5G Evolution: What's Next?

Balazs Bertenyi, Chairman of 3GPP RAN

Abstract

5G networks and devices are being rolled out at an unprecedented speed in all corners of the globe. These rollouts are all based on 3GPP's Release 15 standard that includes the first version of 5G NR specifications and was released in June/2019. 3GPP has since been actively working to expand the functionalities of 5G, and this paper presents the 5G evolution path brought by 3GPP release 16, 17, and beyond toward the full 5G vision.

Timeline for 5G Evolution

The first full release of 3GPP's 5G standards was frozen with Release 15 in two distinct steps: specifications for the Non-StandAlone variant was frozen in December 2018, while specifications for the StandAlone variant were released in June 2019. The network deployments and devices we see being rolled out today are all based on these standards. They include all key functionalities for running an Extreme Mobile Broadband service using the new 5G NR radio technology.

But 5G was meant for much more than enhanced mobile broadband: it was set out to become a complete platform for all communication needs of the next decade encompassing all IoT and vertical industry needs. Reaching these ambitious goals takes several years, and for standards, it takes several releases. This is an incremental process where each new release builds on the capabilities of the previous one. In a typical generational lifespan of 10 years, 3GPP 5G standards are approaching the mid-cycle phase. Release 16 specifications were frozen in March 2020, and Release 17 work is well under way toward completion at the end of 2021, as shown in Fig. 1.

Overview of Key Areas of 5G Evolution

Release 15 provides an outstanding performance in KPIs essential for an extreme mobile broadband service; data rate, spectral efficiency, and latency all surpass what 4G is capable of. Releases 16 and 17 are focusing on expanding the ecosystem to allow the widest possible array of industries to take advantage of 5G. They contain a wide variety of features in all key areas of 5G evolution:

1. Industrial IoT
2. Vertical industries
3. Network deployment and automation
4. Device evolution

Figure 2 presents an overview of these key areas.

Industrial IoT (IIoT)

The goal of creating full support for factory automation and other critical IIoT applications requires effort spanning multiple 3GPP releases. In some markets, the term "Industrial 5G" is used to summarize the Rel-16 and Rel-17 features that form the support of IIoT. These are:

Ultra-Reliable Low-Latency Communications (URLLC): Release 16 completes the URLLC functionality that is critical for IIoT scenarios, such as factory automation. The foundation of URLLC was already laid in Release 15 with support for ultra-low latency by means of the "mini-slot," and Rel-16 adds the ability to achieve unprecedented levels of reliability, down to packet error rates of 10^{-6}, otherwise known as "six nines" reliability, as required by smart factory assembly and control operations.

Time-Sensitive Communication (TSC): TSC is a communication service that supports deterministic and/or isochronous communication with high reliability and availability. It provides packet transport with Quality of Service (QoS) characteristics such as bounded latency and reliability, where end systems and relay/transmit nodes can be strictly synchronized. Rel-16 introduces TSC support based on integration with IEEE Time-Sensitive Networking (TSN), which is the main protocol for factories today. Release 17 takes this to the next level by enabling TSC without reliance on IEEE TSN, thus offering enhanced enablers for the 5G system to support TSC natively for any application and deployment (e.g., music festivals). This will allow deployment of private wider area networks (WANs). Use cases for this type of network include logistics, shipping harbors and program-making and special events (PMSE).

Private Networks: Release 16 includes enhancements for private network deployments that are fully isolated to allow only authorized subscribers to camp in the network (referred to as non-public networks (NPNs) in 3GPP parlance). Release 16 allows support for both an NPN-specific authentication mechanism for UEs without a universal subscriber identity module (USIM) and an Authentication and Key Agreement (AKA) mechanism for UEs with a USIM. In Release 17, private-network support is being further extended by introducing support for

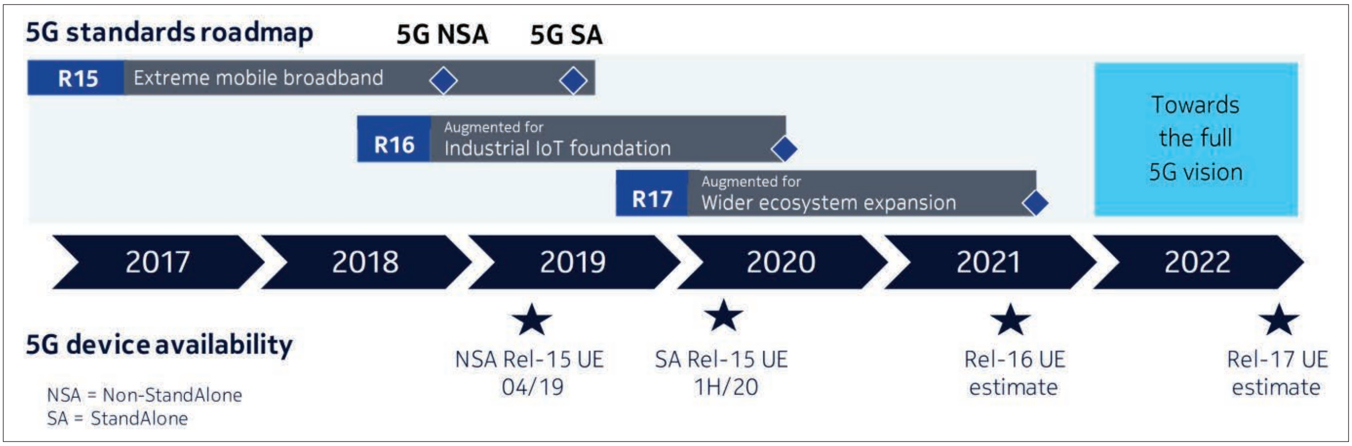

FIGURE 1. 5G standards evolution timeline in 3GPP.

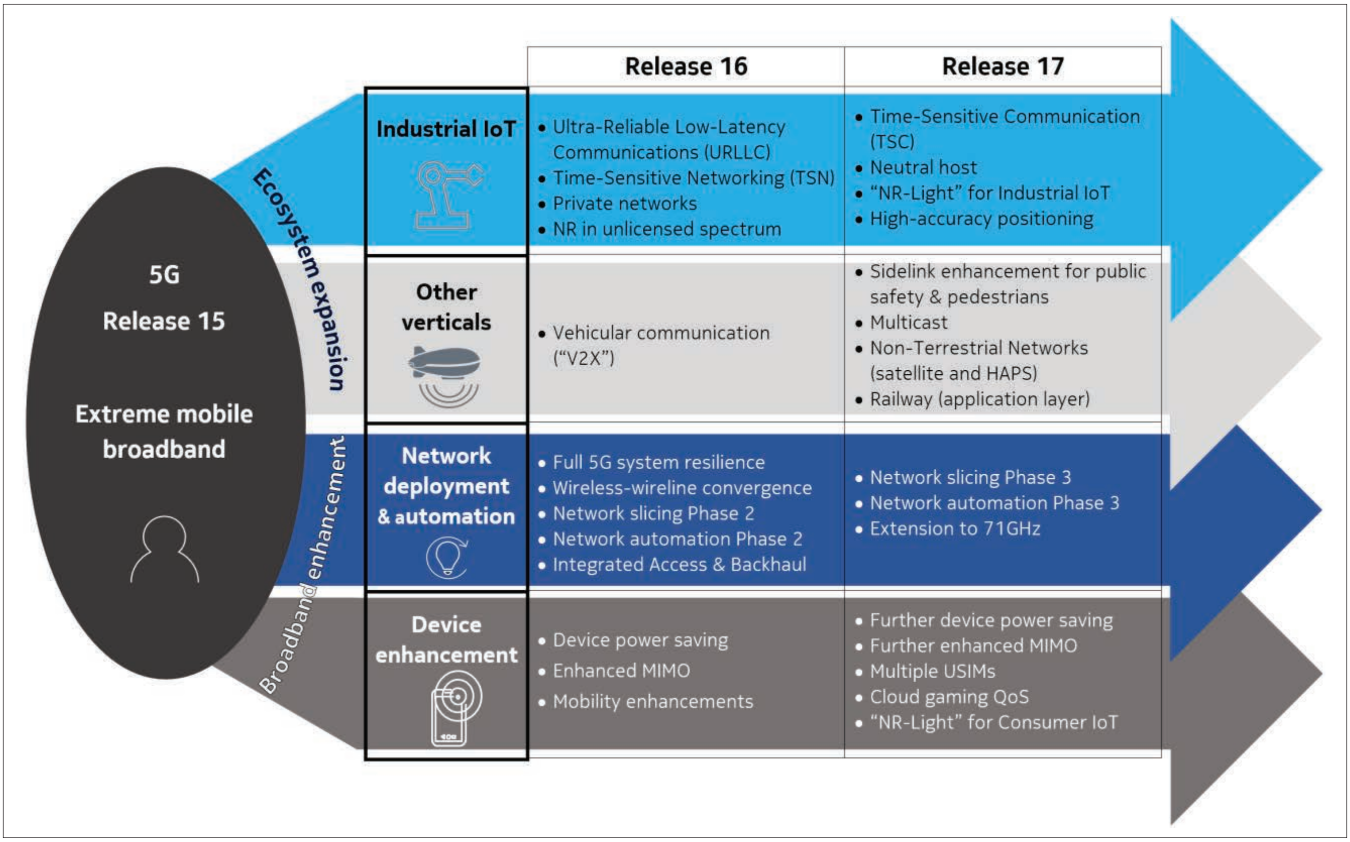

FIGURE 2. Key areas of 5G evolution in Releases 16 & 17.

neutral host models, where the network owner and service provider need not be the same entity. This includes enablers for accessing standalone private networks using credentials from third-party service providers, including public network operators. Furthermore, support for onboarding and provisioning of UEs to access private networks is being introduced.

NR-Unlicensed (NR-U): In order to facilitate deployment of 5G networks for IIoT and other similar enterprise scenarios, Release 16 introduces the features that are necessary for NR to be operated in unlicensed spectrum. Access to unlicensed spectrum provides an important tool to increase capacity for both service providers and private networks. For service providers, NR-U enables access to additional spectrum to improve the cellular network operation by offloading traffic in hotspots, while for private networks NR-U also enables the operation of standalone networks in unlicensed spectrum *without* any access to licensed spectrum. The initial focus of NR-U in Release 16 is on enhanced mobile broadband (eMBB) services in the 5GHz and 6GHz frequency bands. Release 16 NR-U uses the same flexible frame and slot structure and fundamental physical layer design and protocol stack as NR Release 15, hence limiting the magnitude of changes to the UEs compared to licensed band operation. NR-U adds channel access procedures to enable fair coexistence with other systems such as IEEE 802.11 variants or LTE Licensed-Assisted Access (LAA). In Release 17, enhancements for IIoT will include enabling URLLC features to be used in unlicensed spectrum, to better facilitate industrial use cases.

RedCap (Reduced Capability Devices): Release 17 brings a further step in the support of IoT by NR, in the form of supporting optimal operation of device types typical in IIoT services. Such devices will enable use cases like surveillance cameras and industrial sensors to use 5G, in addition to some wearables and consumer IoT devices. Compared to full 5G devices, RedCap devices are likely to be limited to narrower bandwidths, lower peak data rates and fewer antennas in order to reduce cost and complexity, while still enabling advanced connectivity. The characteristics required for these categories of devices, compared to full-blown 5G NR devices and low-power wide-area (LPWA) IoT devices, are illustrated schematically in Fig. 3.

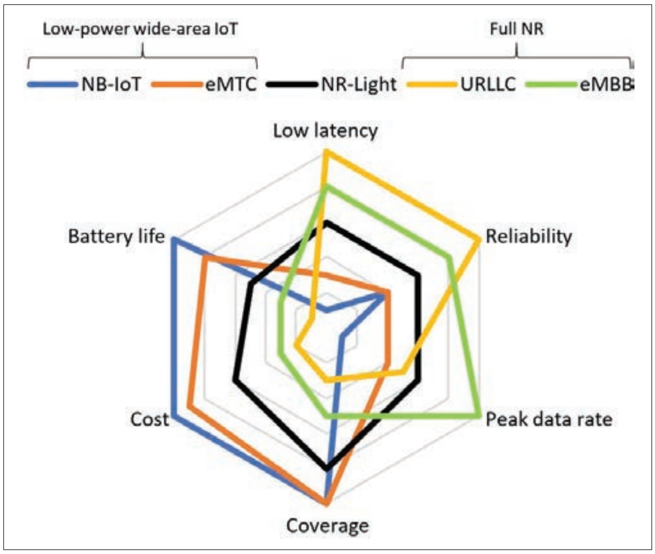

FIGURE 3. Expected capabilities of RedCap devices.

High-accuracy Positioning: Precise and up-to-date device location is an essential requirement for emergency calls, as well

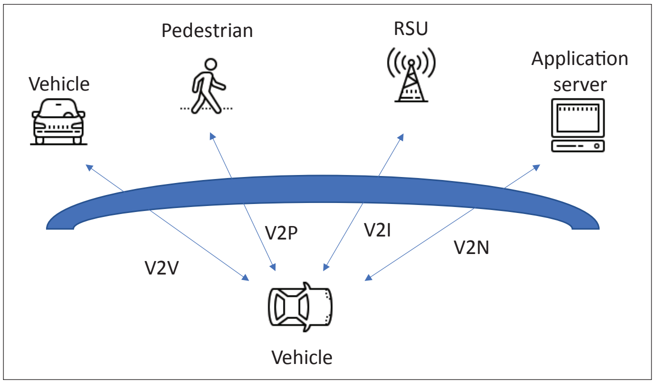

FIGURE 4. V2X communication modes.

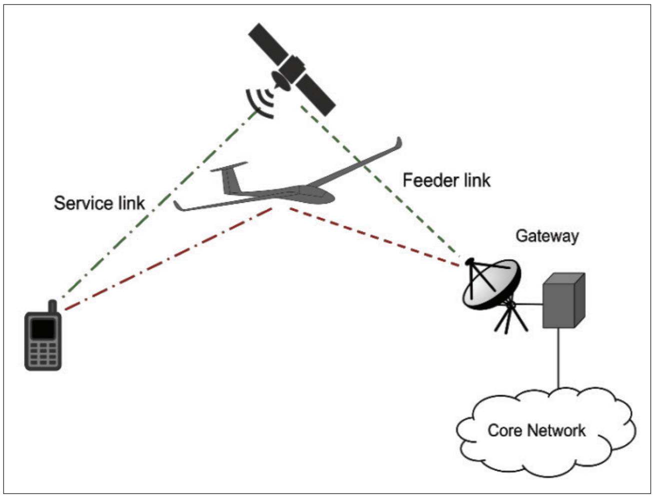

FIGURE 5. Non-terrestrial networking.

as new services like IIoT. Release 15 provides basic positioning protocol support. This is extended in Release 16 to include the capability to locate devices using the 5G radio signals themselves, leading to an expected accuracy in the order of a few metres for the vast majority of users (subject to the deployment scenario), thanks to 5G's wide transmission bandwidth and directional beamforming. In Release 17, further enhancements are being considered, aiming for accuracy levels down to a few tens of centimeters, as required by industrial process applications.

Vertical Industries

While IIoT is the most significant expansion of the 3GPP technology ecosystem addressed by Releases 16 and 17, these releases also introduce features that enable 5G to serve a variety of other dedicated markets where the role of the network operator may be broader than that of typical communication service providers (CSPs). These are loosely referred to as "vertical industries" for 3GPP technology, and the key examples considered are outlined below.

Device-to-device Use Cases, Including V2X[1] and Public Safety: Release 16 introduces a sidelink feature to enable direct communication between terminals. Unlike LTE, the 5G NR sidelink is specifically designed first for the requirements of cellular V2X (C-V2X), including vehicle-to-vehicle, vehicle-to-pedestrian and vehicle-to-roadside unit (RSU) communication, as illustrated in Fig. 4.

While C-V2X in LTE is primarily designed for broadcasting basic safety messages, 5G NR additionally supports more advanced use cases with lower latency, larger payloads and higher data rates, as well as both groupcast and unicast modes. These capabilities provide the foundations for operations such as platooning and remote driving.

It should be noted that wherever network coverage is available, V2X can be supported via the base stations using the URLLC functionality of NR together with edge computing to deliver low latencies. If the sidelink is being used, the physical layer resources can either be scheduled by the gNB[2] or autonomously selected by the UE in a contention-based manner.

In Release 17, enhancements to the sidelink and proximity services are being introduced to support *public safety* applications, for example where emergency workers can benefit from direct device-to-device communication. In particular, sidelink power-saving features are being introduced for handheld devices, which will also benefit pedestrian devices designed to communicate directly with vehicles for pedestrian safety.

[1] Vehicle-to-everything communication.

[2] 5G NR base station.

Multicast Communication: Release 17 is, for the first time in 5G, introducing support for multicast communication, mainly aimed at the use cases of public safety, V2X and railways. These cases require efficient communication to multiple terminals within an area of up to a few cells. The basis for this mode of communication will be Single-Cell Point-to-Multipoint (SC-PTM) transmission, where multiple UEs receive a single physical channel within one cell. This will be combined with a content distribution mechanism and architecture, enabling the same content to be multicast in multiple cells independently.

Non-terrestrial Networks (NTNs): NTNs, in which base stations are deployed on satellites or as High-Altitude Platform Stations (HAPSs), are valuable for supplementing the coverage of terrestrial networks and extending service provisioning to remote areas. The most interesting deployment modes for NTNs are expected to use HAPSs (e.g., balloons or unmanned aerial vehicles (UAVs)) or Low Earth Orbit (LEO) satellites. Release 17 is introducing the minimum set of changes needed to 5G NR to support the challenges of the longer propagation range and delay of non-terrestrial systems. It will be based on a transparent architecture for the non-terrestrial infrastructure, with the possibility of extension to regenerative architectures in a later release. The feeder link (aka backhaul link) terminates in a ground-based gateway that is connected to the 5G core network, as shown in Fig. 5.

Network Deployment and Automation

This 5G evolution area mainly comprises features that enable new deployment capabilities for CSPs, or which provide the means to reduce the operating expenses of running the 5G system.

Network Slicing (NS): Release 16 allows operators to outsource network slice subscription management to third parties who are using the operators' networks to provide services to their own customers. Rel-16 also introduces Network Slice-Specific Authentication and Authorization (NSSAA), wich allows a third party to manage a user's subscription to a particular slice without requiring the operator to be involved in managing transitions (adding/removing end users to/from a slice). For example, BMW, who is offering V2X services, can itself manage its end users (e.g., 5 Series and 7 Series car owners) who get access to its slice. Release 17 will further improve the overall operability and automation of network slice deployments.

Integrated Access & Backhaul (IAB) has been introduced in Release 16 as a key enabler for fast and cost-efficient deploy-

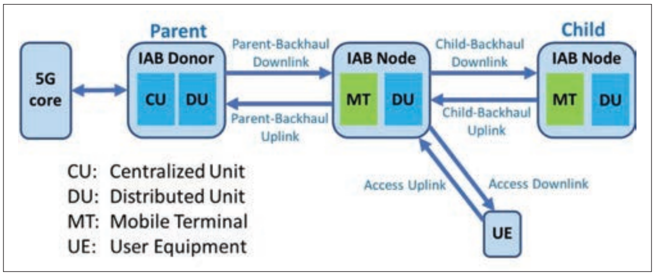

FIGURE 6. Integrated Access & Backhaul (IAB) architecture.

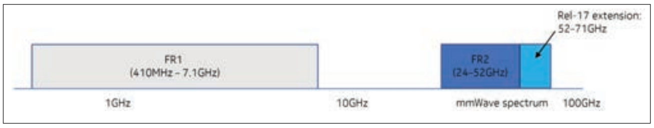

FIGURE 7. Spectrum for 5G.

ments, mainly targeting dense mmWave deployments outdoors. IAB nodes use the same spectrum and air interface for access and backhaul, creating a hierarchical wireless multi-hop network between sites. The hops eventually terminate at a donor node that is connected by means of a conventional fixed backhaul to the core network. IAB nodes may be deployed for four fundamental purposes:

- To remedy isolated coverage gaps
- To provide backhaul where fiber deployment is sparse
- To enhance system capacity
- To bridge coverage from outdoor to indoor

Figure 6 shows the architecture of IAB, with a central IAB node connected to the core network via parent links to a donor node, child links to a node downstream from the central IAB node, and access links to devices served directly by the central IAB node. The IAB architecture leverages the 5G NR gNB logical split architecture, with a Centralized Unit (CU) at the IAB donor node and Distributed Units (DUs) at IAB nodes. An IAB node contains a Mobile Terminal (MT) part that behaves like a UE toward the parent node. On the child links, the DU part of an IAB node behaves like a gNB toward the next-hop IAB node. On the access links, the IAB nodes behave exactly like normal gNBs, providing the NR radio interface for UEs in their coverage areas.

Enhancements to IAB in Rel-17 will improve the options for resource multiplexing between the parent and child links of an IAB node, as well as provide increased flexibility and robustness in the topology and traffic routing.

Extension to 71GHz: In Releases 15 and 16, 5G NR was designed to operate in both conventional cellular spectrum, known as "Frequency Range 1" (FR1), and mmWave spectrum up to 52.6GHz (FR2), as shown in Figure 7. As new spectrum becomes relevant, 3GPP continues to define new bands and band combinations in which 5G can be deployed, aiming to meet the demands of operators worldwide.

A major step in this domain in Release 17 is the introduction of support for 5G NR in the 52–71GHz frequency range, where up to 14GHz of spectrum is available, including both unlicensed spectrum for NR-U and the possibility of licensed spectrum in the 66–71GHz range following the identification of this band as IMT spectrum at WRC-19.

Developing NR to work above its current upper limit of 52GHz requires careful analysis of the transmitter and receiver impairments at these frequencies, as well as the power amplifier characteristics. Additionally, the narrower beams and higher attenuation at these frequencies will provide higher spatial isolation than is seen in lower-frequency bands, and this will need to be taken into account in defining coexistence requirements for unlicensed operation in this frequency range.

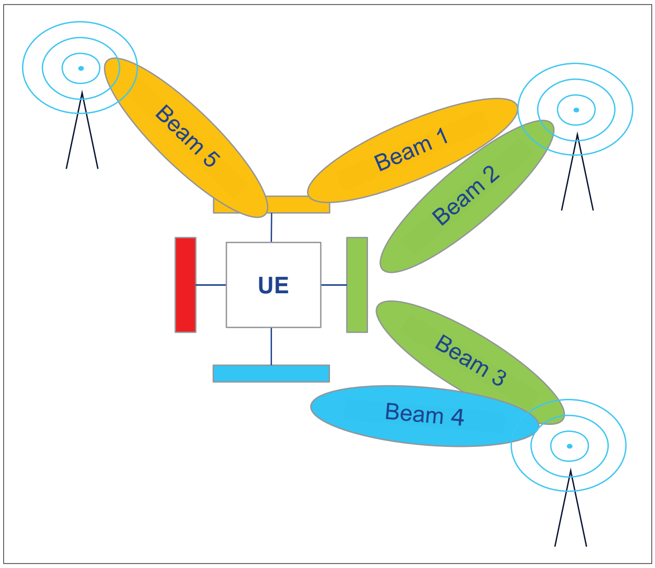

FIGURE 8. User Equipment (UE) with multiple antenna panels, communicating with multiple Transmission & Reception Points (TRPs).

Device Evolution

As 5G deployments become more widespread and mature, Releases 16 and 17 introduce enhancements to the fundamental mobile broadband experience provided by 5G. These include:

Device Power Saving: Device battery life is a key element of the user experience. While many aspects of battery life are dependent on implementation design and tend to improve organically as the technology matures, there are some aspects of the 5G standards that directly impact the energy efficiency of devices. Release 16 introduces signaling from the network to the devices, specifically to help them reduce their "connected mode" power consumption when conditions allow, enabling the device power consumption to be well optimized during periods of active transmission and reception. Release 17 enhancements will focus on power-saving enhancements for devices that are in "inactive" or "idle" modes, thus helping to improve standby battery life.

Enhanced MIMO: 5G has from the beginning provided extensive support for large-scale antenna arrays often referred to as "massive MIMO." The whole design of 5G NR is "beam-based," which means that all the transmissions can be beamformed to enhance coverage and capacity. To support massive MIMO beam management, it is necessary for each transmitter to have accurate information about the characteristics of the radio channel. Signaling of this information incurs overhead, and Release 16 introduces new compression techniques to reduce this overhead when beams are serving multiple users with high data rates simultaneously, thus improving the overall efficiency of massive MIMO operation.

Releases 16 and 17 further introduce the possibility of transmission and reception at multiple Transmission & Reception Points (TRPs) on the network side, leading to a kind of Coordinated Multi-Point (CoMP) operation that is especially relevant for ensuring reliability for URLLC services. Release 17 will also enhance support for devices with multiple antenna panels (as shown in Figure 8), which are particularly effective for increasing data rates when operating at mmWave frequencies.

5G: Beyond the Horizon

The 10-Year Cycle of Gs

The communications industry is characterized by generation cycles of close to 10 years. Each new generation is aligned with global targets set by ITU and unleashes new performance levels, not least by getting out of the constraints of backward compatibility between networks and devices within a generation. New air interfaces such as 3G, 4G/LTE and 5G all reflect the progress in both enabling technologies like compute power and regulation like the assignment of millimeter wave spectrum in 5G. Experience has shown that not all initial ideas for the new generation could be introduced in one go; the large feature hub enabled by 3GPP Release 15 and 16 is a testament to the fact that a phased approach of introducing 5G was needed.

On the other hand, the innovation speed in consumer electronics, webscale businesses and digitalization of industries is strongly tied to building up a critical mass of ecosystem including device availability momentum. Think of the rise and spread of 4k video as an example, where the ecosystem has been forming along professionally produced content and its distribution, TV screens, professional cameras and smartphone cameras. In media formats, large sports events such as the Olympics are catalysts to cross the chasm from early adopters to mass market momentum. Technology breakthroughs in fields such as artificial intelligence give sudden rise to new automation fields. While all these innovations are increasingly interweaved with cloud infrastructure, connectivity fabrics and hence mobile networks, their rise will follow individual timelines.

The next five years will see 5G networks enable use cases beyond mobile broadband to increase, e.g., industrial productivity. This relates to 3GPP features provided with Releases 16 and 17. Toward the end of the decade we will see the initial 6G networks, on which basic research has already started. In this paper we aim to shed some light on what might shape advanced communication networks from 2025 onwards and what will be addressed by subsequent 3GPP releases such as 18 and 19.

Blueprint for Further 5G Evolution

Ubiquitous Coverage: Both IoT and Extreme Mobile Broadband use cases will continue their expansion from population to geographic and volumetric coverage. Use case domains include both connectivity for humans and machines: higher data rates for IoT to support video as a sensor, basic MBB services using fixed wireless access to reach more people. NR Light will add efficiencies for services up to 100Mbps throughput and down to 100ms delay requirements.

Fusion of New Service Profiles: Use cases demanding reliable high data rate services with strictly bounded latency are fast growing. A fusion of eMBB and URLLC characteristics could enable new service profiles.

Time and Space Accuracy: Providing high precision time information through the network without the need for a GNSS satellite link can make applications more robust and less complex. Applications include digital financial transfers, multimedia synchronization, synchronization of industrial production units, and real-time digital twin solutions.

The Next Level of Automation: For the 5G network platform itself, one can foresee AI/ML to modernize and optimize almost any functionality in the 5G network, ranging from linearization of RF power amplifiers to complete Network Automation and optimization of the radio network aligned with open RAN.

Acknowledgements

I would like to express my appreciation to several of my coleagues for their input and hard work in creating this article: Miikka Poikselkka, Matthew Baker, Harri Holma, and Peter Merz.