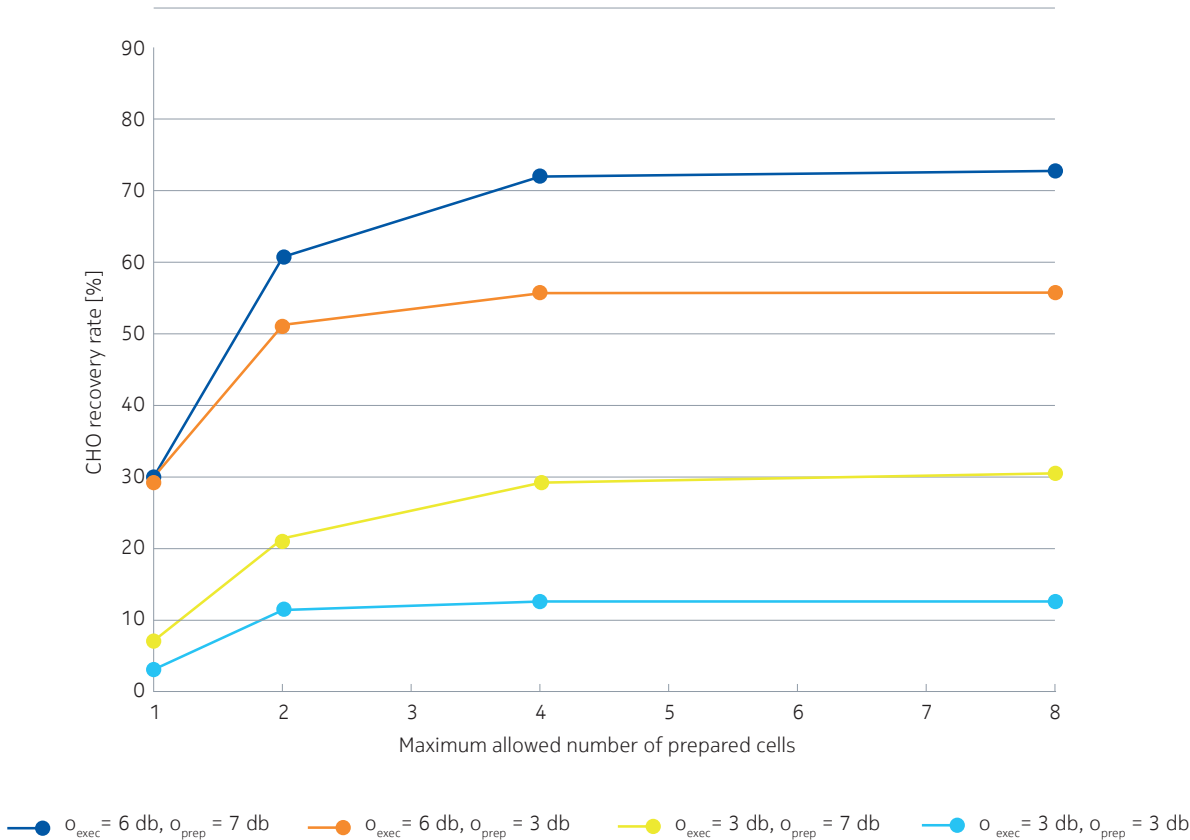


Figure 9 shows the mobility performance of CHO recovery mechanism. The CHO recovery rate, defining the percentage a CHO is executed after a failure instead of re-establishment procedure, is shown as a function of maximum allowed number of prepared cells for different CHO preparation offset o_{prep} and execution offset o_{exec} [12]. A target cell is prepared if it is not weaker than the source cell by more than o_{prep} dB and the CHO is executed if the target cell becomes o_{exec} dB stronger than the source cell. CHO recovery rate shows the percentage of the failures that are resolved by executing CHO instead of performing the costly re-establishment procedures. In those simulations, UMi-street canyon propagation channel model is used, where path-loss, shadow fading, fast-fading, and soft LOS are considered [16]. Each cell is configured with 12 transmit/receive beams and can schedule/activate 4 beams simultaneously for increased spatial multiplexing. The network layout is 7-site hexagon where each site has 120° sectoring (3 cells on one site) and there is 200 m inter-site distance (ISD) between the neighboring sites. In the simulated scenario, the mobility performance of the UEs is investigated for a user speed of 60 km/h.

The initial observation from Figure 9 shows that increasing the maximum number of prepared cells increases the chance that the CHO recovery could resolve the failure problems, i.e., CHO recovery rate increases by ~5% (shown in cyan) up to 45% (shown in dark blue). This is because when more cells are prepared, it is more likely that the cell selection procedure results in selecting one of the prepared cells to which

the UE can perform CHO recovery. Apart from that, in either case, i.e., execution offset of $o_{\text{exec}}=3$ dB or $o_{\text{exec}}=6$ dB, CHO recovery rate increases by 20% when the preparation offset o_{prep} is increased by 4 dB since larger preparation offset enables earlier preparation of the target cells which can be a selected cell during cell selection after failure and CHO recovery. Last but not least, the benefit of the CHO recovery is pronounced more when the CHO execution offset o_{exec} is high, i.e., 6 dB where the higher CHO execution offset postpones the CHO execution and leads to failure. In those cases, UE can use the CHO recovery procedure where CHO recovery rate increases by 45% when o_{exec} is increased by 3 dB. Hence, delay caused by re-establishment procedure is mitigated by using CHO recovery mechanism.

As mentioned before, DAPS handover can reduce the handover interruption time close to zero ms. However, the execution of DAPS handover may not always be successful depending on the radio conditions of the source and target cells. It may happen that the DAPS handover is triggered early by the network leading to random access failure at the target cell. Unlike baseline handover procedure of NR Rel. 15, UE does not experience service interruption during target cell access failure since UE still maintains the connection with source cell. Thus, if the random access fails to the target cell and the radio link of the source cell is not in failure, the UE fallbacks back to the source cell configuration by releasing the protocol stack of the target cell. In this way, the UE remains connected to the network without the need to initiate the re-establishment procedure after a random access failure.

Slim mobility for fast cell switching

In order to reduce the signaling overhead associated with the cell change and reduce the delay in the execution, fast cell switching among candidate serving cells is considered in NR Rel. 18 under L1/2 inter-cell mobility. Herein, dynamic switching is foreseen among candidate serving cells and SCells using lower layer signaling, where the cells are controlled by the same Centralized Unit (CU). This is particularly relevant for FR2 scenarios where dynamic shadowing caused by user and moving obstruction is dominant resulting in high number of handovers.

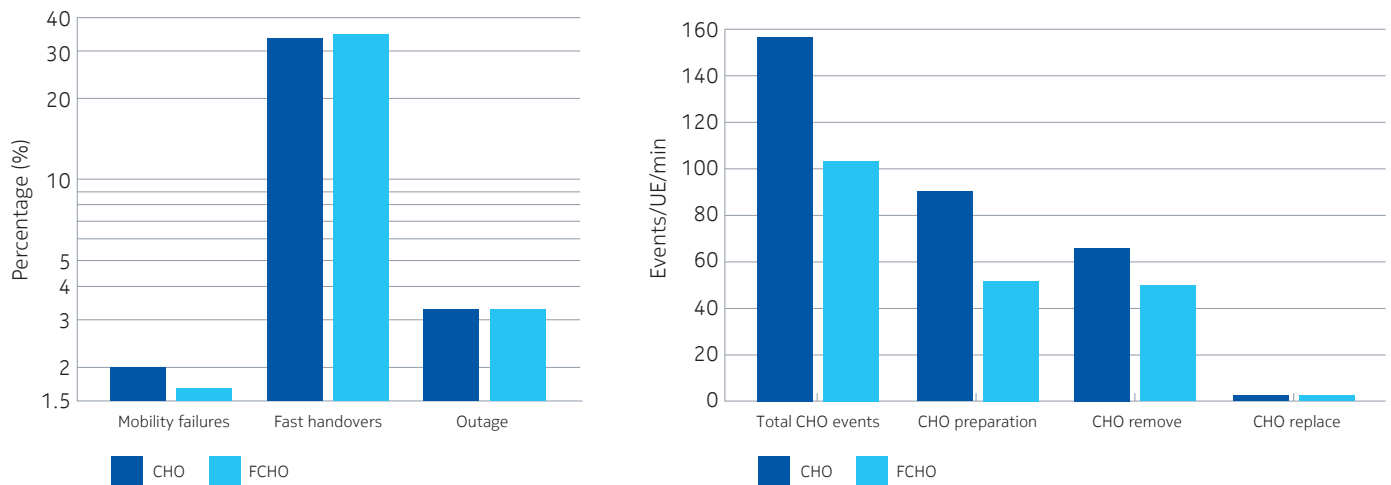
The concept is also candidate to be extended in NR Rel. 18 for L3 mobility where subsequent change of CG (at least for SCG) can be initiated without L3 reconfiguration in NR-DC deployment. Despite the differences in the approach of triggering the cell change and applicable scenarios, a common Radio Resource Control (RRC) signaling would be desired for both enhancements to leverage the synergies.

The solutions for fast cell switching require that the UE is prepared beforehand with multiple candidate serving cells, SCells or CGs in a similar way to CHO/CPAC. The set of cells/CGs to be prepared can be decided by the gNB using measurement reports that are configured to the UE. In contrast to CHO/CPAC, the UE and the network shall maintain at least some of the configured preparations after the cell/CG change is successfully executed. In this way, a subsequent change of the cell/cell group can be triggered without re-initiating the preparation and performing reconfiguration. For this to work, it is necessary that the stored configurations at the UE side remain valid for execution after the cell/CG change, i.e., does not lead to RRC reconfiguration failure when applied, and the UE is always configured with the preparation of the right candidate cells/CGs. For the latter case, the network may need to configure the UE with new cells/CGs and/or release some stored configurations that are no longer relevant for the mobility of the UE.

Applying fast switching for MCG can save the first five steps of Fig. 3 if the UE keeps the configurations of relevant prepared candidate cell groups after the CG change. This can result in saving at least 2 RRC messages on radio interface and 2 other messages on the Xn interface which in turn reduces the delay in triggering the CG change by at least 40-50 ms. For CU-DU split architecture, the benefits are even more pronounced as the signaling between CU and DU over F1 interface and between CU-CP and CU-UP over E1 interface are also saved.

The performance of fast selective switching of MCG based on CHO framework, denoted by FCHO, is compared against CHO in Fig. 10 [17]. The simulation scenario is the same as that used for Fig. 9. In FCHO, the CHO configurations of the candidate target PCells are kept after the handover whereas in CHO they are released by the UE after CHO execution/handover. The comparison is performed with respect to percentage of mobility failures, fast handover (consecutive handovers within short time comprising of ping-pongs and short stays), outage and total number of CHO events. The latter consists of the sum of CHO preparation events (number of times a target cell is prepared), CHO remove events (number of times a prepared target cell is released) and CHO replace events (number of times a prepared cell is replaced by another cell having better radio link quality). As shown in Fig. 10, the total number of CHO events is reduced by approximately 38 % compared to CHO. This is because the UE keeps the CHO configurations of candidate target cells after the handover, and consequently, the network does not have to re-trigger the preparation of the cells which reduces substantially the signaling overhead on the radio and network interfaces. In addition, the percentage of mobility failures decreases since the UE can trigger the handover faster given that it keeps the configurations of the candidate target cells.

Figure 10. Simulation results showing the benefits of fast selective switching of cell group [17]



Machine learning for mobility

It is challenging to find optimal settings for handover decisions with conventional mobility solutions and MRO methods which highly rely on the stationarity of the performance metrics collected from the network. In contrast, Artificial Intelligence (AI)/Machine Learning (ML) algorithms can learn the local context, e.g. radio environment, UE measurements, UE mobility pattern/trajectory, network traffic pattern, service-specific requirements, etc. , and provides the opportunity for a “tailor-made” handover triggering decisions. As such, a much better mobility performance is achieved considering dynamicity of the environment.

AI/ML solutions for mobility would aim at reducing the probability of mobility failures and ping-pongs by triggering the handover on time and selecting the right target cell of handover. Moreover, saving on network and radio resources can be achieved through, e.g., predictions of measurements or the time window in which the UE is expected to perform handover allowing a prepared target cell to make overbooking of the reserved resources.

3GPP Rel. 18 will specify for 5G-Advanced solutions for enabling AI/ML functionality in RAN and Core and the needed inputs and outputs for AI/ML. One of the considered main use cases is mobility optimization which is going to consider the following aspects [18]:

- The location of AI/ML model training and AI/ML model inference.
- AI/ML related signaling aspects between RAN nodes and UE.
- Input data from various network entities and UE for AI/ML model training and inference.
- Output of the model and performance feedback of AI/ML model.

Further information on AI/ML 3GPP study for RAN can be found in [19].

DCCA enhancements

Another mobility related aspect is efficient cell management for cases with Carrier Aggregation (CA) and/or Dual Connectivity (DC). For harvesting the full benefits of CA/DC techniques, it is important to have an agile framework where secondary cell(s) are timely identified and configured to the UE when needed. This is of importance for non-standalone (NSA) deployments where a carrier on NR should be quickly configured and activated to take advantage of 5G. Similarly, it is of importance for standalone (SA) cases where e.g. a UE with its Primary Cell (PCell) on NR Frequency Range 1 (FR1) wants to take additional carriers, either on FR1 and/or FR2 bands, into use. Thus, there is a need to support cases where the aggregated carriers are either from the same or difference sites. The management of such additional carriers for a UE shall be highly agile in line with the user traffic and QoS demands; quickly enabling usage of additional carriers when needed and again quickly released when no longer demanded to avoid unnecessary processing at the UE and to reduce its energy consumption. This is of particular importance for users with time-varying traffic demands (aka burst traffic conditions).

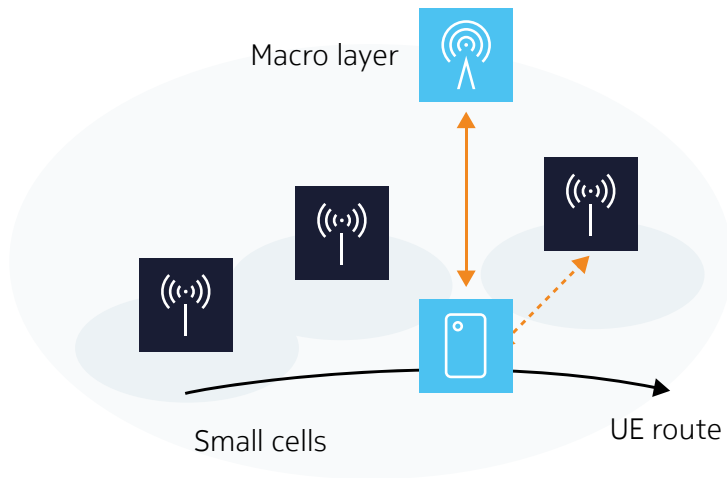
In the following, we describe how such carrier management is gradually improved by introducing enhancements for cell identification, RRM measurements and reduced reporting delays from UEs. As well as innovations related to Conditional PSCell Addition and Change (CPAC) and deactivation of secondary cell groups are outlined.

Early measurement reporting

Offloading traffic to small cells is one of the key enablers in achieving high peak data rates in a heterogeneous cellular network deployment. However, to maintain the benefit of robust mobility offered by the macro cell layer, ideally the connection to the macro cell is still maintained even when bulk of the user data is offloaded and served in a higher capacity small cell. Hence, enabling fast access to the offloading layer has the opportunity to increase the user data perceived throughput by a factor of 5x-10x.

In 3GPP based systems (LTE and 5G NR), multi-link setup getting the best of both worlds can be achieved by CA or DC – which one of these is used depends on the deployment and the back-haul latency between the macro and small cells. An example of heterogeneous network deployment with inter-site small cells (e.g. inter-site CA or DC) is illustrated in the schematic of Fig. 11.

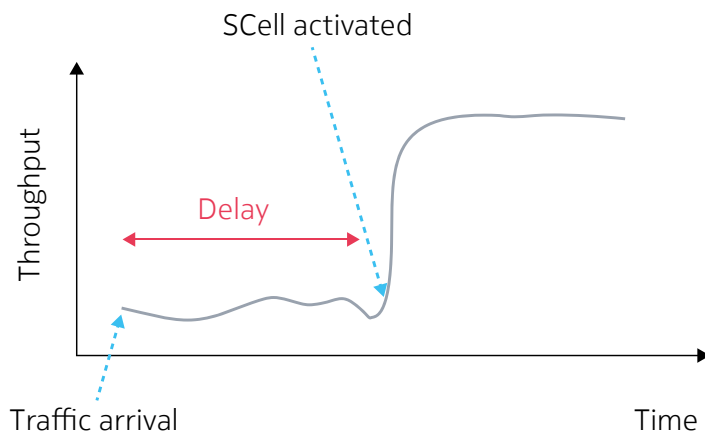
Figure 11 Heterogeneous network deployment with inter-site small cells.



A large part of the delay in accessing the (P)SCell comes from the inter-frequency or inter-RAT measurements (which can be 500+ ms, depending on number of measured frequency layers). This reduces the benefit of CA/DC as the larger bandwidth can be only configured later on, which may degrade connection throughput and latency that the user experiences. The challenge in this case is the delay in getting the best small cell configured for the UE fast enough to maximally benefit out of it, either as a Secondary Cell (SCell) in CA or as a cell in Secondary gNB (SgNB) in DC.

Delay in accessing the small cells represents the main challenge: In LTE, it can take over 500 ms (per measured carrier) to discover, measure (reliably), report, configure and activate a small cell as a SCell in CA. In NR, similar problem is experienced, and the magnitude of the delay in FR1 is similar to LTE. In practice, this means that the potentially very high data rate of the small cell (up to 10-100x data rate compared to the macro cell, depending on e.g. the bandwidth and cell load) is not truly perceived by the user due to the delay in setting up the SCell, as illustrated in Fig. 12.

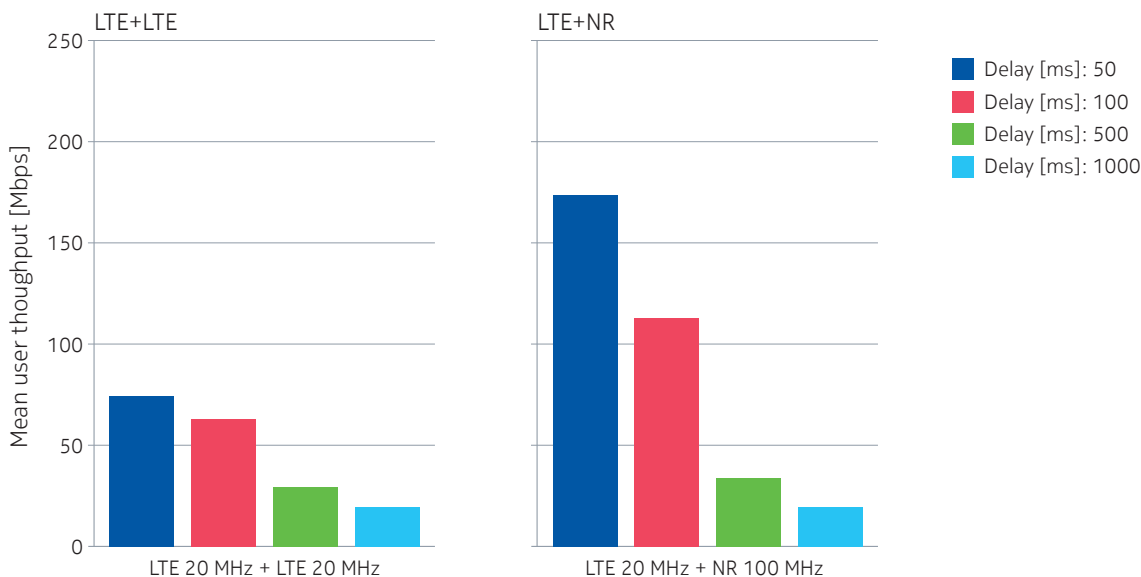
Figure 12 illustration of the observed performance bottleneck resulting from the delay in setting up the CA/DC and getting inter-frequency small cell activated on time.



Nokia pioneered the solutions of the fast access in case of CA/DC, initiating and leading the work in Rel-15 LTE and continuing as co-leader in bringing the LTE approach to Rel-16 NR.

Reducing the measurement delay (i.e. biggest contributor to the (P)SCell setup delay) improves latency of (P)SCell set-up times for all scenarios including the transition from RRC Idle/Inactive to Connected for first time CA/DC setup and (P)Scell addition in RRC CONNECTED. This results in improved UE throughput and paradoxically, even achieves lower UE power consumption because UE easily ends up having shorter active time in CONNECTED. Standardized solution for both Rel-15 LTE and Rel-16 NR has in its fundamental the following structure: UE will perform Idle/Inactive mode cell measurements (or early measurements reporting (EMR)) to enable faster (P)SCell setup immediately after data arrival. This will lead to multiplying the user perceived throughput by a factor of >5x. Example benefits are given in Fig. 13 which shows the mean user throughput [Mbps] as a function of the measurement delay for both LTE + LTE and LTE + NR scenarios, and also illustrates how NR can provide more benefits due to larger amount of available bandwidth.

Figure 13. Performance analysis of fast CA/DC setup (Rel-15 LTE, Rel-16 NR)



The major benefits of CA/DC setup delay reduction are seen in both UE and NW side:

- Reduced CA/DC configuration and activation delays.
- Improved (P)SCell usage leading to more efficient (P)SCell utilisation.
- Faster data offloading from PCell.
- Enhanced resource utilization in the CA/DC network.
- Improved user throughput.
- Enhanced capacity, especially when PCell capacity is limiting and wide BW available for (P)SCell.

As a next step, 3GPP will study in Rel. 18 the impact of FR2 mobility measurement and reporting on the SCell/SCG setup/resume delay for a UE connecting from RRC Idle/Inactive mode. The outcome of the study may result in defining new UE measurement procedures or reporting and related RRM core requirements [13]. As a result, the CA/DC setup delay of SCell/SCG in FR2 will be reduced due to enhanced UE measurements.

Secondary cell activation time improvements

In this section, an overview of two main enhancements for SCell activation time is provided.

Direct SCell activation

With legacy CA, all SCells are in deactivated state upon configuration and after handover, requiring an additional MAC-based activation afterwards. As part of fast CA/DC set-up and configuration, an additional mechanism called Direct SCell Activation was introduced to improve the overall latency of setting up an activated SCell. As with EMR framework, Direct SCell Activation was first introduced in Rel-15 LTE and then later also introduced in Rel-16 NR. The Direct SCell Activation allows to have the SCell configured in activated state immediately, which reduces latency of SCell activation by removing the need for separate SCell activation message. This ensures more efficient CA since the network can retain previously activated SCells as active even during mobility or change of SCell. This was further enhanced in Rel-17 by adding better control of the beam management (i.e. TCI state) for the SCell, allowing more control for the network to select the beam serving the UE in the SCell.

TRS-based SCell Activation

When a deactivated SCell is activated, the UE needs to first receive Synchronization Signal Block (SSB) of the SCell for fine time tracking purposes. The periodicity of SSB transmission may vary and can be up to 160 ms, which introduces a fixed minimum delay to the overall SCell activation delay. Rel-17 NR introduced an additional method to aid the faster CA/DC setup, called temporary RS (TRS) - based SCell activation. The method enables the network to transmit a burst of UE-specific RS to aid the UE in time/frequency acquisition in the SCell, and UE will use additional RS signals for time tracking without having to wait for the cell-specific signals (SSB). As these additional reference signals are provided immediately after the MAC-CE command for SCell activation, they are available for the UE immediately and the UE can measure these without delay, which enables faster SCell activation by eliminating the fixed SSB acquisition delay.

Conditional SN addition and change for fast access

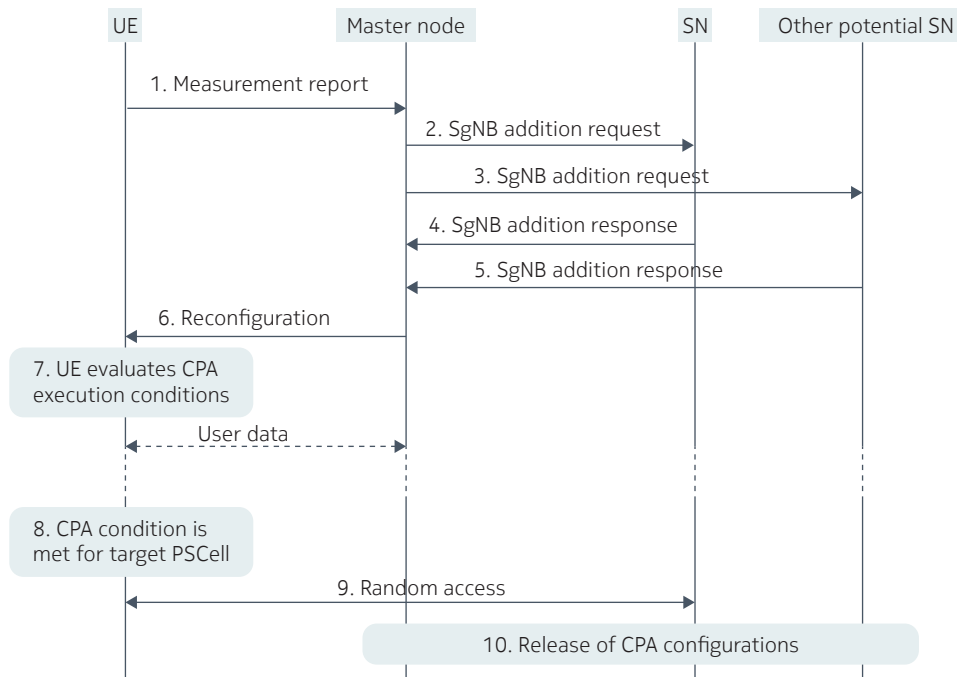
It takes time if the network wants to establish DC for the UE in RRC Connected and use additional radio link towards another node (Secondary Node). Several messages need to be exchanged – between the UE and the Master Node (MN) and also between the MN and the future Secondary Node (SN). Upon receiving an acknowledgement from the SN, MN reconfigures the UE which may then attempt to access the SN by performing random access procedure. As can be noticed, those multiple steps introduce a delay which elapses from the point in time when the radio measurements for candidate SN are to be reported to the MN until the UE is reconfigured. Enhancements have been introduced in 5G to reduce the latency in accessing the SN which in turn allow the UE to benefit as early as possible from the high throughputs of small cells.

Conditional PSCell Addition Change (CPAC) has been defined as a part of 3GPP Release 17. The scope was twofold: to enable conditional setup of SN denoted by Conditional PSCell Addition (CPA) and to allow changing the PSCell in conditional manner denoted by Conditional PSCell Change (CPC). Both solutions (i.e. CPA and CPC) are aimed at reducing the delay to access the SN as the UE can perform random access to PSCell directly after the condition is met, without reporting the measurements to the network and awaiting the corresponding cell change configuration. As stated before, CPC brings additionally another benefit of improving the robustness of PSCell as the UE is prepared in advance with configuration of the candidate target PSCell.

Fig. 14 shows the high-level signaling diagram for CPA [6]. Upon receiving measurement report from UE identifying potential target PSCells, MN triggers the preparation of one or multiple PSCells that are controlled by same or different SNs as shown in steps 2-5. MN provides the configuration of the target

PSCells in step 6 along with CPA execution condition that is associated with each prepared target PSCell. After receiving the configuration, the UE evaluates the CPA execution conditions while continuing the user plane communication with MN. Once the CPA execution condition is met for a prepared target PSCell, the UE executes the corresponding stored configuration and performs access to the target PSCell for quickly establishing DC in step 9, i.e., no further communication with the network (causing delay) before performing access. Finally, MN releases the CPA configurations and associated reserved resources from other prepared SNs.

Figure 14. High-level signaling for conditional PSCell addition (CPA) in MR-DC.



For CPC, there may be two slightly different approaches: MN-initiated CPC and SN-initiated CPC [6]. The former is triggered by the MN without any message exchange with the source SN. The latter is initiated by the source SN and the preparation on the NW's side is performed through the MN. The signaling diagram for both approaches is very similar to CPA of Fig. 14 except that the UE is configured to conditionally change a PSCell rather than adding it for first time. Moreover, there is no specification support for the direct interface between the source SN and candidate target SNs.

Activation of secondary cell group

The NR radio may not be actively used due to e.g., a temporarily unavailable radio connection because of interference or power saving reasons. In these cases, it is desirable to optimize the dual connectivity between LTE and NR such that the NR radio is directed to a power efficient state allowing UE to perform less measurements, thus saving the UE power consumption. This requires having possibility to quickly disable and re-enable the use of LTE-NR dual connectivity. Hence, MR-DC operation might need to be stopped for a while even during RRC Connected (e.g. due to UE power saving), but it would be beneficial to be able to reactivate it fast whenever needed.

In an MR-DC type of deployment, one valid question is what will happen to the second radio leg (i.e. the DC-leg) when the active data transmission (offloading) ends and the need for a second leg would be less. The obvious solution would be to revert back to single connection, where UE will keep the primary leg active which is needed for mobility anchoring (for example) or lower-intensity traffic activity, e.g., keep alive messages. However, this behaviour may not be desirable in case data traffic is bursty and data activity is quite unpredictable. In order to achieve a fast offloading and high peak data rate, the second radio leg needs to be set up fast and activated. On the other hand, when the data traffic is low, then there are clear UE power saving opportunities where UE would benefit of lower measurement effort.

To enable a fast offloading and high peak data rate, there is a trade-off that needs to be achieved between 1) fast offloading in case of secondary radio link activation and 2) UE power saving in case of low activity. Furthermore, compared to release and reconfiguration of the SCG, activating a suspended SCG reduces the latency and signaling overhead. For this, a new mechanism was introduced in MR-DC Rel. 17 work item [20] where gNB can indicate to UE to de-activate at least the NR part of the LTE-NR DC so that UE retains the configuration but does not use it. While SCG is de-activated, UE can be still configured by the network to perform radio link monitoring for primary cell of SCG (PSCell), i.e., to be able to declare SCG RLF (S-RLF), as well as RRC measurements, i.e., to be able to trigger RRM measurement reports. Additionally, UE can be configured to perform beam failure detection to ensure network is aware of beam failure occurrence.

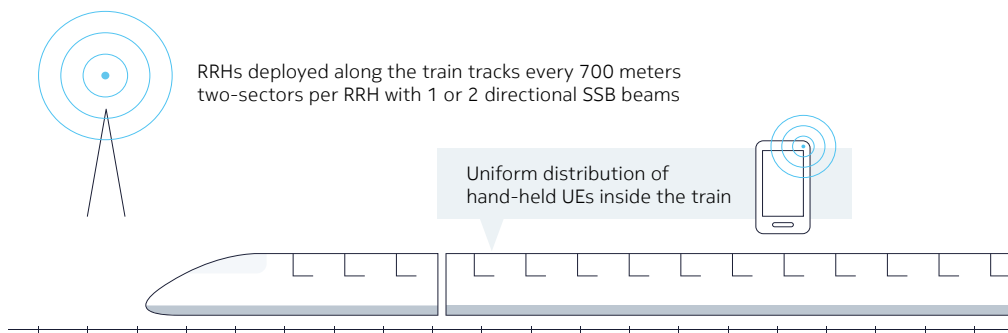
The difference with Rel.16 is the behaviour on PSCell, which in earlier releases was always active, while in Rel. 17 a new mechanism was defined for the case of being deactivated, i.e., to enable SCG de-activation. In practice this means that UE will not monitor the control channel (PDCCH), nor it will have CSI or Power Head Room (PHR) reporting for the deactivated PSCell or SCells of the deactivated SCG. In the event of reception of an SCG activation command from the network, the UE shall access the SCG again. This requires RACH towards the SCG only if PSCell has changed since the last time SCG was active, Timing Advance (TA) timer for the PSCell has expired or UE has detected an RLF or beam failure for the PSCell.

With this mechanism, the UE is able to de-activate the SCG and activate it fast if needed for a fast offloading opportunity, yet at the same time, saving battery if SCG is not in use. Hence, this achieves a positive trade-off between UE power saving and fast setup of DC.

High-speed train scenarios

High-speed train (HST) scenarios are obviously challenging from a mobility point of view due to the high velocity and therefore the needs for frequent handovers of UEs (or cell reselections for IDLE mode UEs). The UE requirements to ensure satisfactory HST mobility performance for FR1 was part of the Release 16 scope with special focus on ensuring sufficiently fast and reliable cell identification requirements, beam management requirements, and cell reselection requirements to support such use cases. Use cases with train velocities of up to 500 km/h with the hand-held UEs inside the train at 3.5 GHz carrier frequency were considered. As for the earlier LTE HST studies, the NR HST FR1 deployments with both so-called open space inter-RRH (Remote Radio Head) distances of 700 meters and Tunnel cases with inter-RRH distances of 300 meters were studied. Considering cases with one or two SSB beams per sector of the RRHs as pictured in Fig. 15, link-level studies confirmed that the higher Doppler shift (at 500 km/h) effects on the RRM measurement accuracy of SSB and CSI-RS did not impact the corresponding accuracy requirements. From advanced dynamic system-level simulations, it is concluded that with proper network configuration, the mobility performance is very good with low probabilities of radio link failures and ping-pong events. In this regard, we recommend that configurations with DRX cycles longer than 160 ms shall not be used for such HST scenarios. Using DRX cycles longer than 160 ms are challenging with the risk of experiencing more failures due to e.g. too late handovers. We also recommend that the SSB based Measurement Timing Configuration (SMTC) period should not be longer than 40ms for FR1 HST scenarios at 500 km/h. Furthermore, applying also short L1 filter length while in DRX is valuable to achieve low radio link failure and time-of-outage rates.

Figure 15: Simple illustration of HST FR1 scenario.



NR HST performance, and the related RF and RRM requirements, for FR2 is currently work in progress as part of Release 17 (to be concluded in 2022). The deployment scenario assumes a carrier frequency of 28 GHz for trains moving at 350 km/h with roof mounted customer-premises equipment (CPE). The CPE can have multiple directional antenna panels, while assuming only one directional antenna panel active at a time. The RRH deployed along the train tracks with 700 meters separation have 1-4 beams. Current simulation results show that good mobility performance is achievable also for such HST FR2 scenarios with hardly any beam failures or handover failures, when using proper network configurations (e.g. of A3 RRM events [4]) and modest use of DRX with recommended DRX cycle length below 80ms. If using longer DRX cycles, the HST FR2 mobility performance is impacted as also observed for HST FR1 cases. The good mobility performance is a result of the use of roof mounted CPE's at the train with directional antenna panel(s) and beamforming at the RRHs which results in good SINR conditions despite the operation at FR2. The Release 17 HST FR2 item is set to be concluded during 2022, considering cases also with Dynamic Point Selection (DPS). The work on HST support in FR2 will also continue in the 3GPP Release 18 [21].

Summary of features

The main mobility and DCCA features in NR are summarized in Table 1 and Table 2, respectively, along with their corresponding 3GPP release and main benefits.

Table 1. Mobility features in NR.

Feature	3GPP Release	Benefits
Baseline handover [2]	Rel. 15	Allow service continuity during handover between cells
Conditional handover [2]	Rel. 16	Improved mobility robustness by reducing the number of radio link and handover failures
Dual-active Protocol Stack (DAPS) handover [2]	Rel. 16	Reduced interruption time during handover down to 2 ms in downlink
Timer T312 [4]	Rel. 16	Early detection of radio link failure and fast recovery of the connection
UE requirements for high-speed train deployment in FR1 [15]	Rel. 16	Defining UE requirements for satisfactory performance in FR1 high-speed train deployment
Conditional handover with SCG configuration [6]	Rel. 17	Enable conditional handover to dual-connectivity connection with target MCG and SCG configurations
UE requirements for high-speed train deployment in FR2 [15, 22]	Rel. 17	Defining UE requirements for satisfactory performance in FR2 high-speed train deployment
L1/2 Centric Mobility [13]	Rel. 18	Reduced interruption time during handover and improved mobility robustness for intra-CU scenarios
Selective Switching of Cell group [13]	Rel. 18	Allow subsequent cell group change without reconfiguration and re-initiation of the preparation
Conditional Handover with Candidate SCGs for CPC/CPA [13]	Rel. 18	Improved access of target SCG during conditional inter-MN handover
Machine learning for Mobility [19]	Rel. 18	Enabling AI/ML functionality in RAN for improved mobility performance

Table 2. DCCA features in NR.

Feature	3GPP Release	Benefits
SCG Addition and Change [6]	Rel. 15	Enable the addition and change of SCG in MR-DC
Intra-SN Conditional PSCell Change [6]	Rel. 16	Improved mobility robustness of primary cell of SCG and fast access to SCG within same SN
Early Measurement Reporting [4]	Rel. 16	Speed up the setup of dual connectivity after establishing or resuming the connection
Direct SCell activation [4]	Rel. 16	Configure SCell as activated in handover, SCell change and SCell addition
TRS-based SCell activation [20]	Rel. 17	Faster SCell access by eliminating the delay in SSB acquisition
Inter-SN Conditional PSCell Change and Conditional PSCell Addition [6]	Rel. 17	Improved mobility robustness of primary cell of SCG and fast access to SCG between different SNs
(De-)Activation of Secondary Cell Group [20]	Rel. 17	Reduced UE power consumption and fast resumption of the SCG
Enhancements for FR2 SCell/SCG measurement reporting [13]	Rel. 18	Fast and reliable setup of FR2 SCell/SCG

Conclusion and outlook

In this white paper, we have outlined the current 5G mobility solutions as well as the related innovations that are coming with 5G-Advanced. Today, 5G offers a powerful toolbox with numerous options for achieving excellent mobility performance under various conditions. That is, achieving efficient cell identification, low probabilities of mobility failures, as well as fast recovery in the rare events of failures. The basic mobility mechanisms rely on solid solutions for network controlled and UE assisted handovers. This is followed by introduction of CHO with higher level of UE autonomous behaviors and DAPS for reduced interruption times. Agile solutions for also quickly taking additional carriers (using CA and/or DC) into use are also introduced, offering attractive benefits for both NSA and SA cases, including cases with usage of FR1 and FR2 bands. Good mobility performance for challenging high-speed train scenarios is also confirmed. In the 5G-Advanced era that starts with Release 18, we foresee even tighter integration of intra-cell beam management and inter-cell mobility functionalities. That will likely include enhanced mechanisms and procedures of Layer-1 and Layer-2 based inter-cell mobility, aiming at zero interruption time handovers also for single-transceiver devices without having to rely on DAPS. Finally, new AI/ML functionalities will be introduced in RAN which will help to improve further the performance of mobility and DCCA and pave the way for more advanced and intelligent mechanisms.

In later stages of 5G-advanced (Rel. 19+), it is expected to introduce new mobility solutions that do not optimize a specific mobility metric but rather multiple ones such as low interruption time, high mobility robustness and throughput. This is particularly relevant for services requiring at the same time high reliability and throughput. For example, make-before-break (such as DAPS) and/or RACH-less features may be introduced for conditional handover or enhancements for DAPS to inter-work with CA/DC may be specified. Another potential area is to improve further the mobility robustness in FR2 considering different multi-panel UE architectures. In particular, the network and/or the UE can be enabled to consider Maximum Permissible Emission (MPE) events (causing UL degradation for UEs transmitting at full power) in mobility related decisions such as handover triggering or admission control. Furthermore, given that the performance of mobility solutions highly depends on the proper configuration of the mobility parameters, it is foreseen to introduce the capability of configuring and changing in a slim way the handover related parameters per group of beams. This would provide the network operator with much more degrees of freedom for controlling the mobility performance in different areas of the cell.

Future releases of 5G-advanced might aim beyond DC and introduce Multi-Connectivity (MC) where the UE may be configured with more than one active secondary node. This would enable the UE to achieve extremely high user throughputs and to perform secondary node change without having to suffer from interruption time. This would be particularly relevant for services like XR/AR, cloud gaming or 4K streaming which may be only provided by secondary node due to high available bandwidth in FR2. In addition, multi-connectivity would enable the network operator to leverage the spectrum available in multiple frequencies to provide extreme mobile broadband connection and high-end user experience.

References

1. 5G-Advanced: Expanding 5G for the connected world, White paper, Nokia: 5G-Advanced: Expanding 5G for the connected world
2. 3GPP TS 38.300 V16.8.0, NR; NR and NG-RAN Overall Description; Stage 2 (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.300/38300-g80.zip
3. 3GPP TS 38.304 V16.7.0, User Equipment (UE) procedures in Idle mode and RRC Inactive state (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.304/38304-g70.zip
4. 3GPP TS 38.331 V16.7.0, NR; Radio Resource Control (RRC) protocol specification (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.331/38331-g70.zip
5. 3GPP TS 38.321 V16.7.0, NR; Medium Access Control (MAC) protocol specification (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.321/38321-g70.zip
6. 3GPP TS 37.340 V17.0.0, Multi-connectivity; Stage 2 (Release 17) https://www.3gpp.org/ftp/Specs/archive/37_series/37.340/37340-h00.zip
7. 3GPP TS 38.401 V16.8.0, NG-RAN, Architecture description (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.401/38401-g80.zip
8. 3GPP TS 38.463, V16.8.0, NG-RAN, E1 Application Protocol (E1AP) (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.463/38463-g80.zip
9. R1-2007151, Moderator summary for multi-beam enhancement: EVM, 3GPP TSG RAN WG1 #102-e e-Meeting, August 17-18, 2020.
10. 3GPP TS 22.261 V18.5.0, Service requirements for the 5G system; Stage 1 (Release 18), https://www.3gpp.org/ftp/Specs/archive/22_series/22.261/22261-i50.zip
11. 3GPP TR 38.838 V17.0.0 (2021-12), Study on XR (Extended Reality) Evaluations for NR (Release 17), https://www.3gpp.org/ftp/Specs/archive/38_series/38.838/38838-h00.zip
12. Stanczak, J., Karabulut, U. and Awada, A., Conditional Handover in 5G: Principles, Future Use Cases and FR2 Performance, doi: 10.48550/ARXIV.2204.01283, 2022.
13. RP-213565, New WID on Further NR Mobility enhancements, 3GPP TSG RAN Meeting #94e Electronic Meeting, Dec. 6-17, 2021, https://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_94e/Docs/RP-213565.zip
14. 3GPP TS 38.323 V16.6.0, NR; Packet Data Convergence Protocol (PDCP) specification (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.323/38323-g60.zip
15. 3GPP TS 38.133 V17.4.0, NR; Requirements for support of radio resource management (Release 17), https://www.3gpp.org/ftp/Specs/archive/38_series/38.133/38133-h40.zip
16. 3GPP TR 38.901 V16.1.0, Study on channel model for frequencies from 0.5 to 100 GHz (Release 16), https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/38901-g10.zip
17. Iqbal, S. B., Awada, A., Karabulut, U., Viering, I., Schulz, P., & Fettweis, G. P., On the Modeling and Analysis of Fast Conditional Handover for 5G-Advanced, doi: 10.48550/ARXIV.2204.06909, 2022.
18. RP-213602, New WI: Artificial Intelligence (AI)/Machine Learning (ML) for NG-RAN, 3GPP TSG RAN Meeting #94e, Electronic Meeting, Dec.6 – 17, 2021.
19. 3GPP TS 37.817 V1.1.0, E-UTRA and NR; Study on enhancement for Data Collection for NR and ENDC (Release 17), https://ftp.3gpp.org/Specs/archive/37_series/37.817/37817-110.zip
20. RP-201040, Revised WID on Further Multi-RAT Dual Connectivity enhancements, 3GPP TSG RAN Meeting #88e, Electronic Meeting, June 29-July 3, 2020.
21. RP-220985, New WI: Enhanced NR support for high speed train scenario in frequency range 2 (FR2), RAN#95-e.
22. 3GPP TR 38.854, V17.0.0, NR support for high speed train scenario in frequency range 2 (FR2).

The background image is a 3D rendering of a futuristic industrial warehouse. On the left, there are tall stacks of cardboard boxes on yellow pallets. The floor is grey with white dashed lines and a series of white triangles pointing towards the right. In the foreground, there are several yellow and black striped cylindrical objects. In the background, there is a blue door and a yellow staircase. The ceiling is high with a grid of white beams and blue structural elements.

NOKIA

Release 16

Towards industrial applications

5G Non-Public Networks

Release 15

Release 16

Release 17

Contents	
Introduction	3
Motivation for SNPN in 5G system	4
Standalone Non-Public Networks	5
Deployment models	6
UE identifiers for SNPN	8
NPN authentication mechanisms	8
Service continuity between SNPN and PLMN	9
Onboarding in SNPN	10
Voice and emergency services support in SNPN and PNI-NPN	11
PNI-NPN and network slicing	13
Conclusions	13
Further reading	13

Introduction

3GPP introduced the concept of Non-Public Networks (NPN) in 5G in Release 16. Two main deployment models were defined: Standalone Non-Public Network (SNPN) and Public Network Integrated Non-Public Network (PNI-NPN). Non-Public Network is basically the terminology defined by 3GPP for deploying a 5G private network. The basic idea was to enable seamless deployments of private networks, either as standalone networks or by integration into a public mobile network, while leveraging the strengths of mobile networks (support for high quality, reliability and seamless mobility) and nearly all 5G features that were standardized in the context of public networks. In this white paper, we highlight the motivation for NPN, describe different deployment models and explain the features that were introduced by 3GPP Releases 16 and 17.

Motivation for SNPN in 5G system

Private networks can be deployed with LTE today. Such networks are deployed using unlicensed or shared spectrum, or dedicated spectrum granted by MNOs for private network use. Regulators have allocated new spectrum for local and private use like 3.7-3.8 GHz in Germany, 3.8-4.2 GHz in UK or 28.2-28.3 GHz in Japan that can be leveraged for 5G private network deployments. However, these deployments are facing some challenges without a globally standardized solution:

1. Each network needs its own PLMN ID (Public Land Mobile Network Identifier) comprised of a Mobile Country Code (MCC) and a Mobile Network Code (MNC) uniquely identifying the network. However, the available number of unique mobile network codes per country is insufficient as it was not designed for a large number of private networks. In addition, there are strict rules on how an MNC in a country is allocated to a network operator by regulatory authorities; some countries allocate MNCs only to networks serving public subscribers.
2. ITU allocated MCC '999' for global use to private networks. This was an important step forward to enable private network deployments, but still the MNC can be a wild card. Thus, the PLMN ID (MCC + MNC) is not globally unique as collisions between two enterprises using the same self-assigned MNC value are possible.
3. When such a network identifier is not unique, this can lead to mobile devices (in 3GPP terms User Equipment or UE) not receiving services. Mobile devices may try to connect to a network that uses the same ID as its "home network" with which it has a subscription. The network using the same ID as the home network will reject the connection attempt and the device remains disconnected. It can also degrade performance of private networks as unauthorized connection attempts to a private network consume network resources leading to higher network load. This can also degrade the performance of critical applications requiring dedicated private network resources like time-sensitive services.
4. Support for devices without USIM and enablers needed to support device onboarding (i.e., enablers for dynamic provisioning of credentials) to ease deployment of private networks.

3GPP specified SNPN and PNI-NPN only for 5G NR and not for LTE E-UTRAN. In case of LTE E-UTRAN, specific enhancements needed for private networks were specified outside 3GPP on top of 3GPP specifications, namely in MulteFire (i.e., for unlicensed), OnGo alliance for CBRS bands in the US.

Standalone Non-Public Networks

Standalone Non-Public Network (SNPN) is a feature specified by 3GPP [1] to alleviate the above-mentioned challenges. With this feature, a so-called Network Identifier (NID) was introduced that can be used in conjunction with any PLMN ID to identify an SNPN. The PLMN ID and NID together refer to the SNPN identifier (SNPN ID). 3GPP allows multiple NID assignment modes [2] to enable varying deployment models:

1. Coordinated assignment mode: network identifiers are assigned using one of the following two options:
 - a) Globally unique NID independent of PLMN ID (Assignment mode 0): The NID is assigned so that it is globally unique independently of the PLMN ID used. In this case, NID uses assignment mode 0. In this case the NID value consists of NID PEN (Private Enterprise Number) and the NID code. The NID PEN is issued to the SNPN provider by the Internet Assigned Numbers Authority (IANA) in its role as the private enterprise number administrator, as maintained at here. The NID code identifies an SNPN within the service provider domain.
 - b) Combination of PLMN ID and NID are globally unique (Assignment mode 2): The NID is assigned so that the combination of NID and PLMN ID is globally unique. In this case, NID uses assignment mode 2. This is for scenarios where the PLMN ID can be regionally shared by private networks (e.g., a shared PLMN ID assigned for use by private networks using the CBRS band) and the NID applies within the scope of the shared PLMN ID.
2. Self-assigned SNPN ID (Assignment mode 1): the SNPN network identifier is chosen by the respective SNPN at the time of deployment; thus, it may not be unique. However, it must use a different numbering space than the coordinated assignment NIDs. In this case, NID uses assignment mode 1.

The SNPN feature enables 5G deployments in licensed, shared (or lightly licensed) and unlicensed spectrum. The NG-RAN has the capability to broadcast both PLMN ID and NID allowing UE(s) in SNPN access mode to consider SNPN network identifiers for network selection. For that purpose, just as with preferred PLMN ID(s), preferred SNPN ID(s) are stored on the device. At the same time, the SNPN RAN can broadcast indicators that prevent unauthorized UE(s) from performing network access.

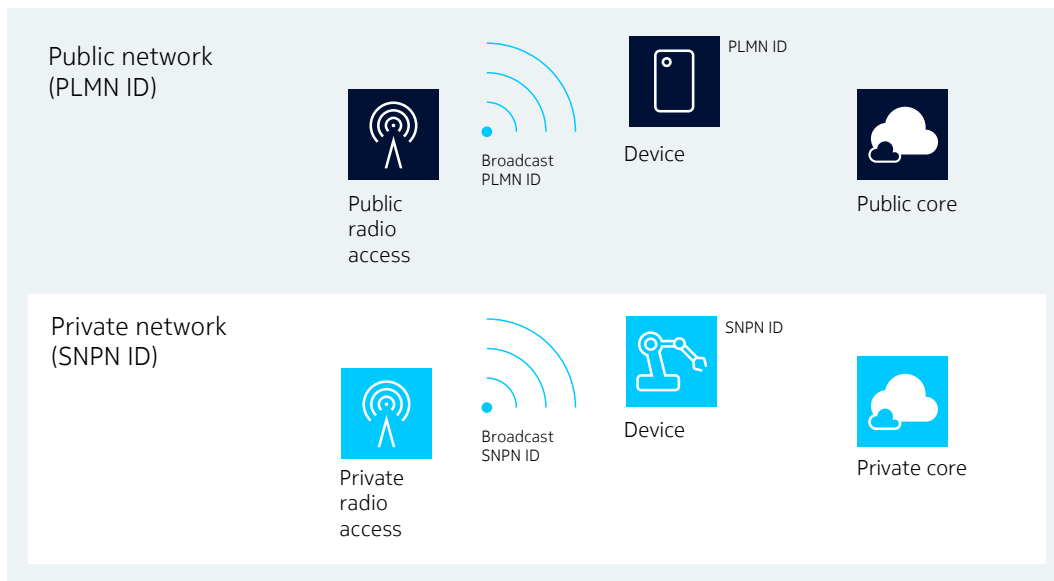
At the same time the UE and SNPN can use 5G AKA, EAP-AKA' or any other key-generating EAP authentication method for mutual authentication. Especially in case of private networks, use of EAP-TLS based on certificates is a potential alternative to AKA-based methods defined by 3GPP.

Deployment models

Possible deployment models enabled with Non-Public Networks are:

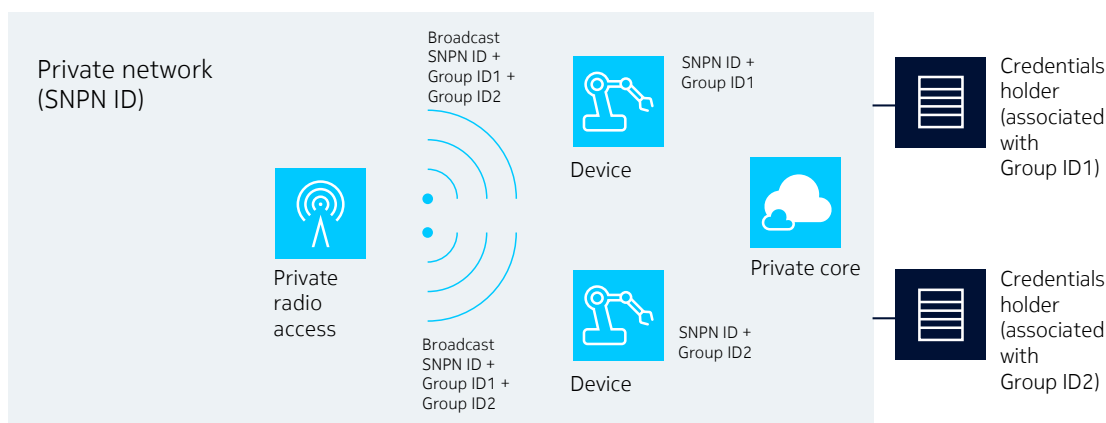
1. Standalone Non-Public Network (SNPN): An SNPN is not relying on network functions like radio access or core network functions provided by a public network operator (PLMN). An SNPN uses PLMN ID plus NID as a network identifier (SNPN ID). The following figure illustrates this type of deployment.

Figure1. Standalone Non-Public Network (SNPN)



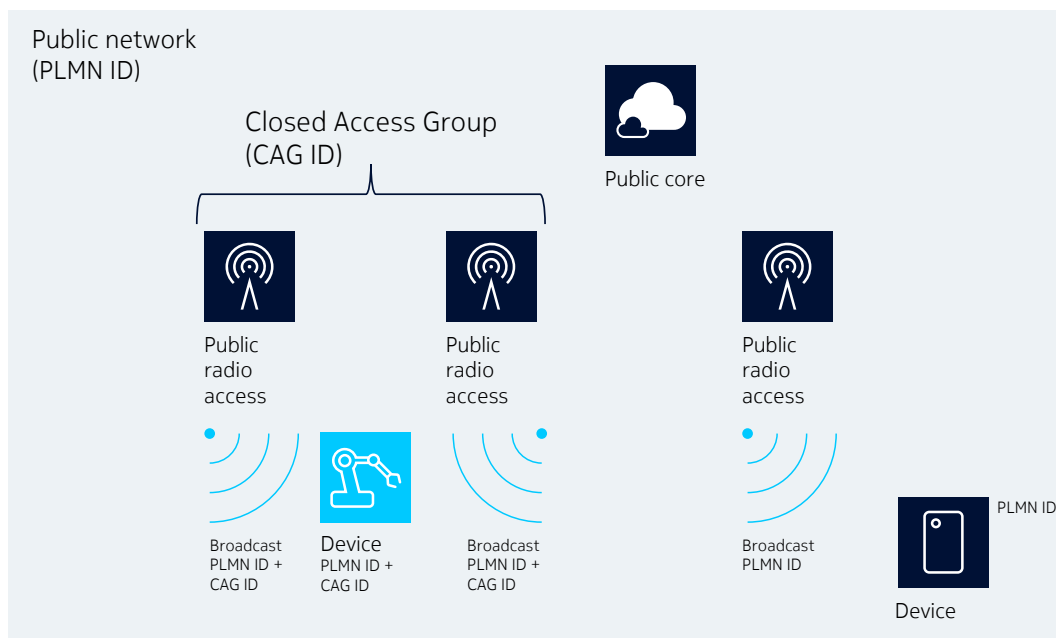
2. Standalone Non-Public Network (SNPN) with credentials assigned by a Credentials Holder (CH): a CH may deploy UDM/AUSF or an AAA server. This option enables support for UE(s) accessing an SNPN using credentials assigned and owned by a CH separate from the SNPN — i.e., the CH performs primary authentication of the UE(s), enabling neutral host type deployments. GIN (Group ID for Network Selection) identifies a group of one or more CH entities.

Figure 2. SNPN with credentials assigned by credentials holder (CH)



3. Public Network Integrated Non-Public Network (PNI-NPN): This option enables private network deployments wholly integrated within a PLMN. A PNI-NPN uses PLMN resources in which the PLMN restricts use of certain cells for NPN subscribers. Access to radio cells of the PLMN is restricted to NPN users by the Closed Access Group (CAG) concept with CAG ID identifying CAG cells. A CAG identifies a group of subscribers who are permitted to access one or more CAG cells associated to the CAG. For the UE, the allowed CAG list indicates the list of CAG cells in which it is authorized to access. PLMN can also allocate a specific slice or dedicate a certain Data Network Name (DNN) to separate traffic of NPN users from the traffic of other PLMN subscribers. As usual, the PLMN ID identifies the network. A CAG cell broadcasts one or multiple CAG identifiers per PLMN. Network selection and reselection are based on PLMN ID. Cell selection, reselection and access control are based on CAG ID. The CAG cell broadcasts information so that only UEs supporting the CAG feature are accessing the cell.

Figure 3. Public network integrated NPN using CAG



Following are the features that are not supported for SNPNs in the 3GPP specifications: EPS interworking, eCall services and home routed roaming. Regulatory requirements such as emergency services, public warning system (PWS) are specified in 3GPP also for SNPNs. All the features specified for regular PLMN are inherited for PNI-NPN, where necessary specific enhancements need for PNI-NPN using CAG have been identified.

UE identifiers for SNPN

A subscriber of an SNPN is identified by a SUPI (Subscription Permanent Identifier) containing an IMSI or a network-specific identifier taking the form of a Network Access Identifier (NAI). When it is identified by an NAI, the realm part of the NAI includes the NID of the SNPN.

When accessing an SNPN using credentials owned by a CH, the SUPI must also contain an identification of the CH. If the CH is an SNPN, then a network-specific identifier based SUPI should be used. If the CH is a PLMN, then an IMSI-based SUPI should be used.

NPN authentication mechanisms

3GPP specifies usage of the EAP authentication framework for primary authentication that was generalized also for SNPN deployment scenarios. EAP-based authentication can be used by an SNPN that may want to support alternatives to the 5G AKA authentication method.

The EAP authentication framework defines the EAP server, the EAP peer and the EAP authenticator. The EAP server is the entity that terminates the EAP authentication with the EAP peer. The EAP peer is the end of the link responding to the EAP authenticator. The EAP authenticator is the other end of the link initiating the EAP authentication. Mapping this to the 5GS architecture, the authentication server AUSF acts as the EAP server, the UE takes the role of the peer, and the SEAF, being a function of the AMF in the serving network, takes the role of a pass-through authenticator.

When 5G AKA is not used as the preferred authentication method, EAP-based authentication can be used instead. In general, EAP supports a variety of concrete authentication methods. 5GS has restricted the usage of EAP to key-generating EAP authentication methods.

5GS distinguishes between NPN authentication in SNPN and PNI-NPN authentication as follows:

- Since access to a PNI-NPN is based on a regular PLMN subscription, 5G AKA and EAP-AKA authentication methods are mandatory to support the UE and the network. The UE is mandated to support a USIM deployed on a UICC. In case of SNPN, 3GPP has lifted these requirements, thus these AKA methods are optional for the UE and SNPN to support to enable different deployment models. An SNPN deployment might still be based on USIMs/UICCs in combination with AKA. Alternatively, other EAP methods like EAP-TLS could also be used and this can be implemented on the device without a USIM. In general, there is no mandatory authentication method defined for UE(s) in SNPN deployment scenarios.

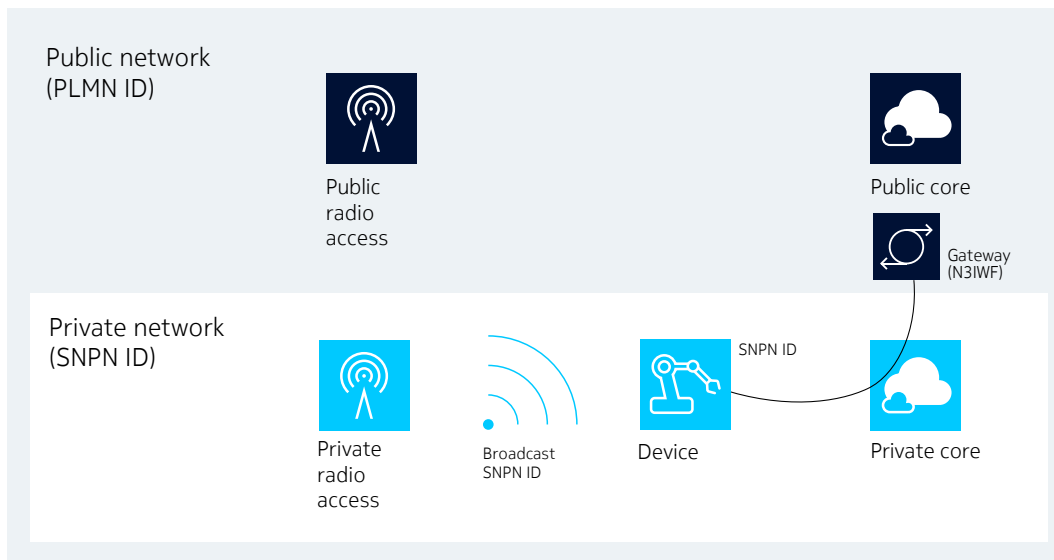
When a key-generating authentication method is used, the serving network identifier (SN ID) is used within the key derivation process as an input parameter to allow verification of the serving network by the home network operator requesting authentication of the UE. While in PLMN cases, the SN ID contains the PLMN ID only, an SNPN also uses the Network Identifier (NID) for identification of the private network. Thus, for SNPN, the SN ID needs to include PLMN ID and NID for key derivation.

Service continuity between SNPN and PLMN

3GPP specifications enable service continuity between SNPN and PLMN even for scenarios where SNPN and PLMN do not have any service level agreement (SLA). 3GPP specified the following options for service continuity to be supported for a given UE between PLMN and SNPN:

1. The ability for the UE to access PLMN services (in this context, PLMN is referred to as overlay network) while being connected to the SNPN RAN and 5G Core (in this context, SNPN is referred to as underlay network)
2. The ability for the UE to access SNPN services (in this context, SNPN is referred to as the overlay network) while being connected to PLMN RAN and 5G Core (in this context, PLMN is referred to as underlay network).

Figure 4. Accessing PLMN services via SNPN



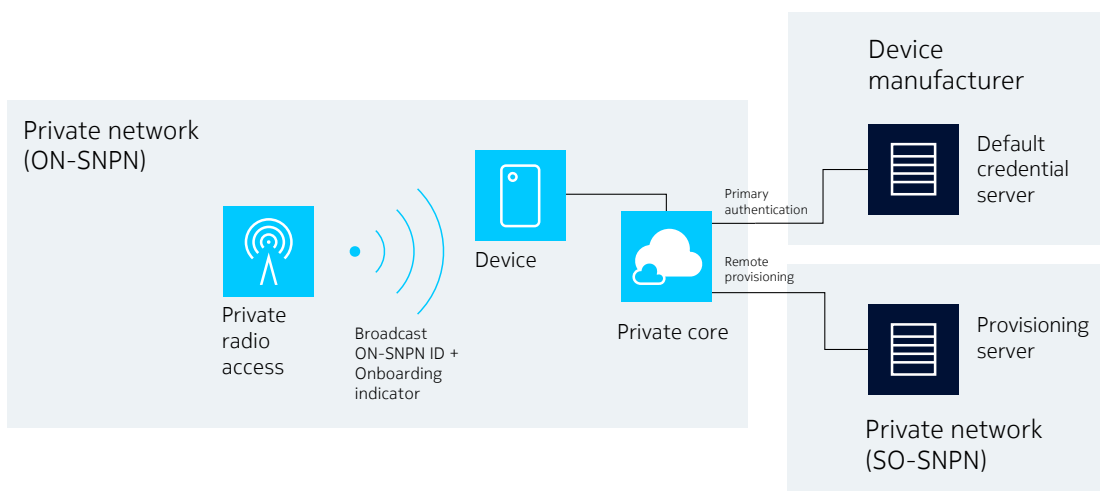
The UE uses the credentials needed to access and register with the respective underlay network and just obtains IP connectivity from the underlay network. It uses IP connectivity and credentials assigned by the overlay network to register with the overlay network via a secure tunnel which is terminated at a gateway in the overlay network (called N3IWF). Using the secure tunnel, the UE can obtain services from the overlay network. Service continuity is also enabled between SNPN and PLMN since the IP anchor is in the overlay network and does not change. Seamless service continuity (with full data integrity) is possible if the UE is dual radio capable and can camp on both networks, thus enabling seamless service continuity — although this is not an essential feature.

Onboarding in SNPN

Onboarding and remote provisioning introduce the capability to allow access and dynamic provisioning for devices in a network. This allows devices to access an SNPN (called Onboarding SNPN or ON-SNPN) without having valid credentials and subscription data for connecting to the SNPN. For example, a device manufacturer can pre-configure devices with default credentials like certificates and ship them to an enterprise where these devices can access (or onboard to) the network. Once the devices are authenticated, referred to as primary authentication, via the manufacturer's infrastructure using a default credential server (DCS) the devices can establish a connection in the ON-SNPN to be remotely provisioned with the subscription and credential data of the actual subscription owner. The subscription owner is referred to as subscription owner SNPN or SO-SNPN. ON-SNPN and SO-SNPN can be the same or different. The ON-SNPN advertises its capability to onboard devices by broadcasting a respective indicator via the radio access network.

3GPP enables establishment of limited user plane connectivity for the device to establish an IP connection with the provisioning server (PVS) for provisioning of devices, however, the actual provisioning protocol between the server and the device is considered over the top of the 3GPP system.

Figure 5. Onboarding and remote provisioning



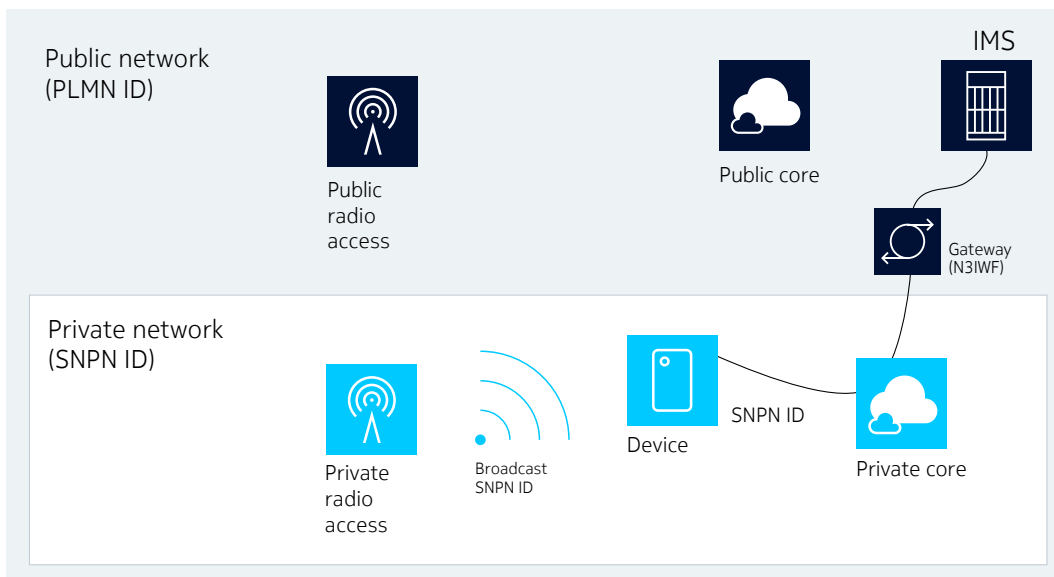
In case of a PNI-NPN, it is assumed that the PLMN providing resources for the NPN has provisioned the device using credentials essential for registering in the PLMN, while credentials for slice-specific authentication can be remotely provisioned after successful primary registration using the default credentials.

Voice and emergency services support in SNPN and PNI-NPN

With 3GPP Rel-17 voice and emergency services support in an SNPN was introduced. Two main deployment options are considered, although other implementation options might exist in real-life deployments:

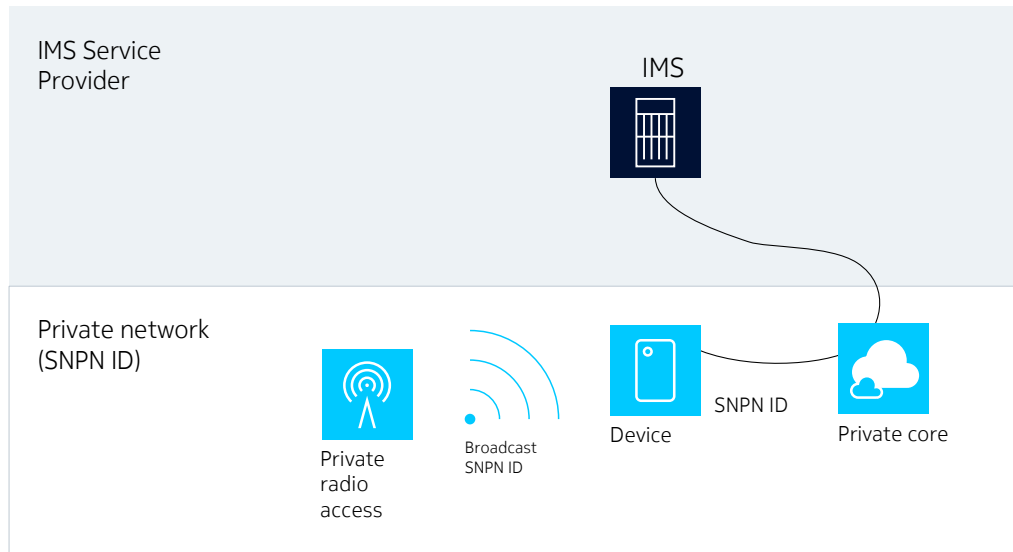
1. SNPN deploys its own IP Multimedia Subsystem (IMS) including Home Subscriber Server (HSS) function to store IMS credentials for its subscribers. The IMS infrastructure can be used for internal voice calls only, or it can also allow SNPN users to make and receive external calls that require, for example, a 'trunking' interface to a public voice provider like a PLMN or fixed network operator. IMS authentication can use credentials stored on the UICC of the device (if available) or can leverage so-called IMS credentials (IMC), which might be a simple username and password pair securely stored on the device.
2. SNPN does not deploy its own IMS infrastructure but relies on an IMS-based voice solution offered by a PLMN or any other third-party entity. In this case, the IMS service provider is responsible to route internal and external calls to or from SNPN devices. Two architecture options were described by 3GPP:
 - a. VoWLAN-like access to IMS services where the device, which is camping on the SNPN, must first establish a secure tunnel to the IMS provider network. The tunnel endpoint is the Non-3GPP Interworking Function (N3IWF). The device can be pre-configured with a tunnel endpoint address and necessary credentials to establish the secure tunnel. The device registers in the IMS and initiates or receives voice calls via the tunnel. Also, media data are sent via the tunnel to and from the device. Emergency calls might not be possible with this deployment option as the IMS network may not be able to retrieve the exact location of the device, which is usually a requirement when providing emergency services.

Figure 6. SNPN UE obtains voice services from PLMN



- b. SNPN provides IMS-related information like the Proxy Call Session Control Function (P-CSCF) address to devices camping on the SNPN, allowing the devices to register at the third-party IMS service provider network. Signaling and media data are directly exchanged between the device in the SNPN and the IMS service provider network. IMS subscription data, which allows SNPN users to be authenticated in the IMS network, are stored in the HSS of the IMS service provider. The IMS service provider is responsible to route internal and external voice calls to and from SNPN devices.

Figure 7. SNPN UE obtains voice services from IMS service provider



The main functional impact to support emergency services within 5GS in an SNPN are as follows:

1. The AMF in an SNPN needs to indicate emergency services support within the registration except for UEs in SNPN-access mode
2. The radio access network of the SNPN needs to broadcast an indicator that the cell supports emergency services allowing UEs in a limited service state to access the network
3. The UE supports the ability to select an SNPN for emergency services. Even if a UE does not have a subscription for an SNPN, if it is in limited-service state, it can determine whether the SNPN cell supports emergency services from the broadcast indicator mentioned above.

Voice support in a PNI-NPN can be provided via IMS since 3GPP Rel-16. The assumption is that the PLMN enables IMS voice and emergency services support for all NPN subscribers within the specific slice and CAG cells reserved for the NPN. Specifically, the P-CSCF address is provided from network functions belonging to the slice to the UE(s) authorized to access the NPN. The IMS subscription data for NPN users are stored in the HSS of the PLMN.

PNI-NPN and network slicing

When a PNI-NPN is made available via a PLMN using an NPN-specific slice, the UE needs to have a valid PLMN subscription to access the PNI-NPN. Thus, a PNI-NPN is one important use case for network slicing in 5G. When using the feature ‘Network Slice-Specific Authentication and Authorization Access’ (NSSAA) the NPN is enabled to allow access to the NPN specific slice for its authenticated users only, i.e., slice specific authentication is performed by the NPN acting as the slice owner.

As network slicing does not prevent UEs from trying to access the network in areas where the UE is not allowed to use the network slice allocated to the PNI-NPN, closed access groups may optionally be used to apply additional access control.

The CAG concept is used in case of PNI-NPNs to prevent UE(s), which are not allowed to access the NPN via the associated cell(s), from automatically selecting and accessing the associated CAG cell(s). CAG is used for access control, authorization at cell selection and configured in the user’s subscription data as part of the mobility restrictions data.

Conclusions

Non-Public Networks in 5G is a powerful concept enabling deployment of private networks either standalone or integrated in a PLMN, depending on the needs and capabilities of an enterprise. Furthermore, SNPN and PNI-NPN inherit nearly all key features that were standardized in the context of public networks such as network slicing, URLLC support and network sharing.

It is also possible for a mobile network operator to use the NPN concept to build its own Non-Public Networks, e.g., using its PLMN ID and allocating a certain NID for an NPN deployed in a stadium, port or on an enterprise’s premises. Thus, the public operator does not need to request a new PLMN ID when deploying networks or sub-networks in certain regions or for certain use cases.

Potential enhancements to SNPN and PNI-NPN are under consideration for 3GPP Rel-18 as part of 5G advanced and the content for Rel-18 is expected to be finalized by the end of December 2021.

Further reading

[1] 3GPP TS 23.501: “System Architecture for the 5G System”.

[2] 3GPP TS 23.003: “Numbering, addressing and identification”.

The evolution of 5G New Radio positioning technologies

Contents

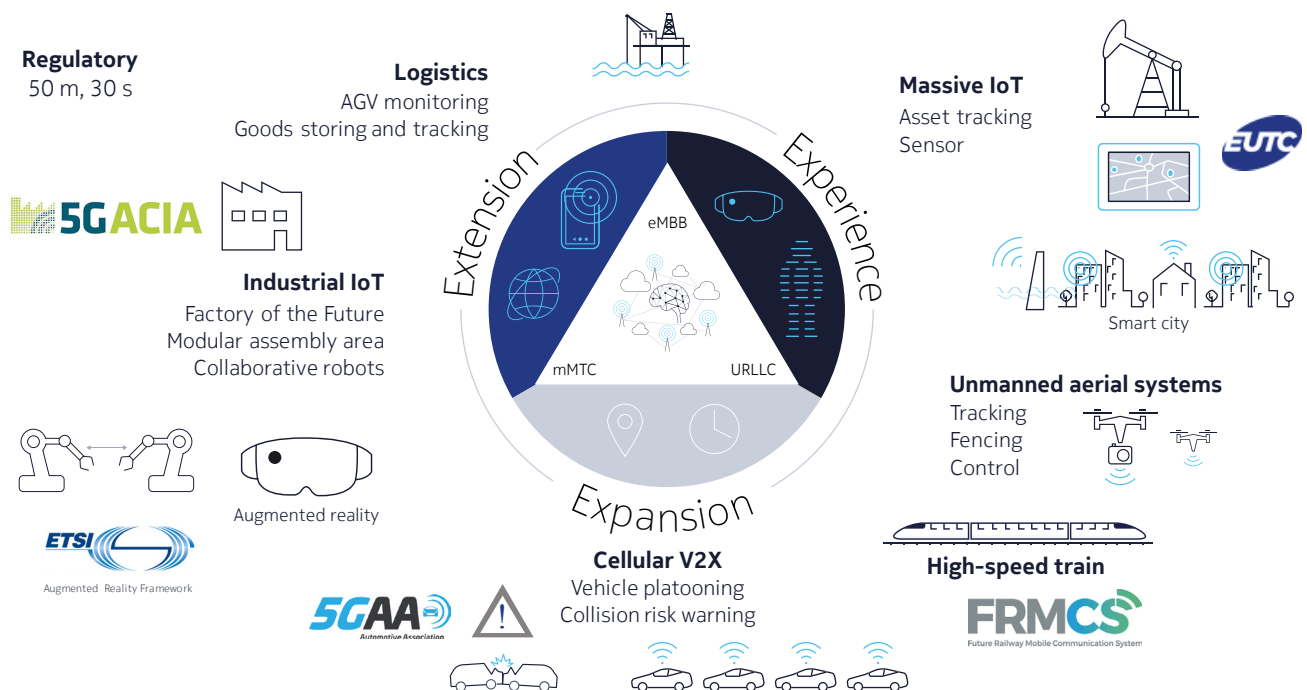
Introduction	3
Use cases and requirements	4
Tools in the toolbox – Release 16 5G NR	5
Refining the tools for the Industrial IoT – Release 17	5
Release 18 and beyond	6
Summary	7

Introduction

Positioning (aka localization) technology utilizing 3G, 4G and Wi-Fi systems has existed for at least the last 20 years. 5G New Radio (NR) Positioning is now evolving as a key feature that enables the operator of a 5G network to position devices for both indoor and outdoor applications with much better accuracy and reliability than those of previous generations.

A wide variety of use cases require positioning services—from locating emergency callers’ locations to autonomous driving to tracking automated forklifts moving in a factory. Many of these use cases, which will expand the applicability of 5G networks beyond traditional mobile broadband, require accurate positioning in order to function. 5G networks offer the exciting possibility to deploy a single technology that addresses both the positioning and data needs of these use cases in an integrated solution, all in an ultra-reliable manner.

Figure 1. Positioning use cases examples



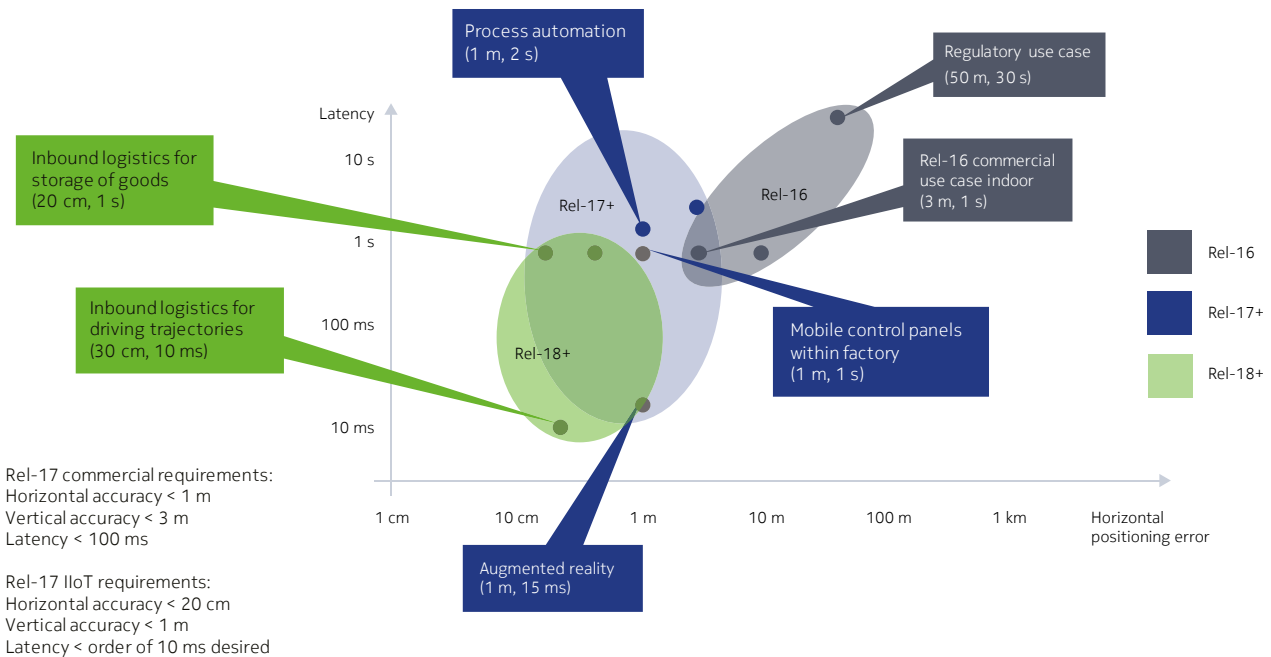
So far 5G-based positioning relies on timing-based, angle-based, power-based or hybrid techniques to locate users. Improved positioning accuracies can be achieved with 5G technologies by taking advantage of the wide bandwidth signals in NR, beamforming capabilities that come from the large number of antenna elements in massive multiple-input, multiple output (MIMO) networks, and the denser 5G network deployments.

Positioning using 5G can complement or act as an alternative to traditional global navigation satellite system (GNSS)-based positioning in many use cases where GNSS coverage is not available (e.g., indoor or dense urban areas). 5G positioning can also be fused with GNSS positioning to improve accuracy, and 5G networks can be used to deliver assistance data in high-accuracy GNSS techniques.

Use cases and requirements

With 5G positioning in a diverse set of use cases comes a diverse set of requirements. Figure 2 shows the evolving requirements as they map to 3GPP Releases. The two key metrics that we focus on here are the latency and horizontal positioning error, but these are certainly not the only metrics. For regulatory use cases, the accuracies required are in the order of tens of meters, while for additional Release 16 (Rel-16) commercial use cases, horizontal accuracies of less than 10 meters for 80% of users in outdoor applications were required. The targeted performance in Rel-17 for Industrial Internet of Things (IIoT) use cases is 20 centimeters horizontal positioning accuracy and less than 1 meter vertical accuracy for 90% of devices.

Figure 2. Positioning requirements across various 3GPP Releases



Along with accuracy, latency is an important requirement; for mobile users, high accuracy is not useful if the estimation is of a past location far from the current location. Regulatory use cases only required latencies less than 30 seconds but latencies down to 10 milliseconds are needed for some IIoT use cases such as tracking driving trajectories on a factory floor.

In addition to accuracy and latency, the coverage area of the localization service is an important characteristic. Unlike Wi-Fi, which only provides limited indoor coverage for the localization service—or GNSS, which is restricted to outdoor applications—5G offers a continuous localization service indoors and outdoors across a wide area network. The accuracy, which depends on the density of radios, can vary across the wide area coverage, e.g., high precision for indoor localization with a high density of radio access points, and coarse precision in the outdoors with a lower density of base stations.

Tools in the toolbox – Release 16 5G NR

During Rel-15, some basic positioning features were introduced to NR that covered the non-standalone deployment case (e.g., using LTE carriers for positioning) and RAT-independent deployment (i.e., non-NR-based methods such as GNSS or Wi-Fi). The wide variety of use cases (e.g., indoor and outdoor), deployment scenarios (e.g., mmWave and sub-6 GHz), and positioning requirements that are covered by 5G led to a diverse set of RAT-dependent techniques being introduced in Rel-16 NR.

The RAT-dependent techniques in NR can be thought of as adding tools to the NR toolbox, which can be used to meet positioning requirements in different use cases and scenarios. In Rel-16, six different RAT-dependent techniques were introduced:

- NR Enhanced cell ID (E-CID)
- Downlink time difference of arrival (DL-TDOA)
- Uplink time difference of arrival (UL-TDOA)
- Downlink Angle of departure (DL-AoD)
- Uplink Angle of arrival (UL AoA)
- Multi-cell round trip time (multi-RTT).

These techniques rely on time, angle and power measurements at either the device side, the network side or a combination of multiple measurements. Some of the methods (e.g., DL-AoD) were introduced to take advantage of the beamforming capabilities of NR and may be better suited for mmWave deployments. Other techniques such as multi-RTT were introduced, based on learning from LTE to address positioning errors caused by synchronization offsets between base stations.

Similar to LTE, NR supports network-based positioning where the measurements are reported by the base stations and/or the users to a central location server, which then computes the position estimate. In addition, in NR, UE-based positioning was introduced for downlink techniques; this is where the user calculates the position estimate based on the self-conducted measurements in conjunction with base station locations, which are provided to the user by the network.

Refining the tools for the Industrial IoT – Release 17

Beyond the emergency call use case, which was the focus of Rel-16, NR positioning is an exciting feature for many future network applications. Many of the most promising (and demanding) use cases are related to the IIoT. To address these use cases in indoor factories, enhancements to the Rel-16 NR positioning solutions are currently being developed by 3GPP. The full scope of the enhancements is still open, but specification work is likely to be done in the following areas:

- Improved accuracy: Sharpening the tools available in Rel-16 to improve the achievable accuracy is underway, potentially including non-line-of-sight identification and mitigation as well as positioning using carrier aggregation.
- Latency reductions: Enhancements to the signaling and procedures are under discussion to meet the latency requirements of at most 100 milliseconds.
- Positioning integrity: Positioning integrity refers to the level of trust associated with the positioning information.
- Scalable and efficient solutions: Rel-17 will support users in an inactive state performing positioning. This enhancement improves the device and network efficiency while allowing networks to scale to larger numbers of users being positioned simultaneously.

Nokia is currently enabling the ARENA2036, which is an experimental future factory to trial innovations for industrial automation with industry partners—working with cutting edge 5G technology, edge cloud and high-speed optical data transmission with an emphasis on precise terrestrial positioning deployed indoors. Nokia is leveraging the realistic industrial testbed of ARENA2036 to evaluate its Rel-17 NR positioning solution. Tracking of materials, products, and the production environment itself is critical to enabling the Fourth Industrial Revolution.

Release 18 and beyond

Beyond Rel-17, and even as 6G takes shape, it is very likely that even further enhancements and expansions of NR positioning will be developed, and positioning and location awareness will be a core feature of these future networks. Releases 16 and 17 have enabled an exciting baseline of solutions that will continue to be improved to address the expanding 5G ecosystem. While it is too early to know which use cases will be addressed in which 3GPP Release, 5G NR may be making enhancements in a few other exciting areas such as:

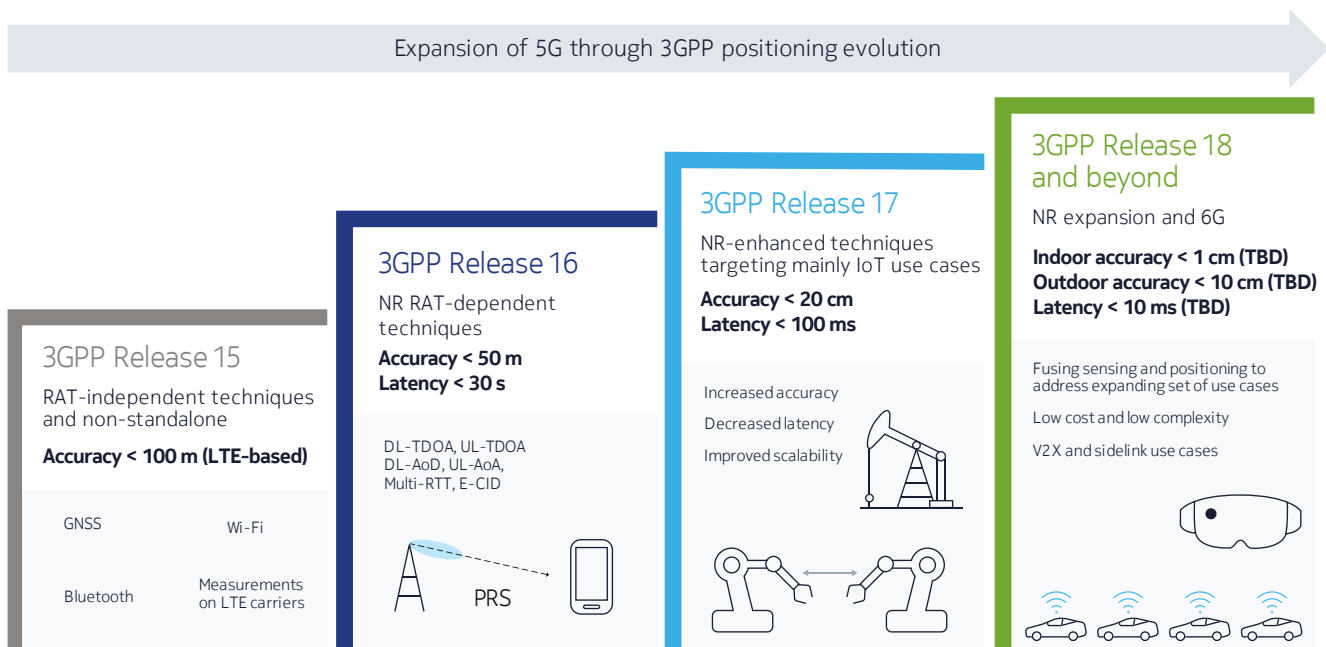
- Carrier-phase-based positioning, which promises higher accuracy
- Positioning using unlicensed spectrum (e.g., NR-U-based positioning)
- Very-low-cost and low-complexity positioning to enable asset tracking
- Sidelink and V2X positioning to address use cases like autonomous driving
- Positioning and pose estimation for extended reality applications.

Even beyond the above enhancements, we envision that 6G networks will have sensing capabilities that will enable the coexistence of communications, positioning and sensing over one network and unlock many more use cases. Moreover, 6G is anticipated to incorporate machine learning solutions in positioning, such that the location coordinates, or relevant parameters, are estimated as an outcome of a process where traditional methods are supplemented with artificial intelligence.

Summary

Positioning of devices has a wide diversity of use cases in current and future cellular network deployments. 5G NR has already addressed many of these use cases in Rel-15 and Rel-16, while further enhancements are still underway in Rel-17. Use cases range from providing emergency services with device locations to moving robots in IIoT factories of the future. As 5G NR continues to evolve and change the world around us, it is clear that positioning services are a necessary and seamless component of cellular networks that will only continue to expand the ways in which 5G and eventually 6G can be deployed.

Figure 3. Positioning features for 3GPP Rel-15 to Rel-18 and beyond



5G time service

Contents	
Introduction	3
5G time service use cases	4
Smart grid	4
Banking and points of sale	5
Media production and events	6
Sensors and testing	6
Consumer	7
Health and medicine	7
Achieving timing services over 5G	8
Over-the-air time synchronization to UTC	8
Time synchronization for 5G industrial IoT	9
5G time synchronization for individual devices	10
Comparison to GNSS/GPS	12
Summary and conclusions	12
References	13

Introduction

Since its introduction in 1933, humans have been able to adjust their clocks using telephone time-of-day or “talking clock” services, where a live or recorded human voice gives the correct local time. Today, time-of-day services are often accessed by computers over the internet to Internet Time Servers running Network Time Protocol (NTP). For cases where the internet is not available, GPS receivers have often been used, but their operation is challenging indoors. Other methods to obtain time-of-day services are considered in Figure 1.

In Release-16 of the 5G standard, 3GPP introduces time synchronization support for 5G over the air, which provides very high accuracy in the microsecond range and with deep indoor penetration. The initial focus for the 5G design has been to meet the strict demands of Industrial IoT (IIoT) and time-sensitive Ethernet networks, but with the next envisioned steps, 5G time-of-day services with microsecond accuracy can empower low-cost and low-power timing devices and clocks. In this white paper, we explain the envisioned functionalities and use cases behind the modern 5G version of “talking clock”.

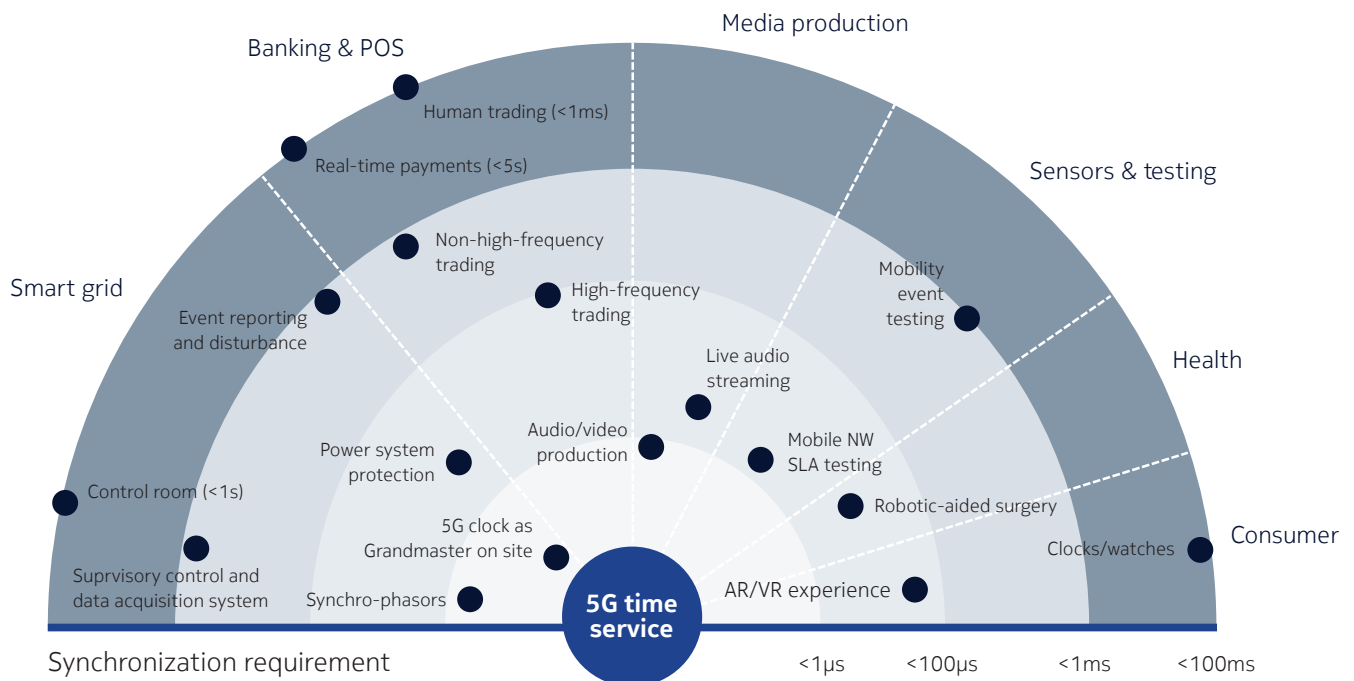
Figure 1 Illustration of the evolution of different methods available to acquire time-of-day synchronization and typical end-user accuracies achieved (inspired by [1])

Time of Day service	Telephone time of day including automated computer time		Radio station time service	GPS time transfer	Internet time service	5G time service
Access method by user	Landline	VoIP, cell phone	Wireless receiver	Wireless receiver	Internet access, ex. PC or time device	Wireless receiver
Typical synchronization accuracy	<30 ms	<150 ms	<50 ms	<100 ns	<100 ms	< 1µs
Years of public operation	1933 - today		1945 - today	1973 - today	1990 - today	2021 onwards

5G time service use cases

In an earlier white paper, time synchronization in combination with advanced Ethernet applications such as IEEE Time Sensitive Networking was described in detail [2]. In this section we focus on different use cases suitable for time synchronization over wide area 5G networks. The example use cases considered in this section have been compared in the radar chart in Figure 2, which shows required synchronization accuracy for each use case. In the following sub-sections, these use cases are discussed in more detail.

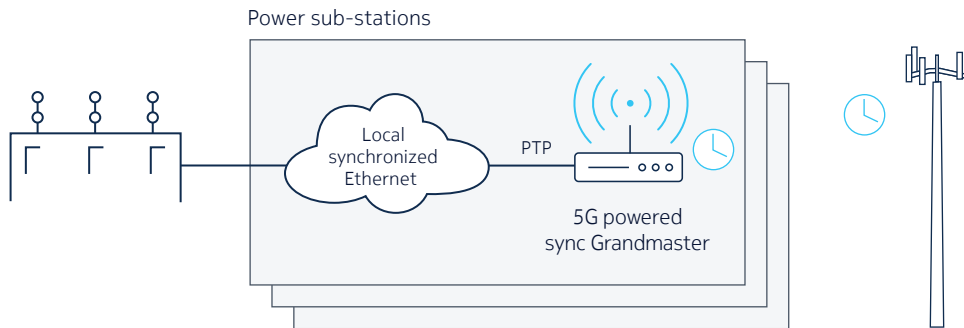
Figure 2. Illustration of various use cases that can be served by 5G time services in terms of synchronization requirements



Smart grid

The smart grid industry relies on timestamping of events for efficient management and operations as well as adjusting power sub-systems. Global Navigation Satellite System (GNSS) is commonly used for time synchronization, for example, in power sub-stations, which often requires the installation of an external antenna for good satellite coverage indoors. In many countries, however, there is already good coverage of cellular networks inside power systems and time synchronization via 5G will offer a convenient alternative or backup to GNSS for better reliability. An example use case is shown in Figure 3 where 5G directly provides time to an indoor located timing device. Depending on the use case, synchronization accuracies in the range of 1 microsecond to a millisecond can be served by 5G, including adaptation to the Precision Time Protocol (PTP) profiles used in the smart grid industry.

Figure 3. Illustration of the use of 5G time synchronization in power sub-stations



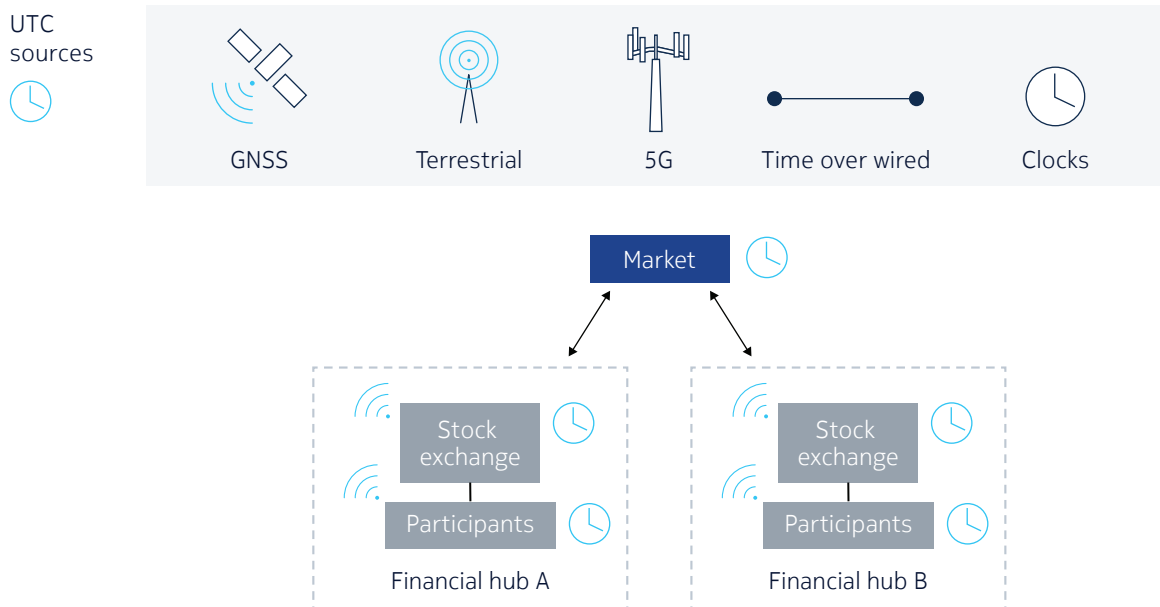
Banking and points of sale

Business activities and financial transactions, such as purchases between a buyer and a seller, rely on timing. Timing is necessary to record activities, execute rules and identify conflict, economic crime, etc. For instance, the points of sale (POS) for a retail or service business can be widely distributed. There may be thousands of independent machines running concurrently that need to be synchronized to a common time source on a real-time basis to ensure the data of all transactions are synchronized together (e.g., between online shopping and physical stores). Coordinated Universal Time (UTC) is the reference for all POS and is used to timestamp any events that take place.

Timestamping accuracy requirements for POS is easily met with 5G. Furthermore, time synchronization is also a key aspect in proof-of-stake blockchain-based approaches that seek to significantly improve the energy and transaction efficiency compared to approaches such as Bitcoin. Here, accurate time synchronization based on GNSS or NTP is associated with some risks such as poor indoor coverage, poor configuration and spoofed sources. These risks can be mitigated using 5G-based methods in parallel or as alternative.

For financial services and high-frequency trading (HFT) where prices change rapidly, the time window for trades is microseconds. Figure 4 shows an illustration of time synchronization in a banking use case with 5G providing a traceable UTC time to the financial hubs. Depending on the trading activity, financial regulations have imposed different requirements in terms of maximum divergence from UTC (ranging from 100 microseconds to 1 second) or granularity of the timestamp (ranging from 1 microsecond to 1 second) [3]. 5G can immediately meet regulation requirements for distributed financial systems as an alternative source to provide traceable UTC.

Figure 4. Illustration of time synchronization in trading and 5G providing traceable UTC



Media production and events

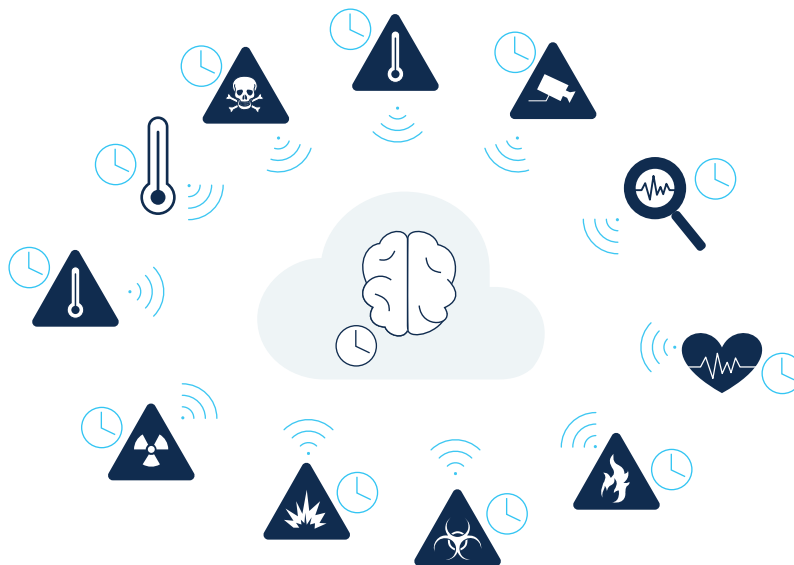
When audio and video are multiplexed for live events, time synchronization is essential to perfectly align different media sources. Media production or streaming devices (microphones, earbuds, instruments, video recorders including drones) rely more and more on wireless communications to generate new experiences and make production more efficient. 5G plays a key role enabling these new use cases [4]. Mixed reality live shows based on 5G are already happening, for example, fashion runway shows, art shows and concerts. In such shows, fast connectivity and low latency are essential. As well, different components, such as headsets, event screens, or live effects, must be accurately synchronized.

Accurate time synchronization via 5G opens new flexibility and reduces the need for local dedicated distribution networks and timeservers to achieve media synchronization. One example is multi-cam editing where external timecoding from a master timecode generator is used to ensure alignment of recordings and an improved workflow. 5G provides excellent indoor timecoding support, and devices can be built using standard output and signaling formats for use in professional or prosumer camcorders (e.g., BNC interface and SMPTE 2xx).

Sensors and testing

Timestamping capabilities in general are essential when troubleshooting complex events where exact timing and order of events is a key aspect. For instance, timing support in distributed sensors and test devices is illustrated in Figure 5. This is especially useful when cloud-based AI/ML frameworks are used to find correlations among a large variety of events observed in different places. Accurate timestamping can be used to automate storage into databases across a large set of inputs for generating sensor fusion as well as advanced correlations, which can lead to new insights in a large range of fields, including medicine, law enforcement, agriculture and autonomous driving.

Figure 5. How events with accurate and common timestamps help AI/ML based learning



Another use case includes drive testing in cellular networks where customized 5G devices are used to test the performance of deployments, including handover failures, data rates and coverage. Despite having GPS receivers and access to the internet, today's commercial test devices can timestamp events with an accuracy in the order of 100 ms only. This means that events cannot be accurately correlated with other events within the operator's network, or across devices when multiple devices are used for testing in parallel. Having access to microsecond level timing within the 5G device enables very accurate timestamping of radio events, thus enabling much deeper testing of network performance as well as advanced device capabilities (e.g., hybrid access devices). Accurate timestamping of both the device and the network also allows for one-way delay analysis compared to today's round-trip-time-based (RTT) methods. For instance, it can be used to test the network's capability to satisfy latency-based service level agreements (SLAs) in each direction and to make advanced troubleshooting of deployment issues.

Consumer

For consumer devices not connected to the Internet, GNSS or radio-station time service is sometimes an in-built option that is used to provide the time. This can include watches, metering and augmented and virtual reality devices (AR/VR). In the latter case, time synchronization becomes an essential service for multi-party AR/VR experiences. For many of these cases, a low-cost 5G based solution can offer benefits for the device and improve the quality of experience (QoE).

Health and medicine

In the health segment, timestamping can be useful for coordinated medicinal treatments. Synchronization is a prerequisite for advanced use cases such as remote robotic surgery.

Achieving timing services over 5G

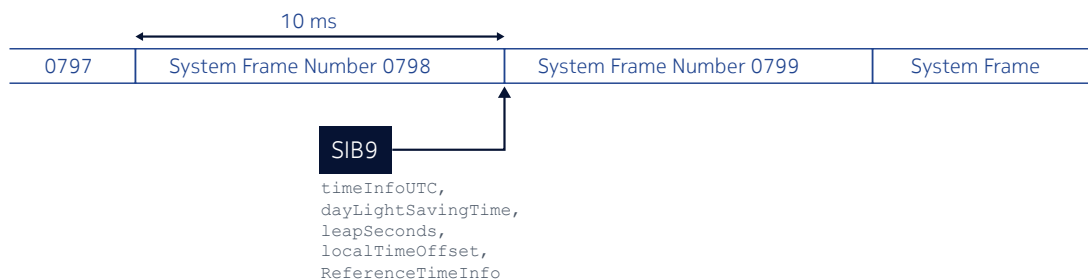
For cellular connectivity, the user equipment (UE) needs to know on which frequency and at what time it can exchange traffic with the base station, thus the UE and base station are synchronized to each other (although the UE does not know the absolute time of the synchronization). With Releases 16 and 17 of 5G, which create support for IIoT and time sensitive communications (TSC) beyond the scope of IIoT, the need for accurate 5G absolute time synchronization (i.e., synchronization to a physical time like UTC) has become much more critical. In the following sub-sections, we discuss the different methods to enable this accurate time synchronization in more detail.

Over-the-air time synchronization to UTC

The process of synchronizing a 5G device to the absolute 5G network time, typically UTC, is done by direct messaging over the air interface. 5G uses a simple but accurate method to communicate time of day to the device. The device (UE) identifies so-called system frames by monitoring the air interface, each with a system frame number (SFN). Contained within the 5G system information block 9 (SIB9) is the information related to GPS time and UTC. SIB9 messages are broadcast to inform devices of the exact time at the boundary between two SFNs. Hence, when the device has locked its internal clock to the system frame pacing, it can use this message to set its absolute time.

SIB9 contains the time information as well as flags to assist the device to understand the local time as shown in Figure 6. Reading the SIB9 and acquiring the time-of-day alignment can be performed by any device even if not subscribed to the 5G service. It only requires a 5G-capable receiver in the device as SIB9 can be broadcast in the cell. For fully subscribed devices, the network may also provide the timing information to each device by means of dedicated signaling (using Radio Resource Control (RRC) messages). More details are available in [5].

Figure 6. Communication of time-of-day over air interface in SIB9 message

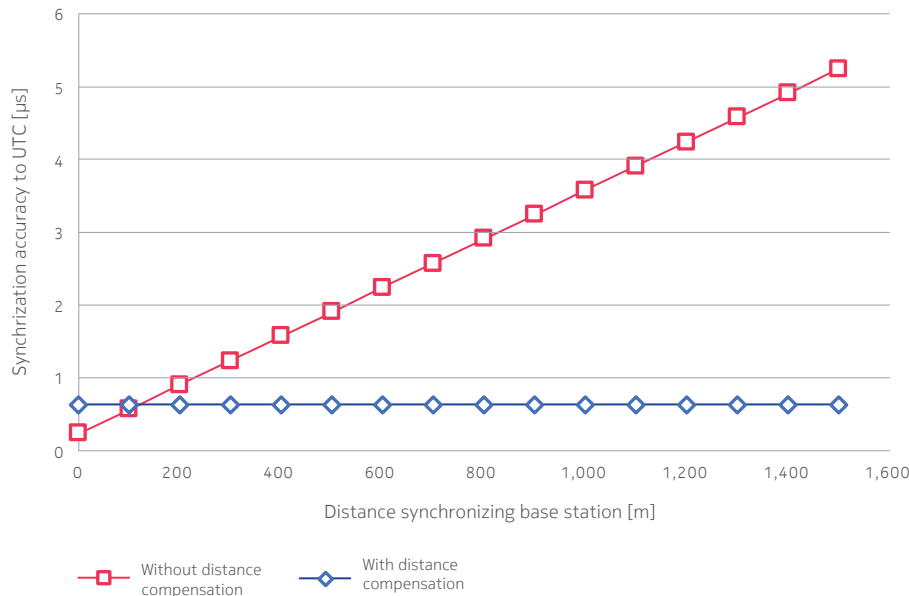


The distance the radio messages travels between the base station and the device causes a time difference between the time when the system frame was sent by the base station and the time when it was received at the device. This time difference is referred to as the radio signal propagation delay. As the propagation delay between the device and the base station delivering the timing information increases, the achieved synchronization accuracy degrades, unless the delivered time information is compensated. Even without compensation, however, good synchronization accuracy can be achieved for very large cells in the kilometer range as shown in Figure 7.

For a connected device, which requires subscription to the 5G operator, the propagation delay can be estimated based on the timing advance procedure, and compensation can be applied to mitigate the error caused by distance. While any connected UE can do compensation on their own, standardized and controlled compensation methods are introduced in 5G Release-17. The performance difference with

and without compensation is shown in Figure 7. For small distances less than 100 meters, compensation should typically be avoided, as propagation delay estimation has a certain error bound.

Figure 7. Achievable synchronization accuracy versus distance to base station providing synchronization



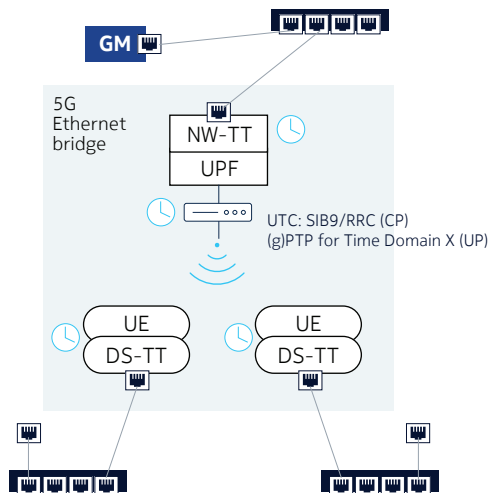
Time synchronization for 5G industrial IoT

In 5G Release-16, full downstream and tight synchronization support was introduced for 5G to serve advanced and time-critical Ethernet services, including IEEE Time Sensitive Networking (TSN) [6]. When used in IIoT, the 5G system can act as an IEEE Std 802.1AS [7] compliant Ethernet bridge, as illustrated in Figure 8. Each device represents one port of such a bridge, and there is one or more ports also in the 5G core network side. The input/output points of the 5G network are the device-side TSN translator (DS-TT) inside the device and the network-side TSN translator (NW-TT) on the 5G network side. The translator functionality implements IEEE Std 802.1AS including compensation for delay and jitter that takes place as individual time synchronization messages traverse the 5G network as user plane traffic. In this way a device can be synchronized to a vertical grand master clock (GM) connected to the 5G network end-point with an accuracy better than 900 ns (without considering additional propagation delay error).

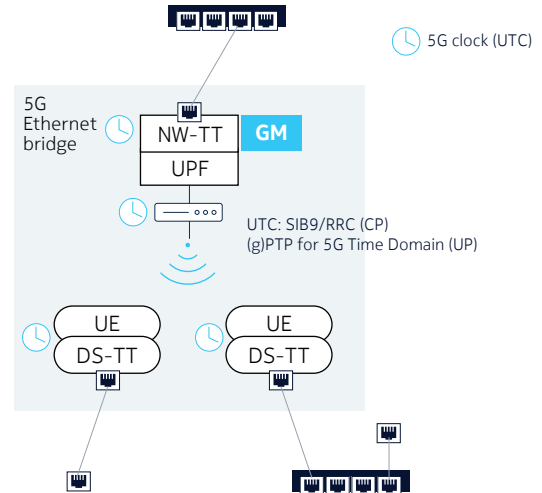
In 5G Release-17, uplink time synchronization support is added so that the GM clock can be located on the device side of the network to synchronize devices that are attached to other 5G devices or sitting behind the 5G network. Additionally, support for best-master clock algorithms (BMCA) is added to support high clock resiliency use cases. Besides IEEE Std 802.1AS support (gPTP), also IEEE Std 1588 PTP [8] support is added in Release-17 to facilitate a broader range of use cases and IP applications.

Figure 8. Illustration of synchronization mechanisms when 5G acts as an Ethernet bridge envisioned for IIoT uses. Devices act as ports of the bridge. DS-TT at the device side and NW-TT at the network side are synchronized to the 5G time domain (e.g., UTC)

When 5G is an Ethernet bridge, with transparent SYNC protocol support (gPTP, PTP over IP/Eth)



When 5G is an Ethernet bridge, with self-contained SYNC protocol support (gPTP, PTP over IP/Eth)



5G time synchronization for individual devices

In IIoT, devices (comprised of UE and DS-TT) are considered as ports on a bridge in an Ethernet network. Many use cases, however, require time synchronization support for a single device camping in the network with basic access to a public 5G network and not part of an Ethernet bridge construct. In 5G Release-17 and Release-18, several such use cases are envisioned. Figure 9 shows five different methods to access time of day from the 5G network and ways to use that information by an application attached to the device (e.g., connected via Ethernet port, a dedicated sync hardware port or via drivers for the operating system). The actual delivery of time information over the air interface using SIB9 and/or RRC messaging is described in the previous section.

The many different methods offer flexible support for a wide range of use cases as well as device designs. The device range can vary from advanced devices, such as DS-TT and advanced synchronization protocols, to very simple devices only capable of reading the time of the 5G system broadcast channels (i.e., SIB9), not requiring a transmitter module, a 5G subscription or even a SIM card slot. Taking the device type as a starting point, Table 1 presents some of the options and the typical synchronization accuracy.

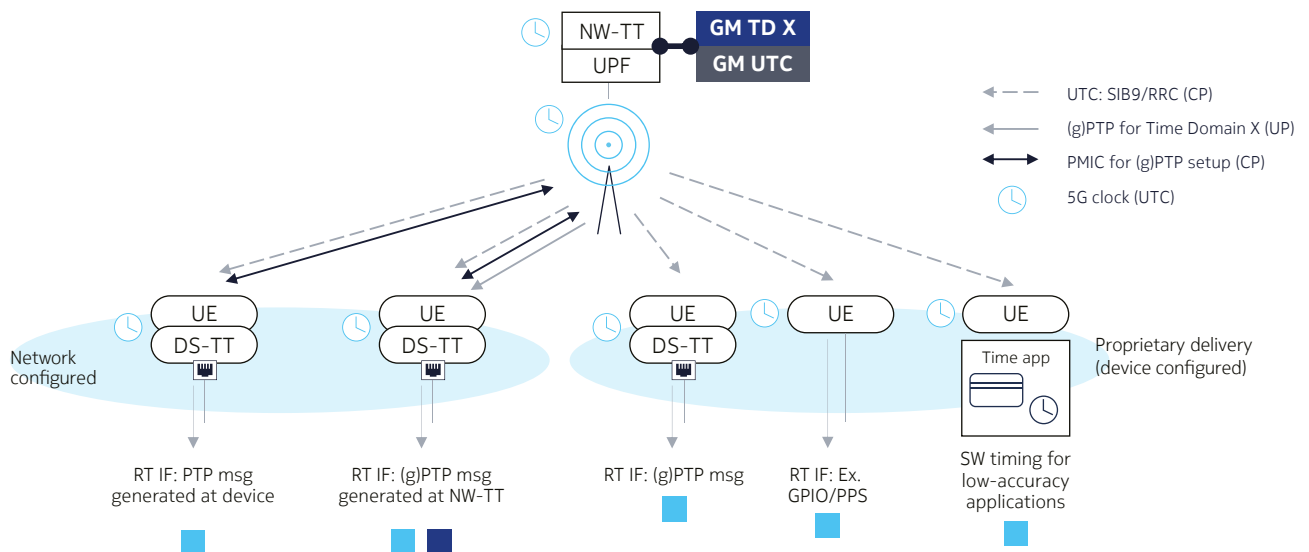
Table 1. Time-of-day synchronization solutions for standalone devices

Device type	5G subs. needed	Sync domain	Method of syncing device	Method for syncing applications	Accuracy level
Full UE/DS-TT	Yes	5G (UTC) only	5G air interface (Control plane (CP) or User plane (UP))	Ethernet, device PTP GM clock	<1µs
Full UE/DS-TT	Yes	Any, incl. 5G (UTC)	5G air interface and received PTP message (CP/UP)	PTP boundary or transparent clock. Compensates and forwards PTP message	<1µs
UE with GPIO	Yes	5G (UTC) only	5G air interface (CP)	Dedicated IF e.g., GPIO/1PPS	<1µs
UE with SW application	Yes	5G (UTC) only	5G air interface (CP/UP)	Via operating system on device	<10µs
Low-cost, low-power device	No	5G (UTC) only	5G air interface (CP), uncompensated	Dedicated IF e.g., GPIO/1PPS	<10µs

Whenever the most accurate results are needed, the network needs to help measure the propagation delay, which requires an active connection (temporary or regular), thus a subscription to the carrier. If lower accuracy of 5-10 µs is acceptable, compensation is not strictly needed. Another factor is whether the device takes care of its own synchronization delivery to its attached clients or if this functionality should be configured by the network. For network configured scenarios, configuration of the (g)PTP operation at the device side is done by exchanging a port management information container (PMIC) between the DS-TT and the 5G system. This type of scenario can be helpful when a company manages a fleet of devices that require automated deployment and monitoring.

Figure 9. Illustration of how standalone devices can achieve time-of-day synchronization from the 5G network and provide accurate timing to their attached clients (RT IF indicates any real-time interface).

When single UE is downstream sync device using any SYNC protocol or delivery mechanism



Comparison to GNSS/GPS

As shown in the use cases, 5G is a viable alternative or supplement to GNSS/GPS for providing time synchronization to UTC or any well-defined time domain. The pros and cons of both solutions are illustrated in Table 2. It is worth noting that the telecommunications sector is one of the big consumers today of GNSS-provided timing and in that sense relies on GNSS for its operations. However, in Release-18, 5G will introduce a timing resiliency service that will make the 5G system robust in case of GNSS failure or degradation. This resiliency will benefit users who perform time-of-day synchronization over the 5G air interface.

Table 2. Comparison of 5G and GNSS for getting time-of-day

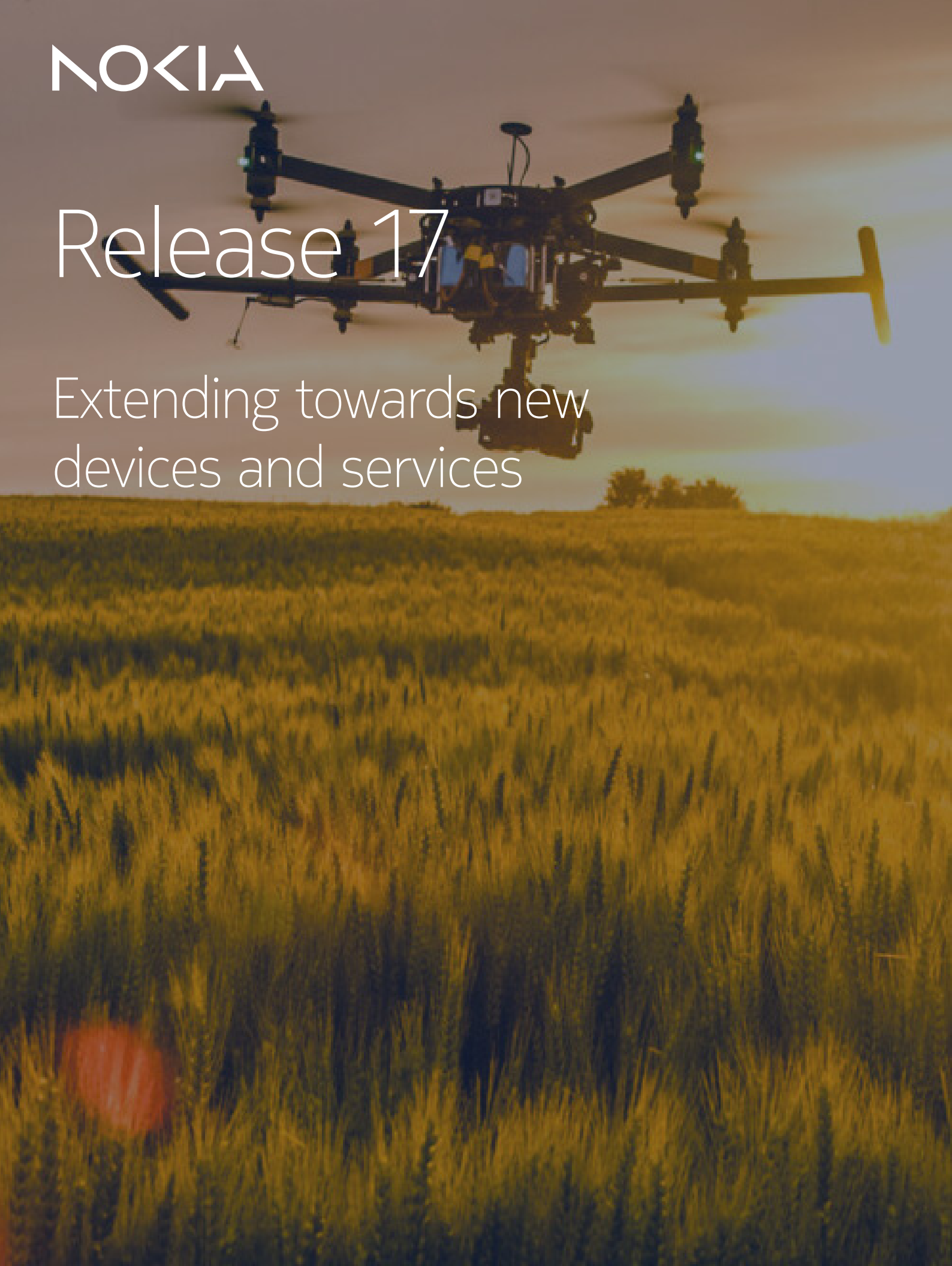
Comparison metric	5G	GNSS
Accuracy (UTC)	<900 ns with subscription, <5-10 μ s without subscription	<100ns
Indoor support	Yes	Limited, may require external antenna and wiring
Populated area coverage	Yes	Yes
Remote area coverage	Limited, later available via 5G non-terrestrial networks	Yes
Service subscription required	Only for most accurate use cases	No
Service resiliency	High with 5G timing resiliency system	Medium, many GNSS risk factors
Power consumption	Medium-high for most accurate timing Very low for <10 μ s accuracy use cases	Medium

Summary and conclusions

Precise time-of-day synchronization over 5G is one of the first examples of wireless networks becoming an essential part of critical infrastructure nationally and globally, serving societies and industries in areas beyond mobile broadband and IoT applications. This white paper has introduced the technological options including possibilities for low-cost device types that can be integrated into a large range of critical applications within industrial, public and consumer segments.

References

- [1] National Institute of Standards & Technology (NIST), “Time Realization and Distribution,” 2021. [Online]. Available: <https://www.nist.gov/time-distribution/radio-station-www/telephone-time-day-service>.
- [2] Nokia, “5G plug-and-produce,” 2021. [Online]. Available: <https://onestore.nokia.com/asset/207281/>.
- [3] European Commission, “Markets in Financial Instruments (MiFID II) - Directive 2014/65/EU, annex to Regulatory Technical Standard 25, level of accuracy of business clocks,” [Online]. Available: https://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160607-rts-25-annex_en.pdf.
- [4] Nokia & Sennheiser, “Low Latency 5G for Professional Audio Transmission,” January 2021. [Online]. Available: <https://www.bell-labs.com/institute/white-papers/low-latency-5g-professional-audio-transmission/>.
- [5] 3GPP, “TS 38.331 NR; Radio Resource Control (RRC); Protocol specification,” 2020. [Online]. Available: https://www.3gpp.org/ftp//Specs/archive/38_series/38.331/38331-g20.zip.
- [6] 3GPP, “TS 23.501 System architecture for the 5G System (5GS),” March 2021. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/23_series/23.501/23501-h00.zip.
- [7] IEEE Std 802.1AS-2020, “IEEE Standard for Local and metropolitan area networks--Timing and Synchronization for Time-Sensitive Applications”.
- [8] IEEE Std 1588, “IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems”.

A drone is shown in flight, hovering over a vast field of tall, golden-brown grass. The scene is captured during sunset or sunrise, with a warm, orange glow across the sky and the field. The drone is a quadcopter with a camera mounted underneath, and its propellers are blurred from motion. The overall mood is serene and technological.

NOKIA

Release 17

Extending towards new
devices and services

5G reduced capability devices

Contents

Introduction	3
5G IoT and reduced capability devices	5
Use cases and opportunities	7
Reduced capability devices	8
Complexity reduction	8
Power saving	11
Device definition and identification	11
Coverage analysis	12
Performance analysis	17
Reduced capability device evolution	19
Conclusions	20

Introduction

Internet of things (IoT) refers to the interconnection and autonomous exchange of data between devices. Three main IoT categories include massive IoT, critical IoT and broadband IoT. In addition, the industry has identified mid-range IoT use cases that are beyond the capabilities of massive IoT devices yet do not require broadband or critical types of services. Examples of these use cases include industrial wireless sensors, video monitoring and wearables such as smart watches and medical monitoring devices. These use cases will benefit from having a new device type that is lower in cost compared to high-end broadband and critical IoT devices. Furthermore, device designs supporting compact form factors are an important requirement for these use cases.

3GPP in Release 17 of the 5G standard (hereafter Rel-17) started standardization development work for reduced capability (RedCap) NR (5G new radio) devices. In this white paper, we provide an overview of the new device type and describe new complexity reduction and power saving features being introduced to support mid-range IoT use cases. Our analysis shows that significant complexity reduction can be achieved. When compared to reference Rel-17 NR devices, it is seen that a total complexity reduction of approximately 70% for FR1 (FDD and TDD) and 50% for FR2 (TDD only) can be achieved.¹

In addition, the white paper presents network coverage analysis to show the impact from the introduction of RedCap devices. Coverage analysis based on typical NR deployment scenarios and target cell edge data rates shows that coverage compensation is generally not needed despite the performance loss from complexity reduction features.

Similarly, network capacity analysis shows that the impact to system spectral efficiency and capacity is minor when the system is not heavily loaded. There is also only minor impact to the performance of enhanced mobile broadband (eMBB) users.

Rel-17 specification work for RedCap devices is targeted to be completed by September 2022. Commercial deployment may be expected 18–24 months after completion of the specifications.

The evolution of RedCap devices will continue with Rel-18, which is the first release of 5G-Advanced. Rel-18 content can be grouped into four areas as shown in Figure 1: Experience (providing new levels of experience), Extension (extending networks into new areas), Expansion (expanding mobile networks beyond connectivity) and Excellence (providing excellent operational support). Providing further complexity reduction in FR1 bands, the goal is to introduce lower-tier devices between massive IoT and Rel-17 RedCap devices, which will be an important part of the Extension block of features.

¹ FDD stands for “frequency division duplex” and TDD for “time division duplex”. FR1 and FR2 represent the two frequency ranges in 5G. FR1 is sub-6 GHz spectrum and FR2, mmWave. Bandwidths include 5–100 MHz (FR1) and 50/100/200/400 MHz (FR2).

Figure 1. 5G-Advanced evolution in four areas



5G IoT and reduced capability devices

Internet of Things (IoT) refers to the interconnection and the autonomous exchange of data between devices and can also be described as a network of physical objects that are connected and exchanging data between themselves via the internet. IoT use cases can be grouped into three categories.

Massive IoT: This refers to IoT use cases or applications that are delay-tolerant and require low data rates. Massive IoT requires low-cost devices, devices with low energy consumption for extended battery life, extended network coverage, and support of massive numbers of devices in the network. Example use cases include asset tracking, telematics, remote monitoring, smart grid and smart city applications. These applications are supported using LTE-M (long-term evolution for machines) and Narrowband IoT (NB-IoT) based on 4G cellular technology [1]. LTE-M is intended for IoT applications requiring higher data rates and can support voice and video services, while NB-IoT can provide very deep coverage and support low-cost devices.

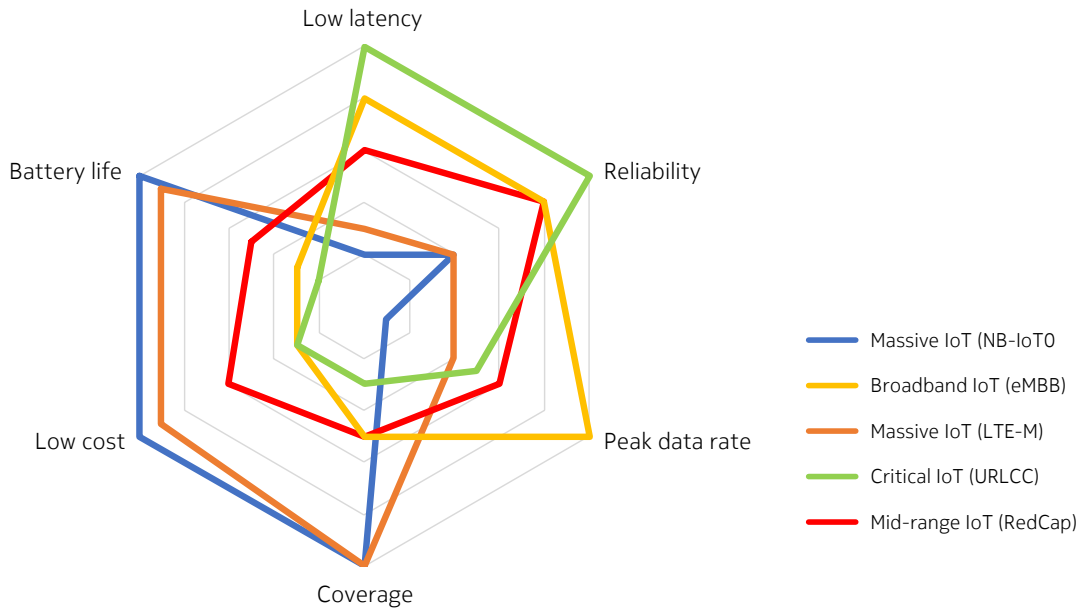
Critical IoT: This refers to IoT use cases or applications that require time-critical data delivery with strict delay constraints and high reliability. Example use cases include factory automation, industrial control, robotics, augmented reality and virtual reality. In 5G NR, these applications are supported using ultra-reliable low-latency communication (URLLC) features. These features have been designed to deliver very high reliability (99.999% uptime) with very low delay (~1 ms end-to-end). Low power consumption, however, is not critical for this IoT category.

Broadband IoT: This refers to IoT use cases or applications that require high data rates and lower delay that are typical of mobile broadband services. Example use cases include industrial gateways, hot spots, and wearables. In general, broadband IoT devices have capabilities similar to consumer mobile broadband devices. For instance, 5G industrial gateways can deliver peak data rates of 4.2 Gbps in the downlink (DL) and 0.45 Gbps in the uplink (UL).

In addition to the above categories, the industry has identified mid-range IoT use cases that are beyond the capabilities of massive IoT devices yet do not require broadband or critical services. Examples of these use cases include industrial wireless sensors, video monitoring and wearables such as smart watches and medical monitoring devices. These use cases will benefit from having a new device type that is lower in cost compared to high-end broadband and critical IoT devices. Furthermore, device designs supporting compact form factors are an important requirement for these use cases. The device requirements for the various IoT categories are illustrated in Figure 2.

Furthermore, ambient power-enabled IoT (also referred to as passive IoT) is being considered in 3GPP. Passive IoT refers to devices without battery or with limited energy storage that are powered by energy harvesting (e.g., from radio waves, light and motion). Examples include radio frequency identification (RFID) tags and near-field communication (NFC) smart cards. Passive IoT is not meant as a replacement to LTE-M and NB-IoT, but to address massive IoT market segments requiring ultra-low power consumption and ultra-low device cost. In the Service and System Aspects group, a study on ambient power-enabled IoT has been approved. The study aims to first study use cases and potential service requirements for passive IoT and then to perform gap analysis between passive IoT and existing requirements. Further studies will be undertaken by the RAN plenary in Rel-19 to identify deployment scenarios and design targets such as coverage, data rate, power consumption, cost and positioning accuracy.

Figure 2. IoT device requirements



In light of the above, 3GPP in Rel-17 started standardization development work for RedCap NR devices. The goal is to deliver lower-cost IoT devices that satisfy the design requirements for mid-range IoT use cases as provided in Table 1. This will allow important categories of use cases such as wearables, industrial sensors and video transmission to be served by lower-cost NR devices thus enabling cost-effective IoT solutions to be deployed in various vertical industries. The design must also support compact form factors. Core specifications to support RedCap devices were completed in March 2022, while performance-related specifications are expected to be completed by September 2022. Commercial deployment may be expected 18–24 months after completion of the specifications.

Table 1. Requirements for NR IoT use cases

Specifications	Critical IoT	Massive IoT	Mid-range IoT
Latency	1 ms	10 seconds	5–10 ms for safety reports, 100 ms for others
Reliability	99.999%	99–99.9%	Up to 99.99%
Peak data rates	N/A	LTE-M: 2.4 Mbps DL, 2.6 Mbps UL NB-IoT: 127 Kbps DL, 159 Kbps UL	Up to 150 Mbps DL, 50 Mbps UL
Battery life	N/A	10 years	1–2 weeks for wearables, several years for industrial sensors
Coverage	N/A	164 dB (maximum coupling loss)	Same as 5G eMBB
Frequency range support	FR1, FR2	FR1	FR1, FR2
Core network support	5GC	EPC and 5GC (for devices supporting N1 mode)	5GC – Standalone
Use case examples	Factory automation, industrial control, robotics, remote surgery	Asset tracking, telematics, remote monitoring, smart grid, smart city, security, healthcare	Wearables, industrial sensors, video transmission

Use cases and opportunities

RedCap devices are designed for use cases that are beyond the capabilities of massive IoT technology such as LTE-M and NB-IoT. This means that they can support high data rates and low latency as shown in Table 1. Some representative use cases include:

- Video transmission for traffic monitoring and security cameras
- Wearables such as smart watches, goggles, smart glasses, on-body health sensors and medical devices
- Industrial wireless sensors for temperature, pressure, proximity, smoke, level and vibration
- Connected car applications like infotainment, telematics, real-time high-definition maps and software upgrades
- Drones for command and control and video transmission

Some of these use cases are illustrated in Figure 3 in the context of smart city and smart factory.

Figure 3. IoT use cases for RedCap devices



Currently, broadband IoT use cases are typically supported using LTE due to wide availability of LTE networks and devices. 5G connectivity for IoT remains nascent due to high device cost and network focus on the broadband segment. The introduction of RedCap devices will help expand the IoT opportunities for the operators, including helping to:

- Expand the addressable IoT market for 5G NR using the existing network footprint by making lower-cost devices available and enabling network software updates
- Support the re-farming of LTE spectrum to NR by providing cost-effective device solutions for IoT service migration
- Provide NR-based IoT solutions to complement time-critical URLLC components in Industry 4.0 for better integration and improved network efficiency.

Reduced capability devices

In this section, we provide an overview of the features introduced to support RedCap devices. They include device complexity reduction techniques, how to support them in the network, and power saving features.

In addition, RedCap devices can operate only in standalone mode and with single connectivity (i.e., no dual connectivity support), and they can only operate in a single band at a time (i.e., no carrier aggregation support). Also note that Rel-15 SSB bandwidth is reused, and physical layer changes are kept to a minimum.

Complexity reduction

In 3GPP, a study was conducted to evaluate different complexity reduction techniques with respect to amount, performance impact, specification impact, and co-existence with legacy devices [2]. The study defined reference NR devices for complexity comparison purpose and included both NR frequency ranges (FR1 and FR2). The evaluation methodology for complexity analysis considers radio frequency (RF) and baseband parts, and the reduction is expressed as a relative cost compared to the reference NR device. The baseband part includes analog-to-digital and digital-to-analog converters (ADC/DAC), fast Fourier transforms and inverse fast Fourier transforms (FFT/IFFT), data buffer, transmitter/receiver processing block, encoder/decoder, etc. Complexity is evaluated based on required baseband operations. The RF part includes power amplifier, filters, transceiver and duplexer. The cost reduction can be achieved by either removing the component or replacing it with a less expensive component.

Based on the outcome of the study, the following complexity reduction techniques were agreed to be introduced.

Reducing the user equipment (UE) RF bandwidth: Reducing the RF bandwidth (BW) can reduce complexity significantly since this option reduces both the ADC/DAC and FFT/IFFT complexity as well as the peak data rates that can be supported by the device. The reduction amount scales with the BW, therefore it is beneficial to select the smallest BW possible. To allow the RedCap UE to continue to operate with legacy channels and signals, the bandwidth can only be reduced to 20 MHz for cmWave (FR1) and 100 MHz for mmWave (FR2). For FR1, reducing the bandwidth from 100 MHz to 20 MHz reduces device complexity by approximately 30%. For FR2, reducing the bandwidth from 200 MHz to 100 MHz reduces device complexity by approximately 15%.

Reducing the number of receive (Rx) branches: This can result in a large complexity reduction as each RF chain constitutes a large percentage of the RF cost. Having only a single receiver chain can reduce the UE complexity by as much as 15–30%, compared to having two, and by as much as 46%, compared to having four. It results, however, in performance degradation for the device in the downlink. In addition to reducing the RF cost, having only a single receiver chain eliminates the need to support spatial multiplexing MIMO (multiple input, multiple output). This can further simplify the baseband operations at the cost of reduced throughput in the downlink.

Half-duplex FDD operation: Half-duplex operation allows for the removal of the duplexer in the RF. The complexity reduction is minor, estimated to be up to 7% per frequency band. However, the saving can be substantial as the device will likely support many bands and so the duplexer can be removed for each supported band.

Relaxing the maximum modulation order in FR1: In NR FR1, legacy devices support 256-QAM (quadrature amplitude modulation) in the downlink. For IoT devices, such very high modulation is not usually necessary and baseband processing savings can be achieved by supporting up to 64-QAM only. This results in a small complexity reduction of approximately 6%. While the complexity reduction is minor, the impact of this relaxation is negligible, therefore, it was agreed to be supported. Note that, in FR2, the maximum modulation order is 64-QAM, thus there is no relaxation in FR2.

Table 2 illustrates the relative complexity reduction in comparison to reference NR devices. From the table, it is seen that total complexity reduction of approximately 70% and 50% can be achieved for FR1 and FR2 devices, respectively.

Table 2. Complexity reduction analysis

Complexity reduction technique	Relative complexity reduction		
	Frequency range 1 (cmWave) FDD	Frequency range 1 (cmWave) TDD	Frequency range 2 (mmWave) TDD
Reducing the UE RF bandwidth	31.9%	33.4%	15.6%
Reducing the number of Rx branches	25.5%	46.0%	30.6%
Half-duplex FDD operation	6.8%	N/A	N/A
Relaxing the maximum modulation order	5.8%	6.3%	N/A
Total device complexity reduction	67.1%	70.7%	47.5%

Table 3 summarizes the key characteristics of RedCap devices including optional capabilities. In FR1, the full duplex FDD (FD-FDD) RedCap device is capable of peak data rates of 85 Mbps in the downlink and 91 Mbps in the uplink. Half-duplex (HD) and TDD devices will have their peak rates reduced accordingly. In FR2, the TDD RedCap device in the 50:50 downlink-to-uplink time split is estimated to be capable of peak data rates of 213 Mbps in the downlink and 228 Mbps in the uplink.

Table 3. Reduced capability NR devices

	Frequency range 1 (cmWave)	Frequency range 2 (mmWave)
Device bandwidth	20 MHz	100 MHz
Antenna configuration	1Tx-1Rx 1Tx-2Rx (optional)	1Tx-1Rx 1Tx-2Rx (optional)
Downlink MIMO support	Yes for device with 2Rx branches	Yes for device with 2Rx branches
Duplex operation	FD-FDD, HD-FDD, TDD	TDD
Maximum modulation	DL: 256-QAM (optional), 64-QAM mandatory UL: 64-QAM	DL: 64-QAM UL: 64-QAM
Peak data rate	FD-FDD, 1Rx: 85 Mbps DL/91 Mbps UL TDD 50:50 DL/UL split, 1Rx: 42 Mbps DL, 45 Mbps UL	TDD 50:50 DL/UL split, 1Rx: 213 Mbps DL, 228 Mbps UL

From the 3GPP point of view, RedCap devices can be supported with moderate changes to the specifications. Key modifications are summarized below.

Reducing the UE RF bandwidth

Due to the reduced device RF bandwidth, RedCap can operate only in a Bandwidth Part (BWP) that is no larger than the device bandwidth. As a result, separate BWP may need to be configured for the RedCap device. In NR, a BWP is a contiguous set of frequency resources that can be configured for a UE. For instance, in FR1, if the carrier BW is 40 MHz, a separate BWP no larger than 20 MHz must be configured for the RedCap device. Legacy UE, however, can continue to use the entire BW and can be configured with

another BWP spanning the entire 40 MHz. The Rel-15 specification already supports configurations of BWP, however, some enhancements were introduced for RedCap devices, as explained below.

In the downlink, a separate BWP may be configured that may not contain the entire cell-defining synchronization signal block (CD-SSB) and control resource set #0 (CORESET#0). This BWP can be used for initial access and when the device is in a connected state. If this BWP is used by the UE in a radio resource control (RRC) connected state, it should contain a non-cell-defining SSB (NCD-SSB). The NCD-SSB may be transmitted less frequently than the CD-SSB, thus lowering the additional overhead. In addition, some devices may have the optional capability to operate without requiring an NCD-SSB. In this case, they can rely on the channel state information reference signal (CSI-RS) for relevant operations.

In the uplink (UL), a separate BWP can also be configured for use both during initial access and in the connected state. The network can disable intra-slot physical uplink control channel (PUCCH) frequency hopping within this BWP so that the physical uplink shared channel (PUSCH) resource is not fragmented.

Reducing the number of Rx branches

The specification mainly impacts the UE performance requirements by reducing the number of Rx branches. In case the RedCap device has two Rx branches, the existing radio access network (RAN4) requirements can be reused. For a device with one Rx branch, new requirements for both RF and radio resources management (RRM) will be defined. They include, for example:

- The reference sensitivity for devices with 1Rx is based on 2Rx reference sensitivity with relaxation (2.5 dB for TDD, FDD; 2.5 dB for 5MHz; 3 dB for 10/15/20 MHz)
- New handover requirements are needed for 1 Rx UE for the following cases: NR FR1–FR1, NR FR2–FR2, NR-E-UTRAN, E-UTRAN–NR, NR FR1–FR2, NR FR2–FR1
- The number of samples for synchronization signal detection for FR1 is extended
- The accuracy level is relaxed for SSB-based layer 3 (L3) measurement with 1 Rx.

Half-duplex FDD operation

Half-duplex (HD) devices cannot transmit and receive simultaneously in FDD. The specification impacts from HD-FDD operation include switching times for the device and device behavior in case of collision between transmission and reception. On the first point, Rel-15 switching times for UE not capable of full-duplex communication are reused for HD-FDD devices. In FR1, the switching times for transmit-receive (Tx-Rx) and Rx-Tx are 13 μ s. On the second point, collision handling behaviors have been defined for HD-FDD devices.

The defined behaviors are based on the existing Rel-15 behaviors for TDD UE and can be summarized, as follows. Where there is a collision between:

- Dynamically scheduled downlink reception and semi-statically configured uplink transmission, prioritize the dynamically scheduled downlink reception
- Semi-statically configured downlink reception and dynamically scheduled uplink transmission, prioritize the dynamically scheduled uplink transmission
- Semi-statically configured downlink reception and semi-statically configured uplink transmission, it is an error case and not expected by the device
- Dynamically scheduled downlink reception and dynamically scheduled uplink transmission, it is an error case and not expected by the device
- SSB and uplink transmission, prioritize SSB
- Random access and downlink reception, leave for UE implementation.

Relaxing the maximum modulation order in FR1

There is negligible specification impact from this relaxation. The existing specifications can be reused to support this feature.

Power saving

Two power saving features will be introduced in Rel-17: extended discontinuous reception (eDRX) and radio resource management (RRM) measurement relaxations for neighboring cells.

In NR, devices in idle or inactive mode must wake up periodically to check for paging and perform measurements. For many IoT use cases, devices may have infrequent data transmissions, thus there is no need to wake up frequently. The first power saving feature is an eDRX procedure that allows them to remain in a power-efficient deep sleep state longer for reduced power consumption. In Rel-17, eDRX is being specified for devices in radio resource control (RRC) inactive and idle states. Two maximum eDRX cycles will be supported: 10.24 s and 10485.76 s.

The eDRX feature is optional for any next-generation NodeB (gNB), which means the gNB implementation must explicitly support eDRX. The eDRX configurations can be different for RRC_IDLE and RRC_INACTIVE. Based on the objectives in the Rel-17 work item description, the paging time window (PTW) and paging hyper-frames (PH) will not be used in eDRX cycles up to 10.24 s. For eDRX cycles longer than 10.24 s, however, the eDRX feature, including the related parameters such as PTW, PH, and the corresponding paging operation defined for E-UTRAN are used as the baseline for both RRC_IDLE and RRC_INACTIVE in NR/5GC. The lower bound for eDRX configuration in RRC_IDLE and RRC_INACTIVE is 2.56 seconds. It is up to the RAN to configure the PTW length for RAN paging, and the RAN PTW length can be different from the core network (CN) PTW length. The maximum PTW length is 40.96 s when the IDLE eDRX cycle is longer than 10.24s. The minimum PTW length is 1.28 s and the step length or granularity of the PTW length is 1.28 s when the idle eDRX cycle is longer than 10.24 s.

The second power saving feature is RRM measurement relaxations for neighboring cells. Two important characteristics for many IoT use cases are that, first, the device is stationary and, second, the device is mostly in an idle or inactive state with infrequent data transmissions. In idle mode, power consumption is dominated by RRM measurements. Therefore, enhancements are being made to allow the UE to relax RRM measurements for neighboring cells. Specifically, new stationarity criteria and not-at-cell-edge criteria will be defined for RedCap devices to allow them to prolong the time between neighbor cell measurements. This RRM relaxation is under network control and can be configured on a device basis.

Device definition and identification

The RedCap device type is based only on the maximum device bandwidth (20 MHz in FR1 and 100 MHz in FR2) and therefore only one device type is defined. Other features, such as the number of Rx branches, can be acquired by the network as part of the UE capability signaling.

In NR, device capability can be acquired by the network in Msg5 of the random-access procedure. Hence, the network would be able to determine whether a device is a RedCap device and its associated capability (e.g., number of Rx branches) in Msg5. The network, however, can configure RedCap devices for early identification using Msg1 or Msg3. For Msg1 early identification, dedicated random access occasions or dedicated random access preambles can be reserved for RedCap devices. For Msg3 early identification, reserved MAC logical channel IDs can be used. Note that, with early identification, the network would only know that the RedCap device has limited bandwidth. It does not know of other limitations such as the number of Rx branches or whether the device is half duplex or full duplex. Additional limitations like these, can be discovered by the network using the RRC UE capability enquiry procedure.

In 3GPP, it was further discussed how to prevent RedCap devices from using radio capabilities not intended for them (e.g., as low-cost eMBB devices). It was agreed that this can be left up to the network. For example, the network can perform subscription validation during RRC connection setup to ensure that the requested services are specific to RedCap devices. Also, the network can use system information block type one (SIB1) to bar RedCap devices.

Coverage analysis

One of the studies on RedCap devices involved analyzing coverage of RedCap devices to determine whether coverage compensation would be needed. It may be noted that the complexity reduction features described in the previous section degrade downlink performance. Furthermore, in some of the use cases under consideration, such as wearables, the device may also have a small form factor. This may limit the antenna efficiency, which then causes an additional coverage loss.

The following complexity reduction features can result in performance loss for a RedCap device compared with a reference NR device.

Reduced number of Rx branches: a RedCap device with fewer receiver antenna branches than a reference NR device would suffer a degradation in downlink performance due to a reduction in combining gain and receive diversity gain.

Reduced UE BW: A RedCap device with smaller BW than a reference NR device may experience some loss in frequency diversity, resulting in performance loss.

Reduced antenna efficiency: Devices with a small form factor may have a small antenna with reduced efficiency, resulting in a loss compared with an NR device. For evaluation purpose, it was agreed that the loss can be up to 3 dB [2].

Three different coverage metrics were used in the link budget analysis for RedCap devices: maximum isotropic loss (MIL), maximum path loss (MPL), and maximum coupling loss (MCL). MIL incorporates antenna gains, thus it enables analysis of the impact of reduced antenna efficiency on coverage. Coverage of the following channels and messages was considered based on a target performance requirement.

- PDCCH – Physical downlink control channel
- PDSCH – Physical downlink shared channel
- PUCCH – Physical uplink control channel
- PUSCH – Physical uplink shared channel
- PRACH – Physical random access channel
- Msg2 – Message 2 of initial access carried on PDSCH
- Msg3 – Message 3 of initial access carried on PUSCH
- Msg4 – Message 4 of initial access carried on PDSCH

The main assumptions for the coverage analysis are listed in Table 4, breaking out the assumptions related to complexity for a reference NR UE (Reference) and a RedCap UE (RedCap). Scenarios in FR1 and FR2 were considered. The target performance metric specified for the uplink (UL) and downlink (DL) data channels was throughput. On the other hand, for the downlink control channel and a few other channels, a block error rate (BLER) was specified as the target metric. The link budget calculations used the signal-to-noise ratio (SNR) required to achieve the target performance metric.

Table 4. Scenarios and assumptions for coverage analysis

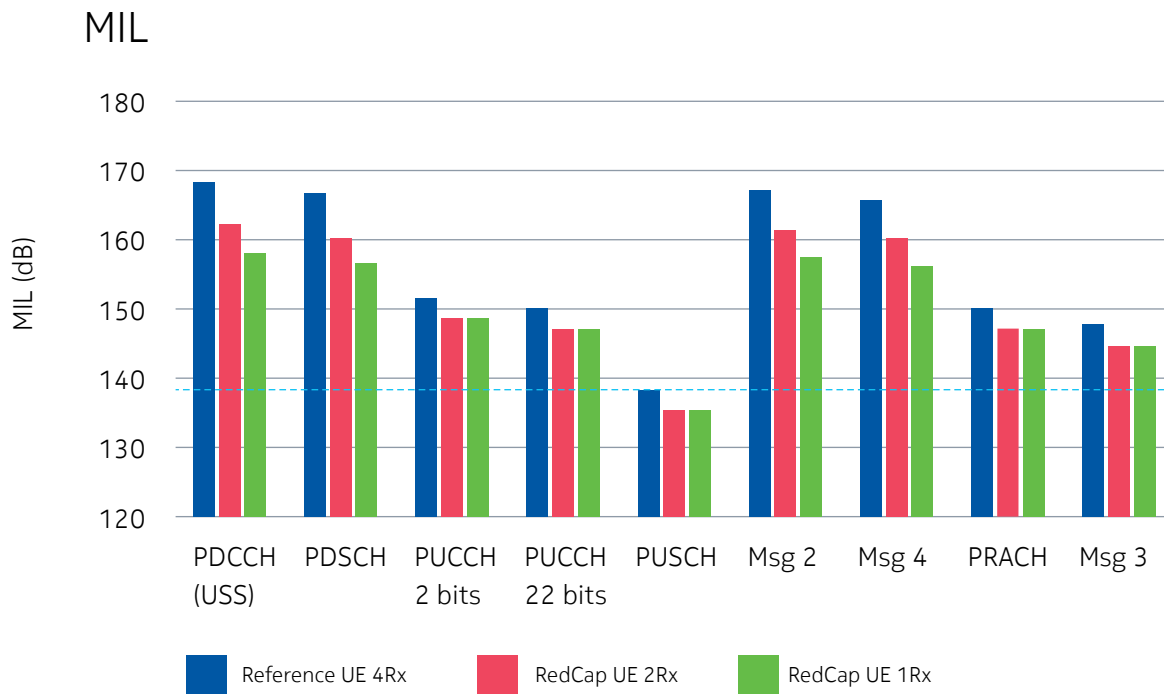
Scenario parameters		FR1	FR2	
Scenario		Urban	Rural	Indoor
Carrier frequency		2.6 GHz and 4 GHz	700 MHz	28 GHz
Duplexing		TDD	FDD	TDD
Subcarrier spacing		30 kHz	15 kHz	120 kHz
UE bandwidth	Reference	100 MHz	20 MHz	100 MHz
	RedCap	20 MHz	20 MHz	100 MHz
Number of UE antennas	Reference	4 DL, 1 UL	2 DL, 1 UL	2 DL, 1 UL
	RedCap	1 or 2 DL, 1 UL	1 or 2 DL, 1 UL	1 DL, 1 UL
UE antenna element gain	Reference	0 dBi	0 dBi	5 dBi
	RedCap	-3 dBi	-3 dBi	5 dBi
PDSCH target throughput	Reference	10 Mbps	1 Mbps	25 Mbps
	RedCap	2 Mbps	1 Mbps	25 Mbps
PUSCH target throughput	Reference	1 Mbps	100 kbps	5 Mbps
	RedCap	1 Mbps	100 kbps	5 Mbps

In the following, we consider the MIL for the following three scenarios.

1. Urban 2.6 GHz (TDD)
2. Rural 700 MHz (FDD)
3. Indoor 28 GHz (TDD)

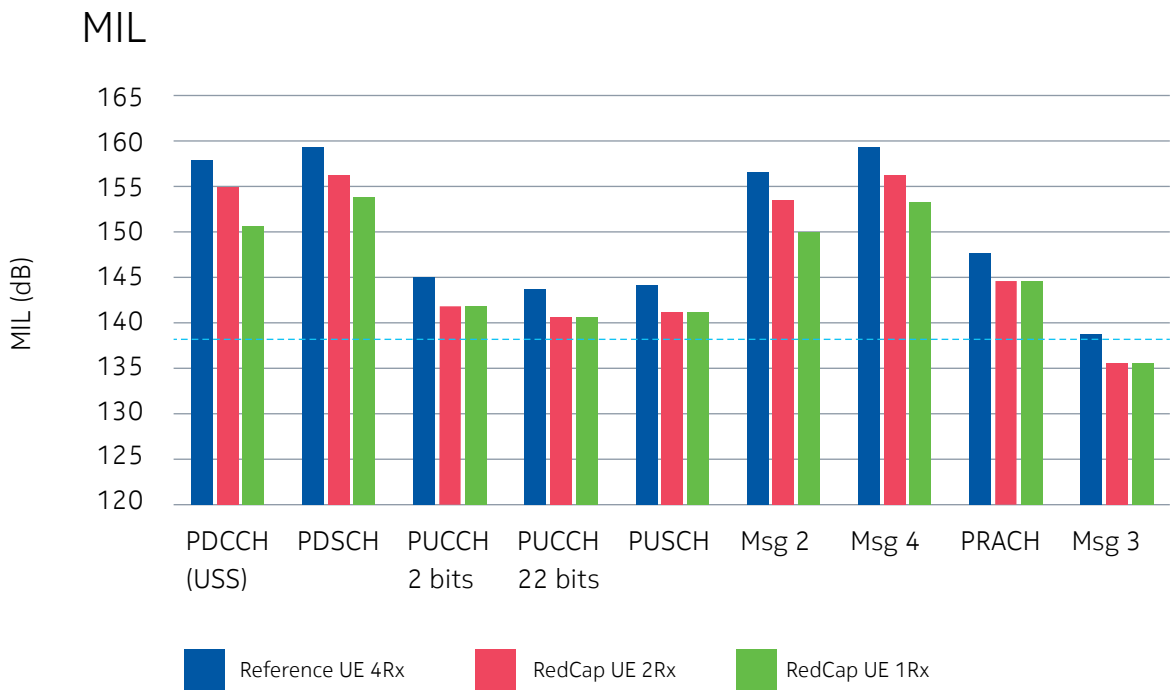
The MILs for all the considered channels in the Urban 2.6 GHz scenario are shown in Figure 4 for both a reference UE and a RedCap UE. RedCap UE with one Rx antenna and two Rx antennas are separately considered. The observed degradation of 3 dB in MIL for uplink channels of the RedCap UE is due to reduced antenna efficiency. On the other hand, reducing the number of Rx antennas causes additional degradation of the MIL on the downlink channels — the total MIL loss is about 10 dB for a RedCap UE with one Rx antenna. To determine the amount of coverage recovery that is needed for the RedCap UE, the limiting channel is identified. This is the channel for the reference UE with the smallest MIL value. As seen from the figure, the PUSCH is the limiting channel in this case, and the corresponding reference MIL is marked with a horizontal red line. Coverage recovery for a RedCap UE is required for a particular channel if the MIL is less than the reference MIL. Thus, in this case, a coverage recovery of 3 dB is seen to be necessary for PUSCH for RedCap UEs with either one Rx antenna or two Rx antennas. For all the other uplink channels, the MIL is better than the reference MIL by a significant margin and hence no coverage recovery is required. Likewise, no coverage recovery is required for downlink channels in this scenario, despite the performance loss due to reduced antenna efficiency and fewer Rx antennas. With more stringent downlink power assumptions, however, it was found that coverage recovery is required for some downlink channels.

Figure 4. Reference NR UE and RedCap UE link budget for different channels in the urban 2.6 GHz (TDD) scenario



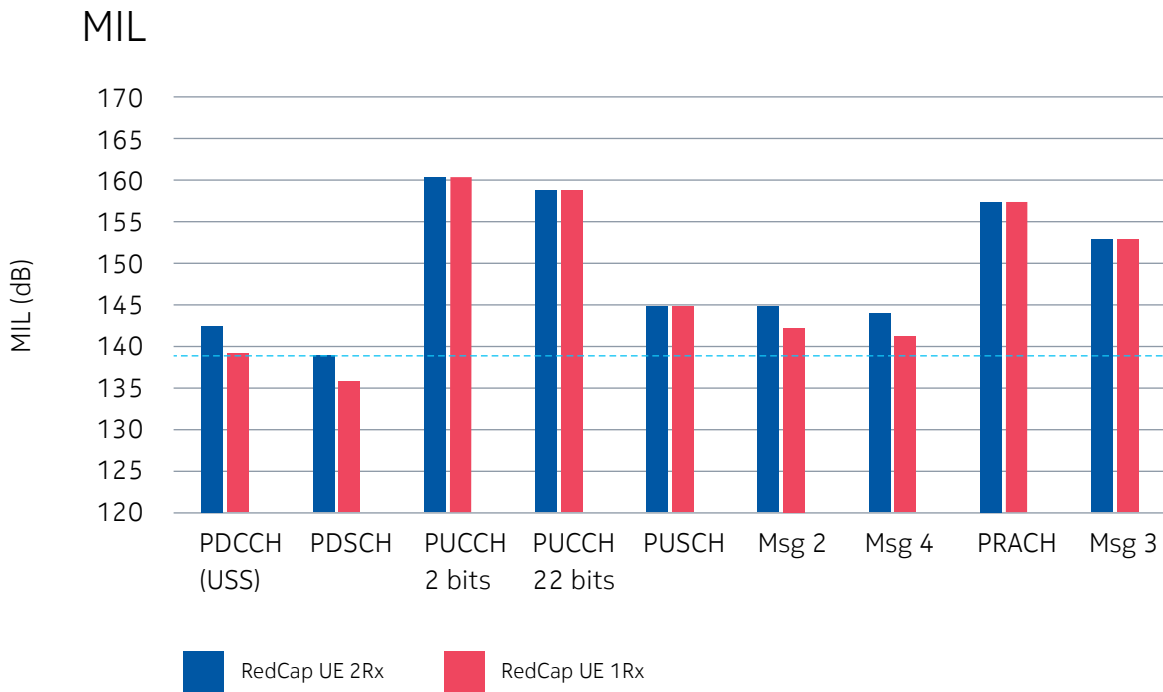
A similar plot of the MIL for the Rural 700 MHz scenario is shown in Figure 5. In this case, it is seen that Msg3 is the limiting channel for the reference NR UE. Therefore, a coverage recovery of 3 dB is required for Msg3 transmission for the RedCap UE corresponding to the amount of reduced antenna efficiency. Again, due to the vastly better MIL for the downlink channels of the reference NR UE, performance loss for the RedCap UE is not enough to cause the MIL to fall below the reference MIL, and hence coverage recovery is not required for the downlink channels.

Figure 5. Reference NR UE and RedCap UE link budget for different channels in the rural 700 MHz (FDD) scenario



The MIL is shown in Figure 6 for the indoor 28 GHz scenario. It is seen that PDSCH is the limiting channel in this scenario primarily due to the transmit power assumptions. For this channel, the RedCap UE experiences a loss of about 4.3 dB, which is the amount of coverage recovery necessary. The channel that has the next larger MIL is PDCCH, but the MIL for the RedCap UE is about the same as the reference MIL and hence coverage recovery is not required. Most of the uplink channels have a significantly higher MIL. At FR2, there is no assumption of reduced antenna efficiency for the RedCap UE. As a result, the MIL for the uplink channels is the same as for the RedCap UE. The MIL of the downlink channels is degraded due to performance loss from reducing the number of Rx antennas to one. The figure shows, however, that the loss experienced by channels other than PDSCH does not reduce their MIL below that of the reference MIL.

Figure 6. Reference NR UE and RedCap UE link budget for different channels in the indoor 28 GHz (TDD) scenario



Various techniques can be considered for improving the performance and recovering coverage of the data and control channels. Among them, the important ones for the data and control channels are as follows:

- Data (PDSCH and PUSCH) use repetition and frequency hopping to increase the received SNR
- Control (PDCCH) uses repetition and large aggregation levels to increase the received SNR and compact scheduling grants to reduce the number of bits to be transmitted.

Based on an extensive link budget analysis, it was determined that in the scenarios considered, a coverage recovery of up to 3 dB may be required for PUSCH and/or Msg3 for a RedCap UE. It was determined, however, that it is not necessary to specify coverage recovery solutions for RedCap devices. Thus, the work item for RedCap device specification does not include any objectives for coverage recovery. There is a separate Rel-17 NR Coverage Enhancement work item, and it was noted that uplink coverage enhancement solutions specified in that work item shall be assumed to also be available to RedCap UEs for coverage recovery.

In summary, RedCap UEs may experience some coverage loss due to their RedCap features. However, the coverage loss can be recovered through legacy NR features as well as Rel-17 coverage enhancement features specified for NR devices. Thus, current network deployments that provide full coverage to NR UEs can also provide full coverage to RedCap UEs. Therefore, site densification will not be necessary to provide full coverage to RedCap UEs.

Performance analysis

The impact of UE complexity reduction on the system-level performance is studied by using the Nokia 5G simulator. The simulation environment is an urban macro-cell network of seven sites with three sectors per site with the inter-site distance of 500 m. The simulations measure the system-level performance of uplink and DL. We assume a TDD system where eight and two subframes, respectively, are allocated for downlink and uplink operations. For simulations, the total system bandwidth of the network is 100 MHz, the subcarrier spacing is 30 kHz, and the carrier frequency is 2.6 GHz. Also, we run simulations for full buffer and file transfer protocol model 3 (FTP3) data traffic models, respectively.

For RedCap UE and enhanced Mobile Broadband (eMBB) UE, the different sets of parameters are used. RedCap UE and eMBB UE, respectively, use 20 MHz and 100 MHz of the total system bandwidth. Also, RedCap UE is assumed to have one or two receiving antennas, and eMBB UE use four receiving antennas. Then, for both RedCap UE and eMBB UE, it is assumed that the number of downlink layers equals the number of receiving antennas. Next, RedCap UE use 64-QAM in downlink and 16-QAM in uplink, and eMBB UE use 256-QAM in downlink and 64-QAM in uplink.

We first run simulations to measure the UE throughput with different RedCap UE ratios (expressed as a percentage of all UE) from 0% to 100%. In the simulations, we use the FTP3 model for bursty traffic with a file size of 500 Kbytes. Also, the traffic arrival rate is adaptively set so that the cell utilization level is medium (resource utilization is between 30% and 50%). First, we observe the 50th percentile user packet throughput of all UE. When the RedCap UE ratio is 0% and 100%, the throughput of all UE is 300.05 Mbps and 22.28 Mbps in downlink and 35.77 Mbps and 7.15 Mbps in uplink, respectively. The results show that deploying additional RedCap UE can decrease the user packet throughput of all UE in the network. Also, we observe the 50th percentile user packet throughput of eMBB UE. When the RedCap UE ratio is 25% and 50%, the eMBB UE's throughput in downlink is 407.42 Mbps and 413.37 Mbps and in uplink, 35.71 Mbps and 36.16 Mbps, respectively. Therefore, the results show that the deployment of RedCap UE has less impact on the downlink and uplink user packet throughput of eMBB UE on the same network. Additional results may be found in [3].

The cell average spectral efficiency is then analyzed for eMBB and RedCap UE in downlink and uplink with the different RedCap UE ratios. In the analysis, FTP3 traffic model is used. The definition of the cell average spectral efficiency is given by:

$$\text{Spectral efficiency [bps/Hz]} = (\text{cell average throughput}) / (\text{bandwidth} \times \text{resource utilization})$$

As a result, the cell average spectral efficiency can be reduced as more RedCap UE are deployed on a network. For instance, the uplink spectral efficiency of all UEs decreases by 20.4% if the RedCap UE ratio changes from 0% to 100%. Also, the downlink spectral efficiency can decrease up to 64.1% if the network only consists of RedCap UEs. Therefore, the deployment of RedCap UEs has less impact on uplink spectral efficiency than on downlink spectral efficiency.

Figure 7. Cell average spectral efficiency of eMBB UE and RedCap UE with full buffer traffic model in downlink and uplink

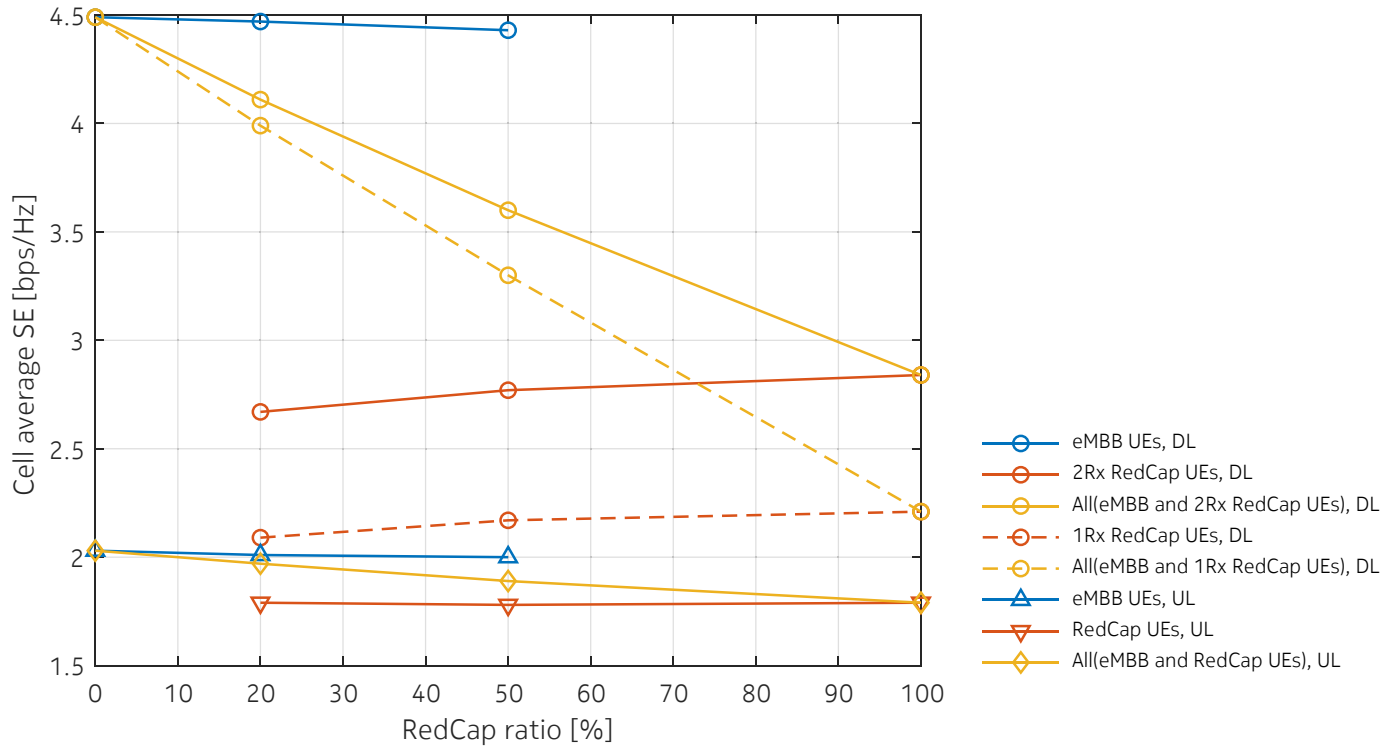


Figure 7 shows the uplink and downlink cell spectral efficiency when the full buffer traffic model is used with the different RedCap UE ratios ranging from 0% to 100%. In Figure 7, we can see that the cell spectral efficiency increases with the number of RedCap UE receiving antennas. For instance, if RedCap UEs increase the number of receiving antennas from one to two, the cell spectral efficiency increment can be up to 27.7% when the simulation environment has the same number of eMBB and RedCap UE on the network.

Figure 7 also shows that the cell spectral efficiency of eMBB UEs is less impacted by the ratio of RedCap UEs on a network especially when increasing the RedCap UE ratio from 0% to 50%, which decreases the downlink spectral efficiency by 1.3% for eMBB UEs. Also, in Figure 7, the cell spectral efficiency of RedCap UEs can increase with the RedCap ratio. For instance, if a network increases the ratio of RedCap UEs from 20% to 50%, spectral efficiency can increase up to 6.4% for RedCap UEs with two receiving antennas in downlink.

Reduced capability device evolution

Table 5 provides an overview of 5G IoT devices up to Rel-17. From the table, it is seen that there is still a large gap in the peak data rates of massive IoT devices compared to RedCap devices. For instance, LTE-M devices can support peak rates of 2.4 Mbps in the downlink and 2.6 Mbps in the uplink, while RedCap devices can support 85 Mbps in the downlink and 91 Mbps in the uplink. In Rel-18, 3GPP is studying further device complexity reduction in FR1. The goal is to introduce lower-tier devices between massive IoT and Rel-17 RedCap devices. The supported peak data rate for the new Rel-18 devices is expected to be around 10 Mbps.

Table 5. 5G IoT Devices

	Massive IoT				Critical IoT URLLC	Mid-range IoT	
	LTE-M			NB-IoT		RedCap	
	Cat-M1	Cat-M2	Cat-NB1	Cat-NB2		FR1	FR2
Downlink peak rate	300 kbps*	2.4 Mbps*	27 kbps	127 kbps		85 Mbps	209 Mbps (50:50 DL/UL)
Uplink peak rate	475 kbps*	2.6 Mbps*	62 kbps	159 kbps		91 Mbps	229 Mbps (50:50 DL/UL)
Coverage (maximum coupling loss)	164 dB					144 dB	
Maximum cell size	100 km		120 km			100 km	
Battery life	10 years						
Latency	Up to 10 s for successful delivery of an application layer packet from the device to RAN				1 ms User plane latency	5-10 ms safety sensor, 100 ms others User plane latency	
Reliability	99% - 99.9%				99.999%	99.99%	
UE bandwidth	1.4 MHz	5 MHz	200 kHz	200 kHz		20 MHz	100 MHz
UE power class	14/20/23 dBm					23 dBm	
Frequency range	FR1					FR1, FR2	

* HD-FDD

Potential solutions for Rel-18 RedCap FR1 devices include device bandwidth reduction to 5 MHz, restricting the bandwidth of the data channels, and relaxed UE processing timelines. Our preliminary analysis shows that approximately 25–30% complexity reduction with respect to Rel-17 device is possible. The lower-complexity device includes 5 MHz bandwidth, half-duplex and supports 64-QAM on the downlink and 16-QAM on the uplink. Further restrictions such as limiting peak data rates can be considered to provide additional cost reduction.

Note that reducing the device bandwidth to 5 MHz will have some impacts. First, when SSB is deployed using sub-carrier spacing (SCS) of 30 kHz, the physical broadcast channel (PBCH) occupies 7.2 MHz, which is beyond the device bandwidth. It is still possible for the device to decode the PBCH using only signals that fall within 5 MHz but the performance will be degraded. Second, only CORESET#0 with 15 kHz SCS and 24 physical resource blocks (PRBs) in size can be supported, so this limits the configuration of CORESET#0.

Conclusions

In this paper, we provide an overview of RedCap devices being introduced in 3GPP Rel-17 to support mid-range IoT use cases such as industrial wireless sensors, wearables and video surveillance. Our analysis shows that significant complexity reduction can be achieved. When compared to reference Rel-17 NR devices, it is seen that a total complexity reduction of approximately 70% for FR1 (FDD and TDD) and 50% for FR2 (TDD only) can be achieved. The specification changes required to support complexity reduction features are seen as moderate, with most modifications related to BWP operations, half-duplex operations, and new requirements for devices with one Rx branch and for NCD-SSB operations.

Network coverage analysis shows that the impact from the introduction of RedCap devices is small. Coverage analysis, based on typical NR deployment scenarios and target cell edge data rates, shows that coverage recovery is generally not needed despite the performance loss in the downlink. This is because the coverage of the downlink channels is significantly better than the bottleneck uplink channel, thus coverage is not impacted even with some performance loss. In FR1, however, reduced antenna efficiency due to device size limitations may mean that coverage recovery is needed. Coverage recovery can be achieved using techniques introduced in another Rel-17 work item on coverage enhancement.

Network capacity analysis shows that the impact to system spectral efficiency and capacity is minor when the system is not heavily loaded. Furthermore, there is also minor impact to the performance of eMBB users.

In addition to complexity reduction, two power saving features — eDRX and RRM relaxation — are also introduced to provide additional power consumption reduction. The first feature can provide significant power savings to devices with infrequent data transmission, while the second feature is beneficial to stationary devices.

Technical specifications for RedCap devices are expected to be completed by September 2022, with commercial deployment expected 18–24 months thereafter. Currently, wearables such as smart watches and goggles are seen as the most attractive initial use case. On the network side, gNBs can simply be upgraded via software to support RedCap devices. Coexistence with other NR devices is seamless since they share the same NR design. Furthermore, the existing network footprint can be used based on our network coverage analysis.

In Rel-18, 3GPP is studying further device complexity reduction in FR1. The goal is to introduce lower-tier devices between massive IoT and Rel-17 RedCap devices. The supported peak data rate for the new Rel-18 devices is expected to be around 10 Mbps.

References

1. R. Ratasuk, N. Mangalvedhe, D. Bhatoolaul and A. Ghosh, “LTE-M Evolution Towards 5G Massive MTC,” 2017 IEEE Globecom Workshops, Singapore, 2017.
2. 3GPP TR 38.875, “Study on support of reduced capability NR devices”, V1.0.0, December 2020.
3. R. Ratasuk, N. Mangalvedhe, G. Lee and D. Bhatoolaul, “Reduced Capability Devices for 5G IoT,” 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2021.

5G empowering Uncrewed Aerial Systems

Contents	
Introduction	3
Motivation for supporting UAS in 5G	4
3GPP Release 17 system architecture enhancements	6
UAV identification	6
UAV authentication and authorization	9
UAV tracking	10
5G-Advanced outlook	11
Conclusion	12
References	13

Introduction

Cellular networks are usually designed to serve users on the ground, for example, base station antennas are typically tilted down to ensure a good tradeoff between coverage and interference. Such network designs, however, may not be optimal for users in the air (e.g., UEs used in drones) and may result in higher levels of interference as the propagation conditions experienced by them are different compared to ground-based users. 3GPP introduced the concept of Aerial UE subscription in LTE Release 15, to identify UEs used with aerial vehicles such as drones or uncrewed aerial vehicles (UAV). At the same time, 3GPP also introduced RAN features like flight path reporting, which helps ensure reliable communications for UAVs over cellular networks. This opens up a large number of use cases for UAVs such as package delivery, power line inspection and emergency response, among others.

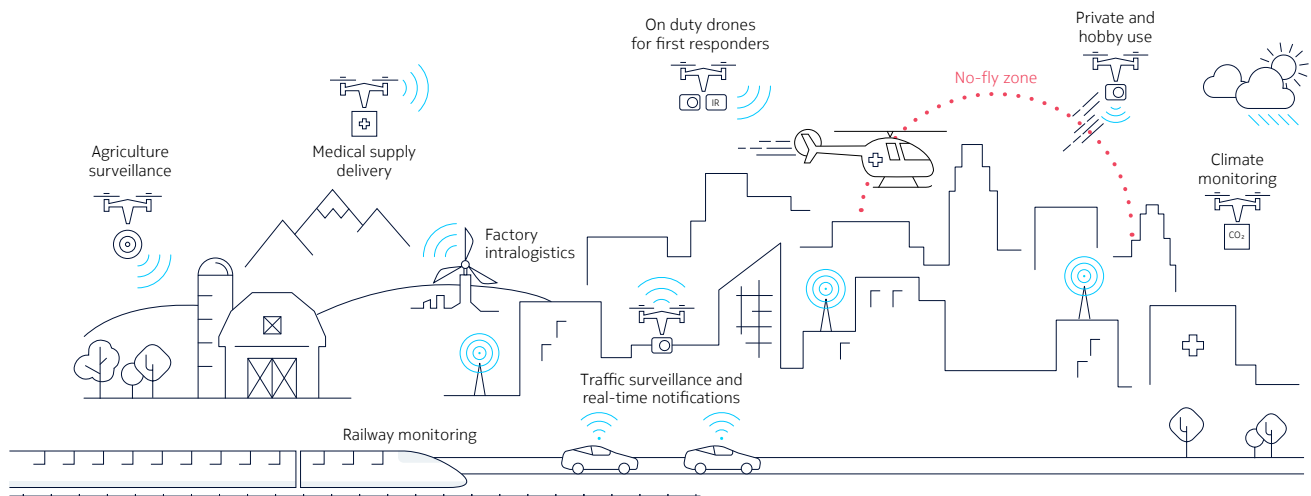
These LTE RAN features, however, are currently not available in 5G NR. In Release 17, 3GPP has introduced system architecture enhancements to support integration of aviation level application servers or systems, e.g., UAS Service Supplier (USS) and UAS Traffic Management (UTM) systems. These provide various services like identification, authentication and authorization, and tracking of UAS. The basic idea is to encourage adoption of 3GPP technologies for various UAS communications and use cases. This can be accomplished by exposing some of the proven 3GPP capabilities, such as identification, authentication and authorization, through industry standard open interfaces for integration with the aviation entities. Further, the Open Generation consortium was created in the US to focus initially on use cases related to operating 5G-equipped drones over the United States [8].

In this white paper, we highlight the motivation for supporting UAS, describe different deployment models and explain the features that were introduced by 3GPP Release 17. We also look at some of the aspects that will be addressed as part of 5G-Advanced in upcoming releases (i.e., Release 18 and beyond) to bring UAV support in NR on par with LTE and, in time, beyond.

Motivation for supporting UAS in 5G

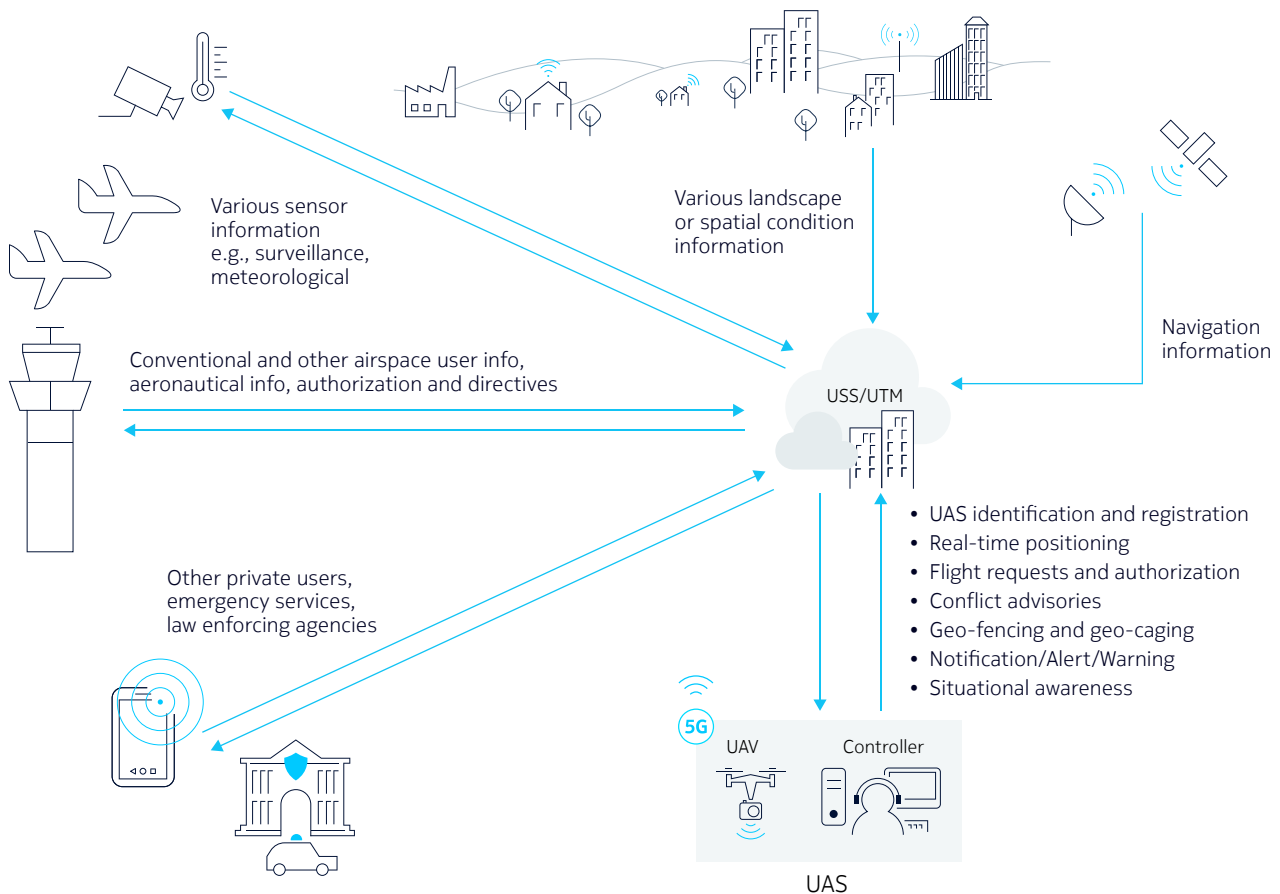
Uncrewed Aerial Systems (UAS) consist of an aircraft or drone, also popularly known as Uncrewed Aerial Vehicle (UAV), and the equipment needed to control it remotely, also called the UAV controller (UAV-C) or sometimes referred to as a ground station controller. The UAS market is anticipated to register a CAGR of around 15% in the coming years to reach USD 47 billion by 2025 [7]. Over the last few years, the UAS market has moved from being primarily driven by either military or individual entertainment devices to a more significant presence in the commercial world. There are numerous use cases for commercial UAS operations, which are illustrated in Figure 1. UAVs are increasingly used in agriculture (e.g., surveying, crop dusting, spraying of insecticides), commercial delivery, property surveillance, traffic monitoring, medical supply delivery and many more. UAVs can also play a significant role in public safety and mission-critical services, for example, to localize and rescue people by gathering visual and sensory information. From the network operations point of view, UAVs can be used for network maintenance and various network measurements. UAVs carrying base station equipment on board (i.e., movable or flying base stations), are also seen as a possible alternative for providing 5G network capabilities to improve connectivity services in a highly crowded environment (e.g., during large events) or for extending coverage in areas with otherwise limited coverage.

Figure 1. UAV use cases



Due to the huge number of potential applications of UAVs in commercial use cases and the anticipated increase in the number of UAVs flying around, the regulatory authorities from most regions and countries are coming up with rules and policies related to UAV registrations, flight permissions, remote identification, tracking and traffic management, including the operation of UAVs both in and beyond visual line of sight (BVLOS). Safely operating a UAV BVLOS requires an extremely reliable command and control (C2) link between the controller and the aircraft, even more in the case of a swarm of drones. 3GPP-defined cellular networks are seen as an attractive way to provide the C2 link for BVLOS flight operations, due to their global presence and inherent capability of providing ultra-reliable and real-time long-distance data transmission capabilities.

Figure 2. UTM Architecture



The Global UTM Association (GUTMA), a consortium of worldwide UAS traffic management (UTM) stakeholders, has published a high-level UTM architecture document. Its purpose is to support and accelerate UTM interoperability with an aim to foster the safe, secure and efficient integration of UAVs in national airspace systems. The UTM system is integrated with various other communication infrastructures (see Figure 2) and needs to collect data or communicate with many other systems to provide the UTM services. Some of the key functionalities of the UTM include UAS registration and identity management, flight plan, permissions and directives, surveillance and tracking, conflict detection and contingency management, and airspace management.

When a UAV is using cellular connectivity, the 3GPP network can be used to facilitate or support the UTM in performing some of these tasks. For example, the 3GPP has a proven identification, authentication and authorization framework that can be extended to the aviation industry for UAS identification, flight permissions and authorization of air space. The 3GPP network can also provide reliable, secure, stable and real-time air-to-ground or ground-to-air communication capability for enabling BVLOS operation of UAS. Supervision systems for UAVs is not well established currently, and there is always the threat of unregulated drones flights. 3GPP can assist the aviation industry with its real-time location and tracking features.

3GPP Release 17 system architecture enhancements

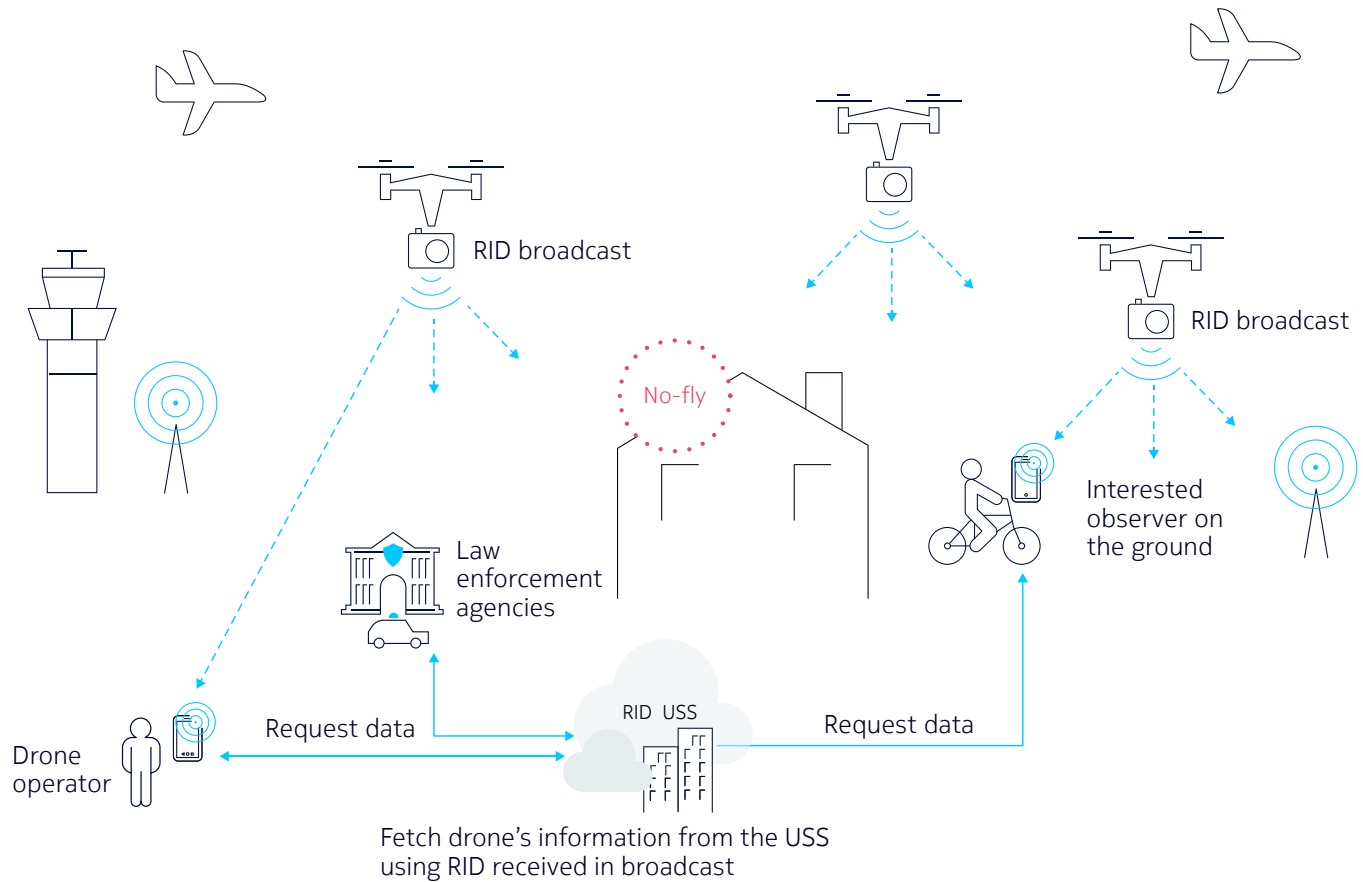
3GPP Technical Specification Group Service and System Aspects Work Group 1 (TSG SA WG1, Services) has specified the service requirements identified for operation of UAS via the 3GPP system, which includes requirements for meeting the business, security and public safety needs for the remote identification and tracking of UAS. Based on this work, in 3GPP Release 17, the Technical Specification Group Service and System Aspects Work Group 2 (TSG SA WG2, System Architecture and Services) has studied the impacts on system architecture and has completed the work of specifying the first phase of architectural enhancements needed to support UAS operations using cellular networks (Support of Uncrewed Aerial Systems Connectivity, Identification and Tracking [1]). The corresponding stage 3 specification (i.e., APIs definition) is expected to be completed by June 2022. The work done by TSG SA WG2 in 3GPP Release 17 has no impact or dependency on the radio access network (RAN) and supports seamless 5GS to EPS interworking [2][3]. In what follows, we look at some of the system architecture enhancements specified in Release 17.

UAV identification

At the aviation level a UAS is identified using a unique identifier assigned by the regional authorities (e.g., CAA Level UAV identifier) or the USS (e.g., USS ID). The unique identifier could be a hardware serial number or registration number or a session ID. This unique identifier can be seen as similar to a car registration number plate. Using this identifier, one can find out information such as the owner of the UAV and its registered address. One of the typical requirements for safely operating UAVs in an airspace is remote identification. Any person on the ground should be able to identify a UAV flying nearby.

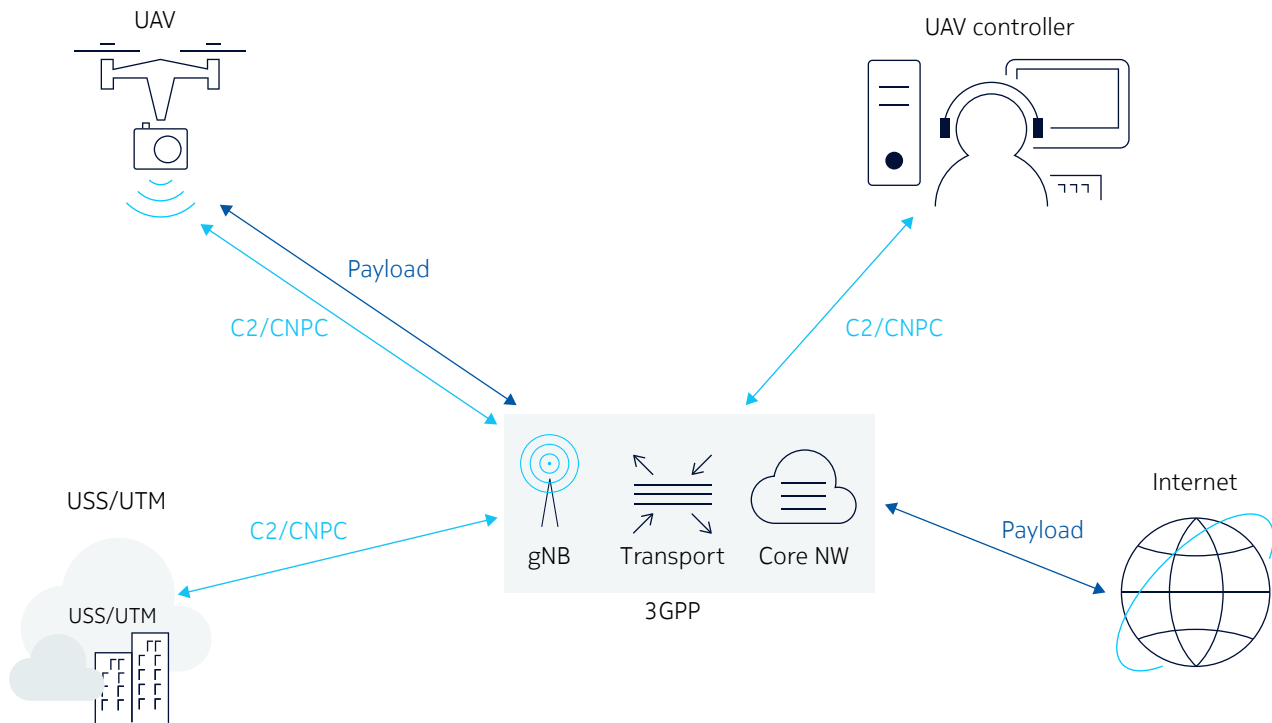
Various regional regulations require the UAV to be able to broadcast its identification information for the purpose of remote identification (RID). The remote identification information contains information apart from the UAV identifier, such as the ground station controller location, its take off location, its current location (including altitude) and emergency state indication. Any person on the ground should be able to intercept the broadcasted information and fetch the necessary information to identify the UAV, where it is going and where its controller is located. This enables people on the ground and other airspace users to identify who is using the airspace around them.

Figure 3. Broadcast remote identification (BRID) of UAV



From an operator's network (or 3GPP) perspective, however, this remote Identification (RID) of UAVs is not sufficient. The network operator (3GPP) needs to identify when user equipment (UE) is used with a UAV, so that differentiated services can be provided to the UE when it is accessing the network for various UAS operations. The UAV may use an operator's network for various types of communication. These communications can be grouped into two broad categories based on their requirements. First is a payload communication, which can be any application data such as surveillance data collected by a surveillance drone, 4K video or sensor data. From the perspective of safe UAS operation, this category of (payload) communication can be considered as non-critical and can be delivered depending on the type of application. The second category of communication is control and non-payload communication (CNPC), sometimes also referred to as command and control (C2) communication. This is critical for UAS operation and safe airspace usage and requires very high reliability. It must also have very low end-to-end latency.

Figure 4. 3GPP network can apply different policies and QoS control for different types of communication with the UE



The 3GPP network can differentiate these various types of communication (see Figure 4) based on factors like the Data Network Name (DNN) or the Network Slice that the UE is trying to access. 3GPP Release 17 has defined aerial UE subscription to distinguish UEs that are allowed to be used with aerial vehicles such as UAV. The subscription enhancement also includes an indication in the subscription information of the data network name (DNN) and network slice to indicate if the DNN or slice is used for such critical UAS communications, such as C2/CNPC or communication with USS.

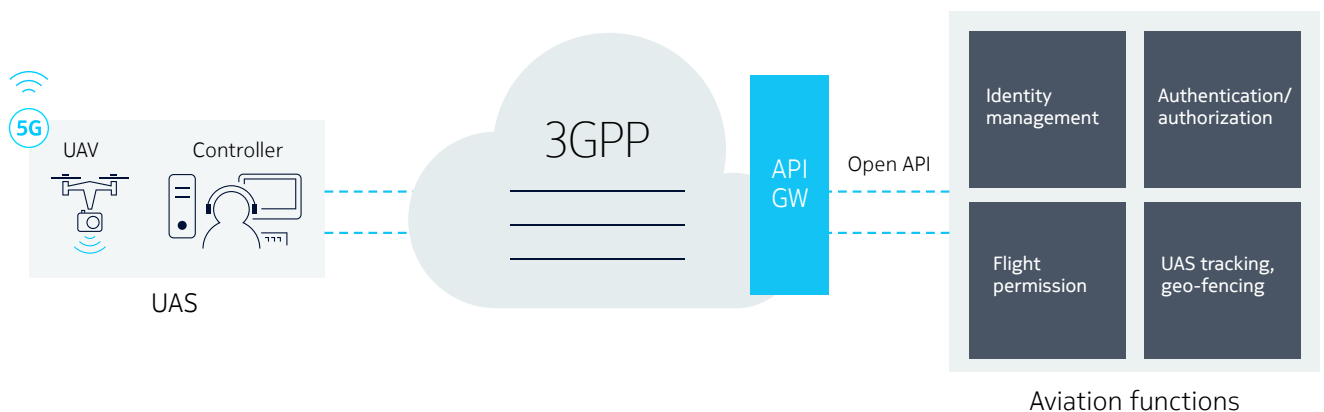
UAV authentication and authorization

The regulatory and aviation authorities in many countries have defined, or are in the process of defining, rules for registration of UAVs with a central entity. Additionally, there are region- and country-specific regulations for prior flight path approval and flight permissions before flying a UAV, sometimes also referred to as “no permission, no take-off” (NPNT). Currently all processes, such as authorization of UAVs and flight permissions, require direct human interaction and may not be sustainable considering the huge potential growth in the number of UAVs and other flying objects (e.g., urban air mobility or UAM) in the future.

There is a need for a more dynamic way to authorize a UAV to use the airspace. From the 3GPP network point of view, when a UE with “aerial subscription” is trying to access UAS-specific services through the network, it may want to verify that the UE is used with an authorized UAV. Some regional regulations may also require the network to authenticate and authorize the UAV identity by an aviation-level application server (e.g., USS/UTM) before providing any cellular service to the UE used with the UAV. If 3GPP connectivity is also used for a C2 link between the UAV and the UAV controller, the 3GPP network can also provide mechanisms where a third-party application can control whether and when the C2 communication can be allowed. For example, the 3GPP network can give the control to the regional aviation-level applications managing the airspace (e.g., USS/UTM) whether to allow the UAV to get airborne and have a C2 link setup with the UAV controller. To support such use cases, 3GPP Release 17 has defined an authentication and authorization framework that can be utilized by the aviation industry to automate and provide dynamic authorization for flight path and airspace usage.

The 3GPP Release 17 system architecture is enhanced to provide an industry-standard open API-based framework (see Figure 5) for the aviation entities managing the UAV registration, flight planning and authorization, using the exposure services framework. The APIs provide an optional capability at the 3GPP network to first get the UAV identities authenticated by an external aviation entity before allowing any connectivity services to UE(s) used with UAV. It also enables dynamic flight path planning and authorization to use the airspace and controlling the pairing of the UAV and UAV controller link for C2 communication. The framework can also be used to request specific quality of service (QoS) characteristics and policy configuration for the communication link. The framework is designed independently of the application protocol, and the 3GPP network can transparently relay the authentication key exchanges and other authentication and authorization data between the UAV and the aviation level entity (e.g., USS/UTM), in a secured manner.

Figure 5. API-based service exposure from 3GPP

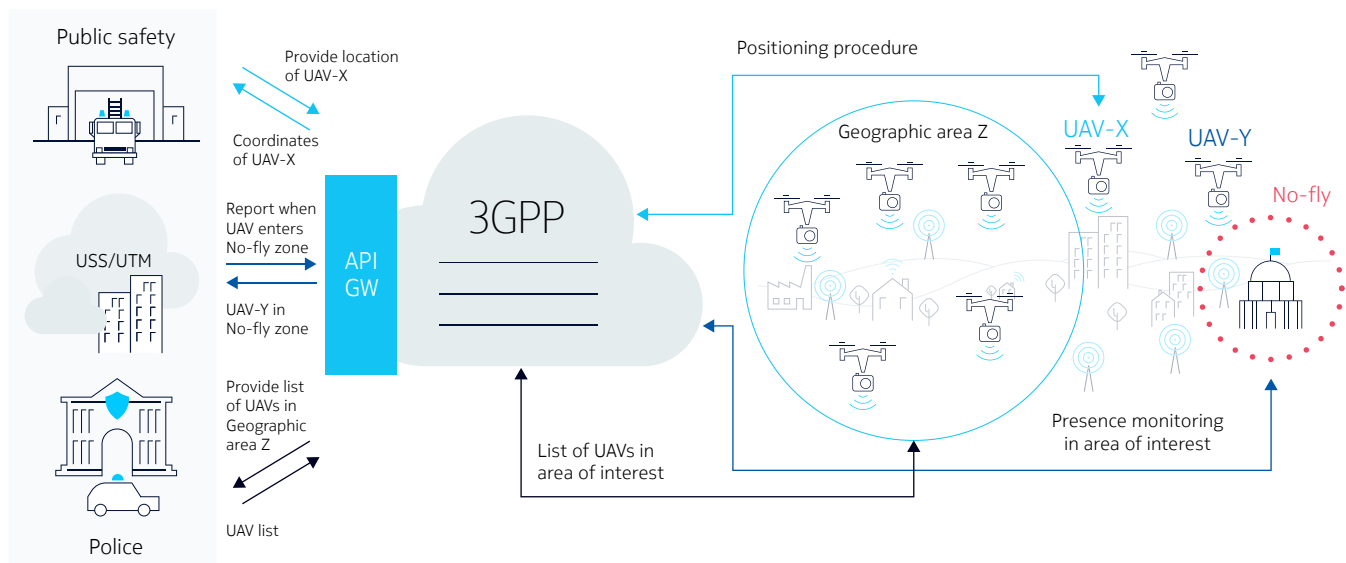


As a ubiquitous 5G network may not be available everywhere, a common identification, authentication and authorization framework has been designed for both 5GC and EPC to provide seamless mobility between 5G and 4G networks that is independent of the access network and has no impact on legacy EPC network elements.

UAV tracking

As the UAV is using a UE for cellular connectivity, the 3GPP system can support the aviation entities responsible for planning and managing the airspace, controlling the air traffic, and tracking the UAV(s) by exposing various location services and event monitoring capabilities. 3GPP Release 17 provides enhancements to the existing location services and event exposure framework to provide capabilities catering to the unique requirements from the aviation industry (see Figure 6).

Figure 6. Location and tracking services provided by 3GPP



Location services: This can be used by an authorized application server outside the 3GPP network to request current (and periodic) location of a UE, using an open API. The USS/UTM or other aviation-level application and any public safety or law enforcement agencies can use the network-provided location information to complement the existing mechanisms for tracking a UAV.

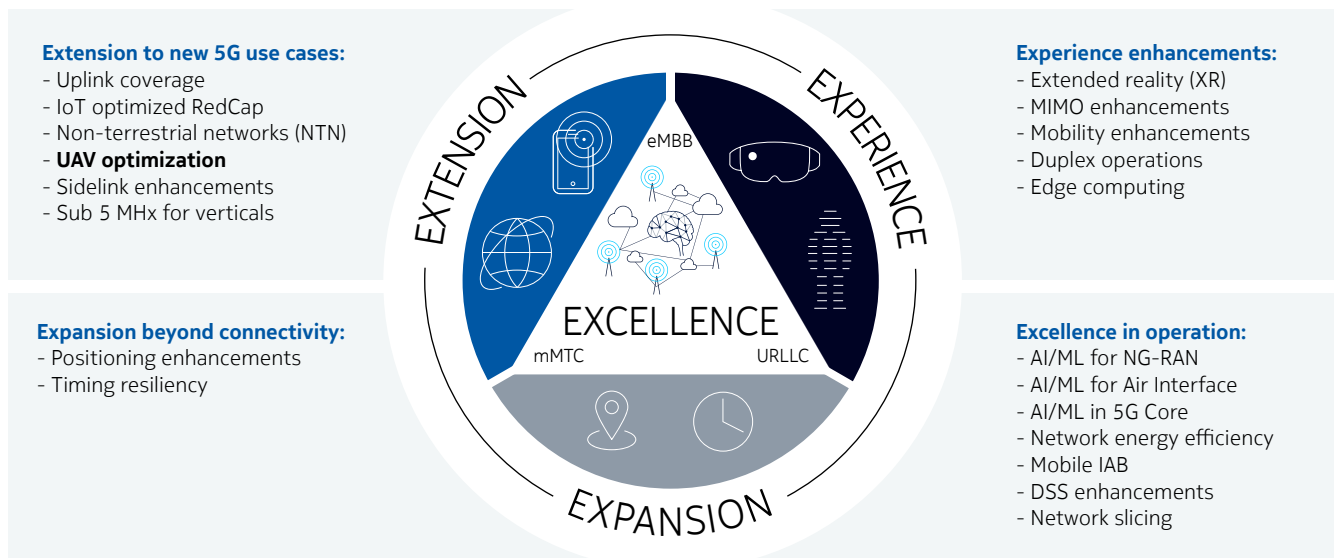
Presence Monitoring: The 3GPP network can be configured to monitor the movement of a UE used with a UAV and report when the UAV moves into, or moves outside of, a specified geographic area. This can be useful to support use cases like geo-fencing or no-fly-zone monitoring. This feature can be used in two different ways. First, it can monitor and prevent a UAV from entering sensitive and no-fly zones such as airports, police stations and government complexes. Second, it can monitor and restrict the movement of a UAV within a specified geographic area or corridor, for instance, allowing the UAV to fly only over an agriculture field in rural areas or to restrict the movement of the UAV within a pre-defined drone corridor.

List UAVs operating in an area of interest: Let us assume that a police van notices many UAVs flying in an area. It would be a tedious task for it to read the RID broadcasts from each UAV and then find out the total number of UAVs in the area or their identity. 3GPP Release 17 provides open APIs that any authorized application server outside the 3GPP system (e.g., hosted by USS/UTM, public safety or the police) can use to query the 3GPP network and find out the list of UAVs served by the network in a specific geographic area.

5G-Advanced outlook

3GPP has already defined a set of functionalities and features that will be a part of the 5G-Advanced Release 18 package. These functionalities can be grouped into four areas: providing new levels of **experience**, network **extension** into new areas, mobile network **expansion** beyond connectivity, and providing operational support **excellence** (see Figure 7). Extending UAV support in Release 18 will be an important part of the ‘extension’ block of features. This will not only enable 5G-Advanced networks to expand their reach to new areas geographically but, in the case of UAV, to new service areas.

Figure 7. 5G-Advanced, Release 18



While the UAV-related studies in Release 15 for LTE showed that cellular networks could be used for connecting UAVs, 3GPP Release 17 provides further architecture enhancements for supporting identification, authentication and tracking of UAS, and it allows both LTE and NR to provide extra information to the UTM. This is only a first step towards improving the safe and secure integration of UAVs with cellular communication capabilities into the network. 5G-Advanced (starting from Release 18) will introduce holistic support for UAVs. Both the radio (5G NR) and core network will provide carrier-grade UAV managed services capabilities, taking them beyond what is already provided by LTE.

3GPP TSG SA WG2 has agreed on a study item in Release 18, where it will investigate how to leverage existing mechanisms of device-to-device direct discovery and device-to-device direct communication for meeting remote identification (RID) broadcast requirements and for UAV-to-UAV direct communication to support requirements related to detect-and-avoid (DAA) mechanisms that are currently being developed by the aviation community (e.g., GUTMA). The study will also investigate mechanisms to support C2 communications between the UAV and the UAV controller using the device-to-device (D2D) direct communication.

The 3GPP TSG RAN has also agreed on a Release 18 work item, with Nokia as rapporteur, for 5G NR enhancements to support UAV. The RAN is generally highly optimized for users on the ground and further enhancements will be introduced in Release 18 on the RAN side. Similar features as introduced in release 15 for LTE for UAV will be brought to NR first, such as reporting of flight path, location, height and speed and measurement optimizations. Subscription-based identification will then be specified and whether

enhancements are needed to improve co-existence between users on the ground and UAVs, as UAVs at heights of 50–100 meters cause considerably more interference than users on the ground. One solution for interference mitigation from UAVs is to utilize beamforming capabilities on the UAV. It will also be investigated whether enhancements are needed for UAV identification broadcast or if existing mechanisms are sufficient.

Looking beyond Release 18, 3GPP will likely expand support for emerging use cases and requirements coming from different fora of the aviation industry. One such forum is the Aerial Connectivity Joint Activity (ACJA), which is a joint initiative by GSMA [6] and GUTMA (Global UTM Association) [4]. ACJA is defining various UAS models where both the UAV and its controller will use cellular connectivity [5]. Further, the UAVs can also be grouped into different categories and mission types such as public safety, commercial payload vehicle and critical surveillance. UAVs are also seen as a future means for urban mobility (e.g., drone taxis), also popularly referred to as urban air mobility (UAM). Each category of UAVs will have different communication and KPI requirements. The 3GPP network will need to provide support for these UAS models and provide differentiated services and prioritization of services based on various UAS or UAM categories and mission types.

In summary, while current cellular networks can be used to serve UAVs, Release 17 has introduced enhancements to fulfill the increased requirements coming from the aviation authorities. Release 18 of NR and other system features will improve the co-existence with terrestrial users even further and extend support for UAVs into new areas and use cases.

Conclusion

5G cellular networks are a natural connectivity choice for UAVs due to the huge, anticipated growth in the number of UAVs and the realization of new use cases, which will require ultra-reliable and real-time communication. Nokia has promoted the use of 5G for UAVs in ACJA (see our earlier white paper). While current cellular networks can be used to serve UAVs, safe integration and operation of UAVs needs system features like UAV identification, authentication and flight authorization and tracking. Adding these various system features in combination with RAN features will accelerate the adoption of cellular communications for critical drone communications such as command and control (C2), broadcast remote identification (BRID) and networked remote identification (NRID). These additional features will also enable new use cases and service capabilities, as well as dynamic airspace management, flight planning, authorization, tracking, conformance monitoring and conflict detection and contingency management. Finally, Release 18 of NR, where Nokia is the rapporteur for the RAN work item on UAV enhancements, will also bring further improvements for co-existence of aerial users with terrestrial users.

References

- [1] 3GPP TS 23.256: “Support of Uncrewed Aerial Systems (UAS) connectivity, identification and tracking”.
- [2] 3GPP TS 23.501: “System Architecture for the 5G System”.
- [3] 3GPP TS 23.502: “Procedures for the 5G System (5GS)”.
- [4] UAS Traffic Management Architecture (https://www.gutma.org/docs/Global_UTM_Architecture_V1.pdf)
- [5] Aerial Connectivity Joint Activity (<https://gutma.org/acja>)
- [6] GSMA Advanced Air Mobility (<https://www.gsma.com/iot/aviation/>)
- [7] Drones Market - Growth, Trends, and Forecast (2020 - 2025), 2020 (https://www.reportlinker.com/p05881491/?utm_source=GNW)
- [8] OPEN GENERATION UNCREWED AIRCRAFT SYSTEMS (UAS) USE CASES
https://f.hubspotusercontent20.net/hubfs/7754670/Open_Generation/Open%20Gen%20Reports/ME_OpenGen_UAS_UseCases_Feb2022.pdf?__hstc=145223019.31cf68d6b1413bbc283aad8a64325266.1644257089135.1644257089135.1644257089135.1&__hssc=145223019.1.1645931400109&__hsfp=2185856287&hsCtaTracking=ec0d1dbc-6896-4867-bb72-b50ee6239b71%7Cf3a167e7-9bba-41d6-9079-a2d8c9a9de4d

Abbreviations

1PPS	One Pulse Per Second	CORESET	Control resource set
3GPP	3rd Generation Partnership Project	CP	Control Plane
5GC	5G Core	CPE	Customer-premises equipment
5GS	5G System	CSI-RS	Channel state information reference signal
AAA	Authentication, Authorization and Accounting	DAA	Detect and Avoid
ACJA	Aerial Connectivity Joint Activity	DC	Dual connectivity
AI	Artificial Intelligence	DCCA	Dual connectivity carrier aggregation
AMF	Access and Mobility Management Function	DCS	Default Credential Server
API	Application Programming Interface	DL	Downlink
AR	Augmented Reality	DL-AoD	Downlink Angle of Departure
AUSF	Authentication Server Function	DL-TDOA	Downlink Time Difference of Arrival
BLER	Block error rate	DN	Data Network
BMCA	Best Master Clock Algorithm	DNN	Data Network Name
BVLOS	Beyond Visual Line of Sight	DPS	Dynamic point selection
BW	Bandwidth	DRX	Discontinuous reception
BWP	Bandwidth part	DS-TT	Device-side TSN translator
C2	Command & Control	E-CID	Enhanced cell ID
CA	Carrier aggregation	eDRX	Extended discontinuous reception
CAA	Civil Aviation Authority	eMBB	enhanced mobile broadband
CAG	Closed Access Group	EPC	Extended packet core
CBRS	Citizens Broadband Radio Service	EUTRAN	Evolved UTRAN
CD-SSB	Cell-defining synchronization signal block	FD	Full duplex
CG	Cell Group	FDD	Frequency division duplex
CH	Credentials Holder	FR	Frequency range
CHO	Conditional handover	FR1	5G frequency range 1 (sub 6 GHz)
cmWave	Centimeter wave	FR2	5G frequency range 2 (mmWave)
CNPC	Control and Non-Payload Communication	GIN	Group ID for Network Selection
		GM	Grand Master
		gNB	Next generation NodeB

GNSS	Global Navigation Satellite System	Msg	Message
GPIO	General Purpose Input/Output	N3IWF	Non-3GPP Interworking Function
GPS	Global Positioning System	NB-IoT	Narrowband internet of things
gPTP	Generic Precision Time Protocol	NCD-SSB	Non-cell-defining synchronization
GSMA	Global System for Mobile communications Association	NFC	Near-field communication
GUTMA	Global UTM Association	NID	Network Identifier
HARQ	Hybrid automatic repeat request	NPN	Non-Public Network
HD	Half duplex	NR	New Radio
HFT	High-Frequency Trading	NRID	Networked Remote Identification
HSS	Home Subscriber Server	NTP	Network Time Protocol
HST	High speed train	NW-TT	Network-side TSN translator
IIoT	Industrial Internet of Things	P-CSCF	Proxy Call Session Control Function
IMS	IP Multimedia Subsystem	PBCH	Physical broadcast channel
IoT	Internet of things	PCell	Primary Cell of Master Cell Group
L1	Layer 1	PCF	Policy Function
LTE	Long Term Evolution	PDCCH	Physical downlink control channel
LTE-M	Long term evolution for machines	PDSCH	Physical downlink shared channel
MAC	Medium access control	PH	Paging hyper-frame
MCC	Mobile Country Code	PHR	Power Headroom
MCG	Master Cell Group	PLMN	Public Land Mobile Network
MCL	Maximum coupling loss	PLMN ID	Public Land Mobile Network Identity
MIL	Maximum isotropic loss	PMIC	Port Management Information Container
MIMO	Multiple-input multiple-output	PNI-NPN	Public Network Integrated Non-Public Network
ML	Machine Learning	POS	Point of Sale
mMTC	Massive Machine-Type Communications	PRACH	Physical random-access channel
mmWave	Millimeter wave	PSCell	Primary Cell of Secondary Cell Group
MNC	Mobile Network Code	PTP	Precision Time Protocol
MPL	Maximum path loss	PTW	Paging time window
MR	Multi-RAT	PUCCH	Physical uplink control channel

PVS	Provisioning Server	SSB	Synchronization signal block
QAM	Quadrature amplitude modulation	SW	Software
QoE	Quality of Experience	TA	Timing Advance
RAN	Radio Access Network	TD	Time Domain
RAT	Radio Access Technology	TDD	Time division duplex
RedCap	Reduced capability	TSC	Time Sensitive Communication
Rel.	Release	TSN	Time Sensitive Networking
RF	Radio frequency	Tx	Transmit
RFID	Radio frequency identification	UAM	Urban Air Mobility
RID	Remote Identification	UAS	Uncrewed Aerial System
RLF	Radio link failure	UAV	Uncrewed Aerial Vehicle
RRC	Radio Resource Control	UAV-C	UAV Controller
RRH	Remote Radio Head	UDM	Universal Data Management
RRM	Radio resource management	UE	User Equipment
RTT	Round-Trip-Time	UL	Uplink
Rx	Receive	UL-TDOA	Uplink time difference of arrival
SCell	Secondary Cell	UMTS	Universal mobile telecommunications system
SCG	Secondary Cell Group	UP	User Plane
SFN	System Frame Number	UPF	User Plane Function
SIB	System Information Block	URLLC	Ultra-reliable low-latency communications
SIB1	System information block type one signal block	USS	UAS Service Supplier
SINR	Signal to interference noise ratio	UTC	Coordinated Universal Time
SLA	Service Level Agreement	UTM	UAS Traffic Management
SMF	Session Management Function	UTRAN	UMTS terrestrial radio access network
SMPTE	Society of Motion Picture and Television Engineers	V2X	Vehicle to everything
SMTC	SSB-based Measurement Timing Configuration	VR	Virtual Reality
SNPN	Standalone Non-Public Network		
SNR	Signal-to-noise ratio		

Additional resources

Blogs

[Now the real work on 5G-Advanced begins](#)
[5G-Advanced's system architecture begins taking shape at 3GPP](#)
[How 5G-Advanced will help quench the world's growing thirst for capacity](#)
[Here's our first real glimpse at 5G-Advanced](#)
[Now it is time for drones to connect to mobile networks](#)
[Creativity and passion drive Devaki Chandramouli in her standards work](#)
[5G will open new possibilities in positioning](#)
[Building the 5G highway for connected cars](#)
[5G and the sustainable future: a look to 2025](#)
[6G technology leadership in the US](#)
[HAPS: Connect the unconnected](#)

Videos

[How to extract value from your 5G network? 5G-Advanced discussion](#)
[5G-Advanced: Expanding 5G for the connected world](#)
[Rel-16 RAN progress](#)
[5G evolution](#)

White paper

[5G reduced capability devices](#)
[5G empowering Uncrewed Aerial Systems](#)
[How network adaptations for 5G devices will lead to superior battery life](#)
[5G time service](#)
[A comparison of 4G and 5G authentication methods](#)
[5G Non-Public Networks](#)
[Controlling drones over cellular networks](#)
[The Evolution of 5G New Radio Positioning Technologies](#)
[Low Latency 5G for Professional Audio Transmission](#)

Thematic internet pages

[5G-Advanced: Expand and transform your connected world](#)
[Standards and patents](#)

Discover more about 5G and 5G-Advanced

NOKIA

About Nokia

At Nokia, we create technology that helps the world act together.

As a trusted partner for critical networks, we are committed to innovation and technology leadership across mobile, fixed and cloud networks. We create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

Adhering to the highest standards of integrity and security, we help build the capabilities needed for a more productive, sustainable and inclusive world.

Nokia is a registered trademark of Nokia Corporation. Other product and company names mentioned herein may be trademarks or trade names of their respective owners.

© 2023 Nokia

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

CID 212629