

5G beyond radio access: a flatter sliced network

David Soldani

Sommario

Questo articolo descrive i principali scenari di utilizzo e i relativi requisiti tecnici per il 5G, presentando le diverse opzioni architetturali, tecnologie abilitanti, le bande spettrali definite dal 3GPP e le relative roadmap. Vengono affrontati sia gli aspetti radio (copertura, capacità, latenza e affidabilità) sia quelli della core network (slicing, architetture, protocolli). Viene illustrato in dettaglio la differenza tra LTE e 5G, sia a livello radio che a livello rete core. Inoltre viene sottolineato come con il 5G il Total Cost of Ownership (TCO) scenda sostanzialmente e come si creino nuove opportunità di ritorni, specialmente nel settore business to business (B2B), in particolare con offerte di Network Slice come Servizio (NSaaS) a prezzi competitivi per una varietà di settori verticali.

Abstract

This article defines what 5G is, the main usage scenarios and related technical requirements. It also presents the different architecture options, key enabling technologies and spectrum bands specified by 3GPP, and roadmap for the different 3GPP releases. It then introduces the main technology components, end to end, and key functionalities of New Radio (NR), Next Generation Radio Access Network (NG-RAN) and 5G Core (5GC). The NG-RAN performances, in terms of coverage, capacity and latency and reliability with respect to LTE (3GPP Release 14 and previous releases), are also discussed. This is followed by a deep dive into 5G slicing concepts, architectures, protocols and procedures for the composition, embedding, commissioning, setup and teardown of network slice instances, from terminals to external data networks. As a part of this framework, the new aspects of QoS management at flow level, end to end, are also described. The 3GPP Phase 1 (Release 15)

specifications for enhanced Mobile Broadband (eMBB) are completed in the second half of 2018. The descriptions of 3GPP Phase 2 (Release 16) work items, with more functionalities for Ultra-Reliable Low Latency Communications (URLLC) and Massive Machine Type Communications (mMTC) and additional architecture options, are made available in the first half of 2020. Other 3GPP releases with further enhancements and support of frequency bands above 40 GHz will follow. 5G systems are expected to be deployed at 3.5 GHz reusing the same sites (radio grid) as LTE at 1800 MHz. In 3GPP R15, 5G NR supports band from 400MHz to 40GHz, provides 10-20 times better cell capacity than LTE, and latency below 10 ms, being 1ms the target of NR, depending on the protocol layer split between Radio Unit (RU) and Radio Control Unit, i.e. the deployed fronthaul interface. Service oriented architecture, Shared Data Layer (SDL), and terminal assisted network slicing are also supported with the introduction of 5GC in a standalone (SA) configuration. 5G slicing enables new information and role models, and network multitenancy using a single physical infrastructure. With 5G the total cost of ownership (TCO) plunges substantially and new sources of revenues, especially for business to business (B2B), surge significantly, as communication service providers have an easy way to offer Networks Slice as a Service (NSaaS) at competitive price to their tenants from a variety of vertical sectors.

Keywords: 5G, Spectrum, SDN, NFV, NSaaS, Slicing, Computing, mMIMO, Management and Orchestration, Multi-connectivity, New Radio, 5GC, NG-RAN

1. What is 5G?

5G defined as the International Mobile Telecommunications (IMT) for 2020 and beyond (5G) will expand and support diverse usage scenarios and applications with respect to current mobile network generations, purposed primarily for voice, mobile internet and video experience.

The agreed scenarios for 5G include: 1) “Enhanced mobile broadband (eMBB)” addressing human-centric use cases for access to multimedia content, services and data; 2) “Ultra-reliable-low latency communications (URLLC)” with strict requirements, especially in terms of latency and reliability; and 3) “Massive machine type communications (mMTC)” for a very large number of connected devices and typically transmitting a relatively low volume of non-delay-sensitive information [1].

5G technologies will efficiently enable new secure, dependable, ultra-reliable, and delay-critical services to everyone and everything, such as cognitive objects and cyber physical systems. To realize this vision, 5G capabilities will include [2]: A new flexible and efficient wireless interface, access schemes, and

other enabling wireless and network technologies, as well as a new plastic network architecture, supporting multi-tenant and new role models. With network slicing, different end-to-end logical networks with isolated properties are provided and operated independently [3]. These enable operators to support different use cases, with devices able to connect to multiple slices simultaneously, and monetize network slice instances as a service [4].

Two architectures options are possible [5], [6]: The radio access will be based on New Radio (NR) in a *Stand-Alone* (SA) configuration of 5G systems - Devices, Next Generation (NG) Radio Access Network (NR), NG Core (or 5GC) - or Non-Standalone (NSA) NR in an Evolved Packet System (EPS), which is a dual connectivity 5G System deployment with an Evolved-UTRA, as the anchor Radio Access Technology (RAT), and NR as the secondary RAT, in a non-standalone configuration in Evolved Packet System. The different architecture options are illustrated in Figure 1.

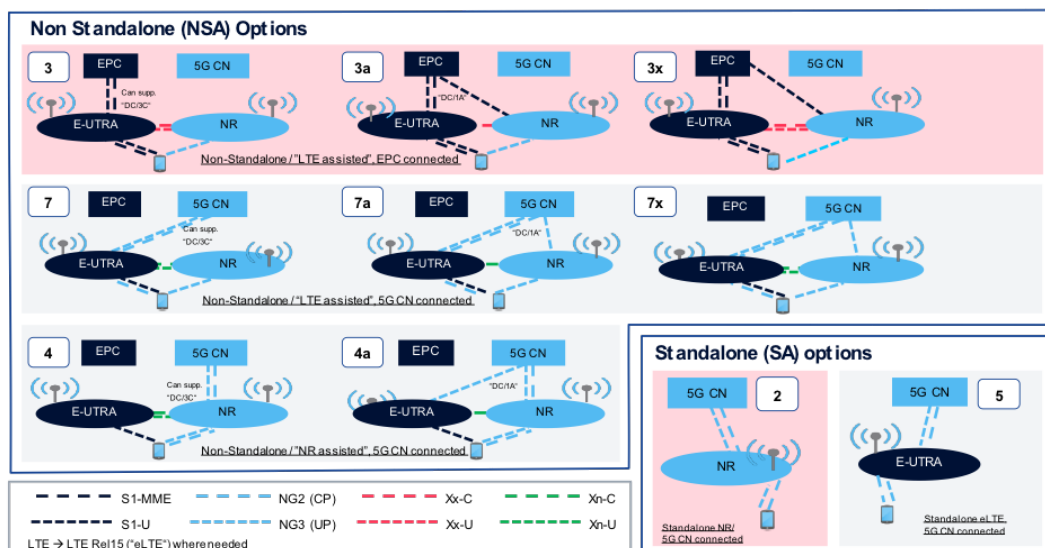


Figure 1
5G Network architecture options: NSA and SA. (© Nokia 2018)

A high-level view of the 3GPP roadmap is depicted in Figure 2. 3GPP has defined two phases for the 5G system standardization: 5G Phase1, or 3GPP Release 15 (R-15), will be finalized during 2Q 2018; the 5G Phase 2, or 3GPP Release 16 (R-16), will be completed in 1Q 2020. In 3GPP R-15, the 5G Mobile System will support eMBB and some aspects of URLLC usage scenarios. In 3GPP R-16, the 5G NR will be enhanced with additional features for eMBB, URLLC and mMTC, currently left for further studies. As of today, in 3GPP, there is a strong industry interest in completing the Non-Standalone version of New Radio specifications, Option-3 family, by Dec 2017-Mar 2018, and in finalizing the Standalone 5G architecture, Option-2, by June-September 2018. All other architectures will be specified at later time.

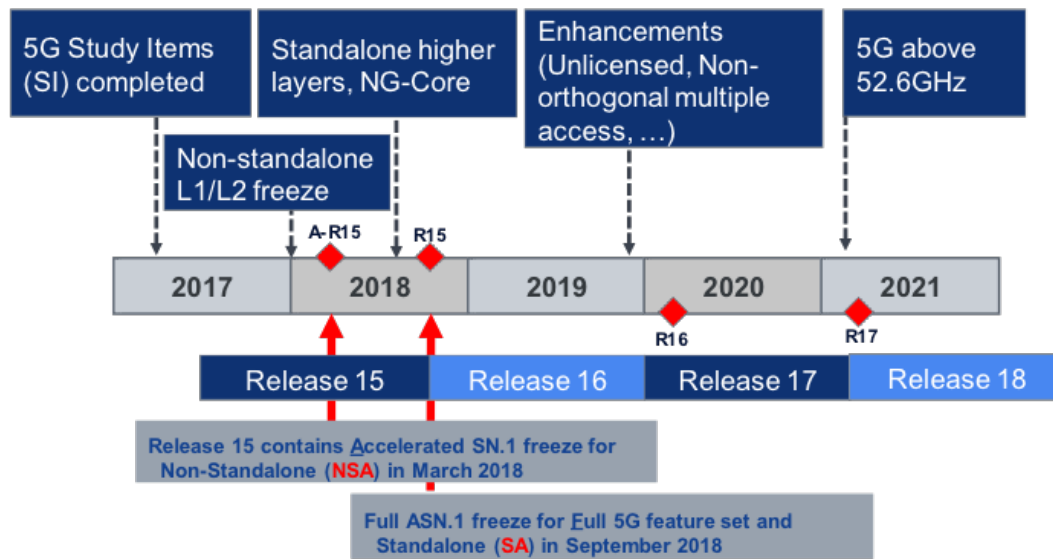


Figure 2
3GPP 5G releases timeline till 2022. (© Nokia 2018)

The main new technology components of 5G are shown in Figure 3: New spectrum; Massive MIMO (mMIMO) beamforming; Network slicing; Multi-Connectivity; and cloud and edge computing [2], [7].

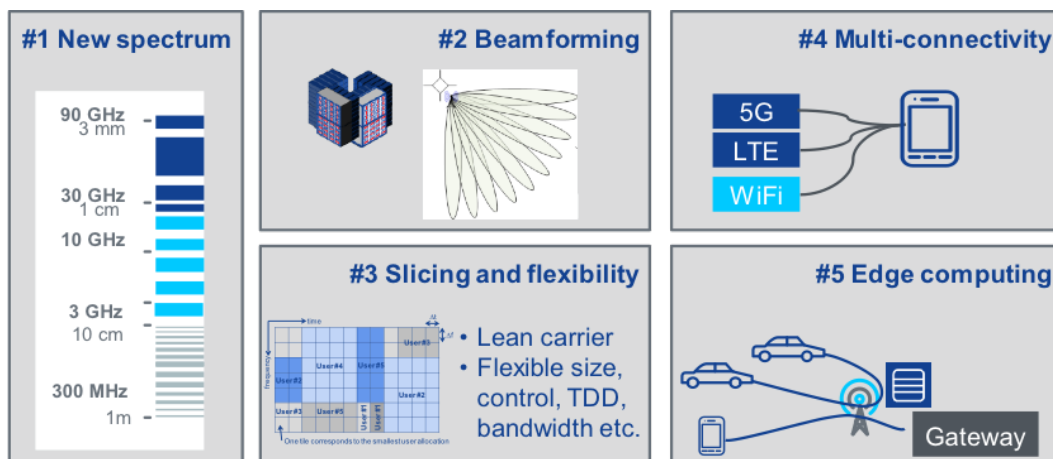


Figure 3
The key 5G enabling technologies [2]. (© Nokia 2018)

1.1 Wireless technologies

5G is expected to increase spectrum efficiency and support contiguous, non-contiguous, and much broader channel bandwidths than available to current mobile systems. 5G radio will be the most flexible way to benefit from all available spectrum options from 400 MHz to 90 GHz, including licensed, shared access and license exempt bands, FDD and TDD modes, including Supplementary Uplink (SUL), as well as narrowband and wideband Carrier Components (CC). The recently approved 3GPP New NR bands (n) in Frequency Range 1 (below 6GHz) and Range 2 (mmWave) are depicted in Figure 4. The LTE re-farming bands will use LTE numbers with n prefix. The 37-40GHz band will be also to defined by the 3GPP.

	Band number	UL	DL	Duplex mode
Range 1	n77	3.3 – 4.2 GHz	3.3 – 4.2 GHz	TDD
	n78	3.3 – 3.8 GHz	3.3 – 3.8 GHz	TDD
	n79	4.4 – 5.0 GHz	4.4 – 5.0 GHz	TDD
	n80	1710 – 1785 MHz	N/A	SUL
	n81	880 – 915 MHz	N/A	SUL
	n82	832 – 862 MHz	N/A	SUL
	n83	703 – 748 MHz	N/A	SUL
	n84	1920 – 1980 MHz	N/A	SUL
Range 2	n257	26.5 – 29.5 GHz	26.5 – 29.5 GHz	TDD
	n258	24.25 – 27.5 GHz	24.25 – 27.5 GHz	TDD
	n259	31.8 – 33.4 GHz	31.8 – 33.4 GHz	TDD

Figure 4
Examples of 3GPP 5G approved bands

Millimeter wave spectrum above 20 GHz can provide bandwidth up to 1-2 GHz, which offers data rates up to 20 Gb/s and extreme mobile broadband capacity. The high bands are mostly suitable for local usage, such as mass events, indoor and outdoor hotspots and for Fixed Wireless Access (FWA).

Spectrum at 3.5 GHz, 4.5 GHz and 4.9 GHz will be used for 5G coverage and capacity in urban areas reusing the existing sites. At those frequencies, the bandwidth can be up to 100 MHz per operator, and even up to 200 MHz with the re-farming of some of the existing bands. As shown in Figure 5, the 5G coverage at 3.5 GHz, when using mMIMO beamforming, can be like LTE at 1800 MHz. The expected uplink gap (≈ 3 dB) in link budget can be compensated by sharing the LTE band with NR (NR-LTE multiplexed in frequency domain in uplink FDD mode, and 5G at 3.5 GHz in downlink TDD model; or NR-LTE overlapping transmission in uplink using the LTE band and TDM between LTE and NR).

Low bands, below 1 GHz, are needed for wide area rural coverage, for ultra-high reliability and for deep indoor penetration. Extensive coverage is important for new use cases such as IoT and critical communication. The low band could be 700 MHz, which is available in many countries, or 900 MHz, today mostly used by 2G and 3G systems. In the USA, another option for 5G low band is 600 MHz [2].

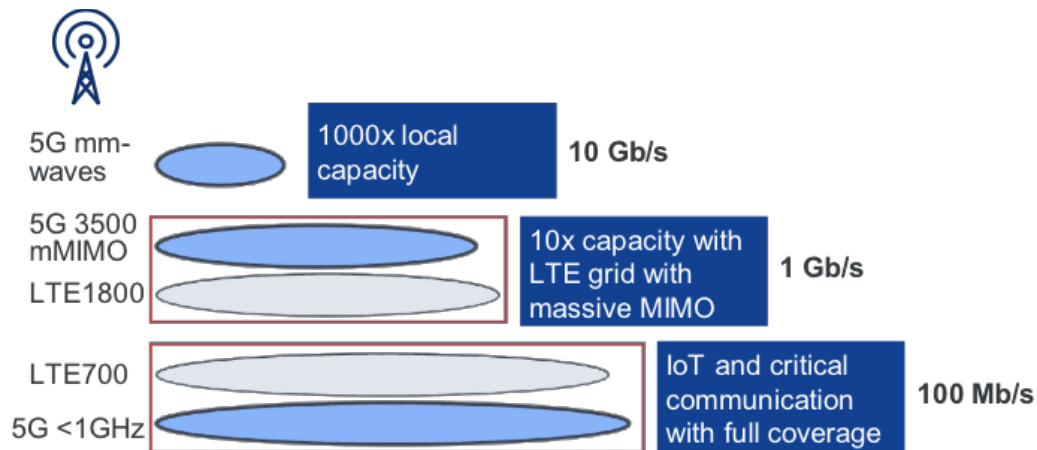


Figure 5

5G Coverage footprint: Combination of low and high bands. (© Nokia 2018)

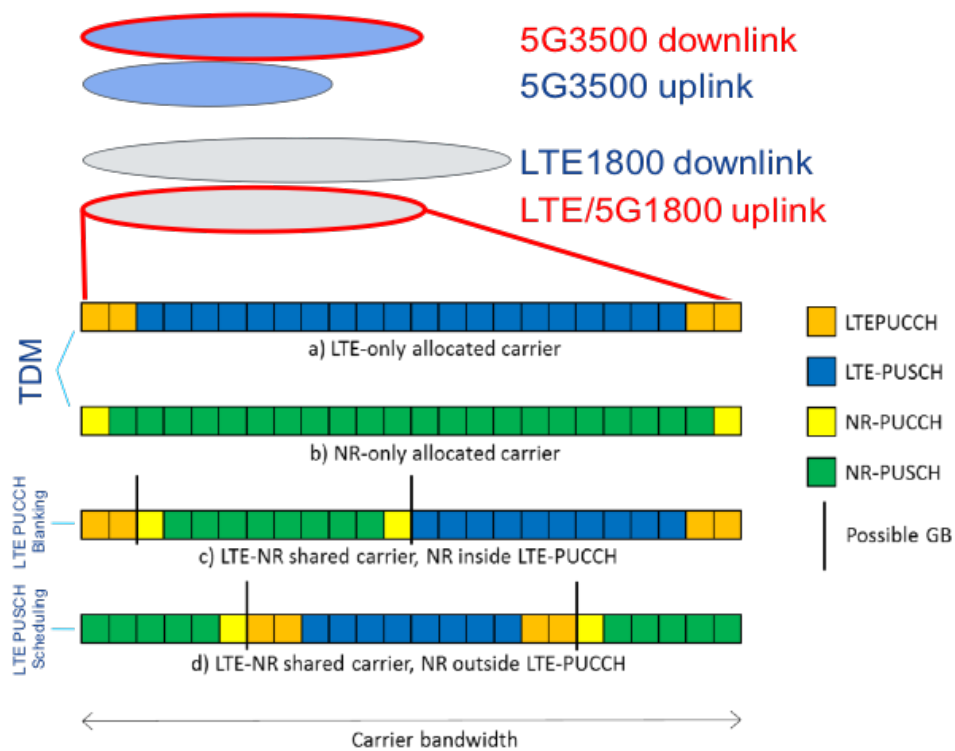


Figure 6

Low band FDD for coverage (uplink sharing) and 3.5 GHz TDD for capacity. (© Nokia 2018)

The new 5G radio is for all spectrum options. To this end, 5G supports a *flexible numerology*, which consists of different *Sub Carrier Spacing* (SCS), *nominal Cyclic Prefix* (CP), and *Transmission Time Interval* (TTI), or *scheduling interval*, depending on bandwidth and latency requirements. An overview of the 5G numerology is shown in Figure 7. Sub-carrier spacings of 15 kHz to 120 kHz, and the corresponding cyclic prefix of 4.7 to 0.6 μ s and scheduling interval of 1 ms to 0.125 ms, are defined for different *Carrier Components* (CC), which may vary from 5 MHz to 100MHz, below 6GHz, and from 50 to 400MHz, above 6GHz [7].

Numerologies with normal CP (subframe = 1ms)

Parameter	Slot Configuration 0						Slot Configuration 1		
Subcarrier spacing [kHz]	15	30	60*	120*	240**	480	15	30	60
OFDM symbol duration [us]	66.7	33.3	16.6	8.33	4.17	2.08	66.7	33.3	16.6
Nominal CP [us]	4.7	2.41	1.205	0.60	0.30	0.15	4.7	2.41	1.205
Nominal max BW [MHz]	49.5	99	198	396	397.4	397.4	49.5	99	198
Max FFT size	4096	4096	4096	4096	2048	1024	4096	4096	4096
Min scheduling interval [symbols]	14	14	14	14	14	14	7	7	7
Min scheduling interval [slots]	1	1	1	1	1	1	1	1	1
Min scheduling interval [ms]	1.0	0.5	0.25	0.125	0.0625	0.0312	0.5	0.25	0.125

* 1st Product; ** Synchronization Signal (SS) Block

a)

Numerologies with extended CP (subframe = 1ms)

Subcarrier spacing [kHz]	Symbol Duration [us]	Ext CP [us]	Nom max BW	FFT Size	Scheduling Interval [sym]	Sched Interval [slot]	Scheduling Interval [ms]
60	16.6	4.2	198	4096	12	1	0.25

b)

Figure 7

5G numerology overview with 1ms subframe: a) Normal CP; b) Extended CP.
(© Nokia 2018)

For optimal radio performance, the higher the carrier frequency, the higher the allowed carrier component and subcarrier spacing, the lower the corresponding cyclic prefix and scheduling period. If we consider a typical 5G deployment at the 3.5 GHz band, the bandwidth could be 40-100 MHz, the subcarrier spacing 30-60 kHz and minimum scheduling period 0.125 ms. The corresponding numbers in LTE are 20 MHz bandwidth, 15 kHz subcarrier spacing and 1 ms scheduling period. If the slot length is more than 0.125 ms in the narrowband cases and low latency is required, then so called 'mini-slot' can be used in 5G, where the transmission time may be reduced to 2, 4 or 7 OFDM symbols. It is also possible to combine multiple slots together.

Massive MIMO (mMIMO) will be an integral part of 5G from day one, including common and control channels with beam forming and tracking. With mMIMO the *number of transmitting antenna elements is much higher than the number of MIMO streams* (layers). In practice, mMIMO means that the number of controllable antenna elements is more than eight.

5G radio supports 8 Layer Single User (SU)-MIMO or 16 Layer Multi User (MU)-MIMO in the downlink, and 4 Layer SU-MIMO in the uplink, with the possibility of dynamic

switching in both directions. MU-MIMO means that parallel MIMO data streams (or layers) are transmitted to different users at the same time-frequency resources. A typical example of 16 Layers configuration is 2layer/UEx8UE MU-MIMO.

Beamforming (BF) offers the advantages that the same resources can be reused for multiple users in a cell: It allows Space Division Multiple Access (SDMA), maximizing the number of supported users within that sector: It minimizes interference and increases cell capacity. As a result, Massive MIMO and *Active Antenna Technologies* (AAT) give higher spectral efficiency and provide much more capacity on existing base station sites. The technology can also enhance link performance and increase the coverage area. 5G radio design is fully optimized for massive MIMO using three basic techniques for forming and steering beams [8], as illustrated in Figure 8:

- *Digital beamforming*, where each antenna element has a transceiver unit with the adaptive Tx/Rx weights in the baseband, enabling frequency selective beamforming. Digital beamforming boosts capacity and flexibility and it is mostly suited to bands below 6 GHz.
- *Analog Beamforming* implements only one transceiver unit and one RF beam per polarization. Adaptive Tx/Rx weighting on the RF is used to form a beam. This is best suited for coverage at higher mmWave bands and offers low cost and complexity.
- *Hybrid beamforming* is a combination of analog and digital beamforming. When some beamforming is in the analog domain, the number of transceivers is typically much lower than the number of physical antennas, which can simplify implementation, particularly at high frequency bands. This technique is suited to bands above 6 GHz.

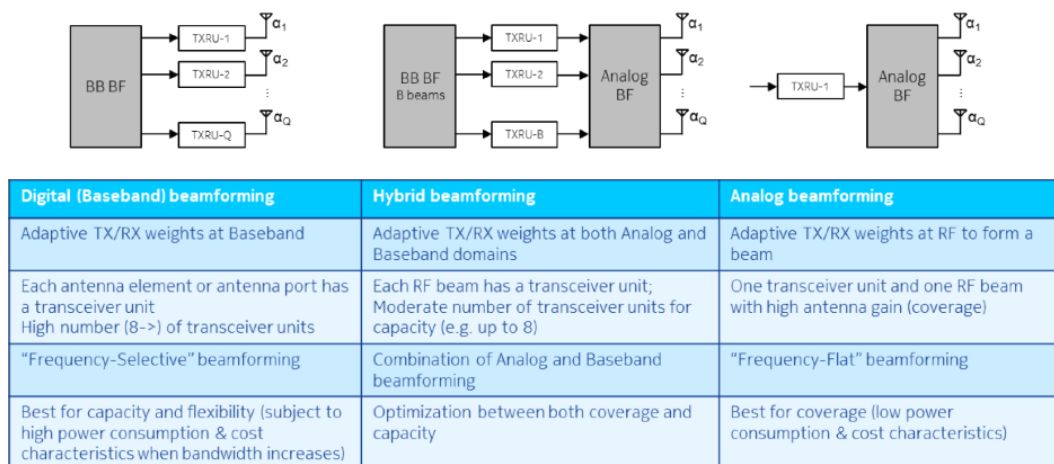


Figure 8

Digital, Analog, and Hybrid beamforming: Advantages and drawbacks. (© Nokia 2018)

The radio interface protocol stack is also optimized for cloud and distributed computing, with a flexible front-haul split, between high-layer protocols, i.e.

PDCP and RLC, and low layer (L1) peer entities, using the Enhanced Common Public Radio Interface (eCPRI). This will allow access providers to relax the transmission capacity and the utilization of Ethernet. For example, see Figure 9, assuming 100 MHz band, a 3-sector site with 64TX/RX mMIMO and 16 layers, the interface between the radio unit (RU) and edge cloud (radio access unit, RAU) requires: 1Tb/s with no split, using a CPRI interface; 150 Gb/s with low layer split and enhanced CPRI (eCPRI) interface; and 1-10 Gb/s with high layer split. The latency with low layer split is expected to be below 0.1 ms, as with CPRI; and 5 ms with high layer split, which is still satisfactory for most of the ultra-reliable low latency communications.

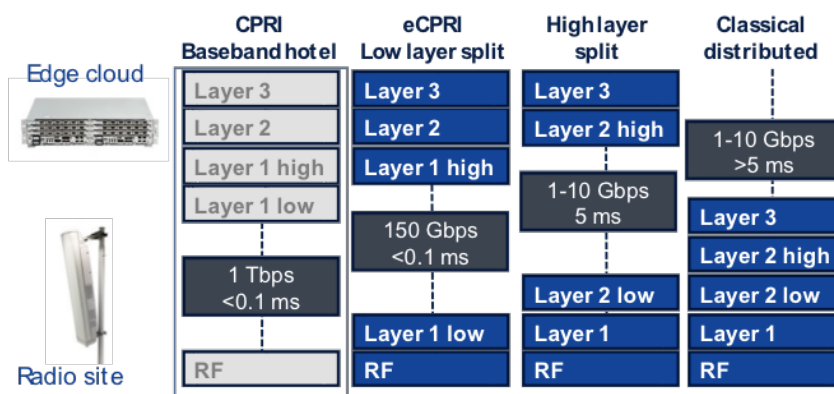


Figure 9

5G radio architecture options and performance for 100 MHz, 3-sector, 64TX/RX mMIMO, 16 Layers. (© Nokia 2018)

Compared to LTE1800 with 2x2 MIMO, as baseline, 5G gives 10-20x additional capacity, being 2-4 more spectrally efficient. For example, see Figure 10, LTE with 20MHz band achieves 40 Mb/s cell throughput, which corresponds to a spectrum efficiency of 2b/s/Hz; 5G with mMIMO beamforming at 3.5 GHz and 100MHz band reaches 400-800 Mb/s cell throughput, corresponding to a spectrum efficiency of 4-8 b/s/Hz.

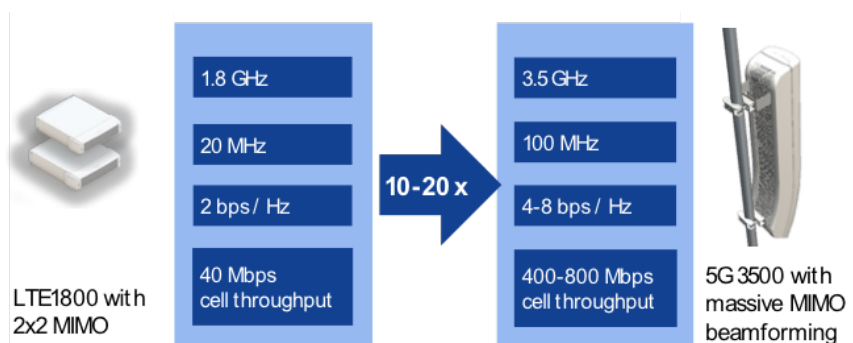


Figure 10

Example of 5G capacity gain with respect to LTE. (© Nokia 2018)

The radio interface latency with 5G is also dramatically reduced compared to previous releases. The target is 1 ms in idle and connected mode, with and without uplink allocated resources. 5G solutions to low latency are: Shorter transmission time interval, connected inactive state and contention based uplink.

1.2 Network technologies

5G is not only about radio! Additional benefits will come with the introduction of the 5G core (5GC), which supports many new enabling network technologies. Among other fundamental technology components, as depicted in Figure 11, the 5GC is characterized by a *layered and service oriented architecture*, with control plane (CP) and user plane (UP) split and *shared data layer* (SDL), for subscription, state and policy data [9]. It also supports [6]: *User plane session continuity*, while the terminal moves across different access points; *interworking with untrusted non-3GPP access*; a comprehensive *policy framework* for access traffic steering, switching and splitting; and *wireless-wireline convergence*.

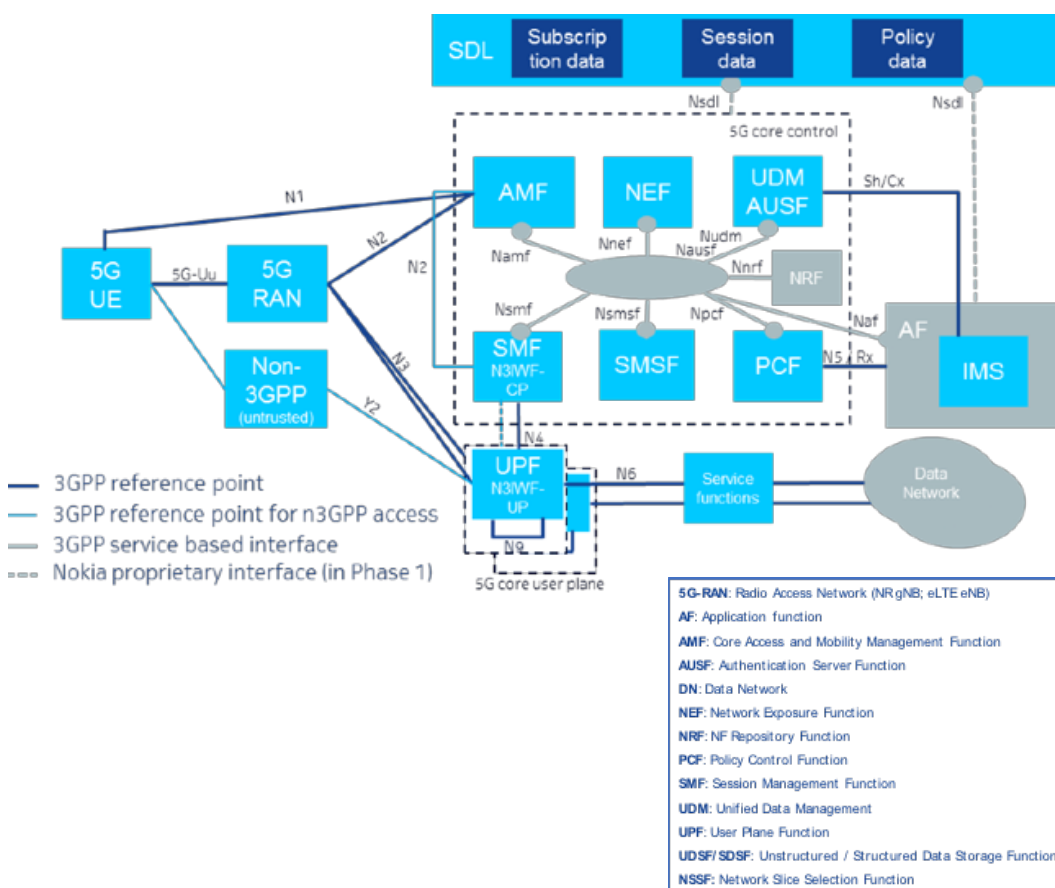


Figure 11
5G Core (5GC) overview [5]. (© Nokia 2018)

The separation of control and user planes provides deployment flexibility and independence. The distribution of core functionality, especially user plane, closer to the radio, i.e. to the edge cloud, enables the placement of applications closer to the end-user with improved throughput, reducing the transport network load, and latency, decreasing the physical distance and number of transport hops. The service based architecture – Including the related Network Repository Function (NRF) for 5GC control plane functions – allows flexible addition and extension of functions including the proprietary ones. The SDN to all network functions (and micro-services within the functions) enables common session resiliency and geo-redundancy models, and a shared interface for accessing all session data (e.g. for analytics). The 5GC is also agile to deploy new or updated functions in a “DevOps” style. The Network Exposure Function (NEF) enables the integration of telco functions with service providers, enterprises and operator’s own IT systems to create new use cases and opportunities to monetize. The core slicing and related Network Slice Selection Function (NSSF) enable a flexible assignment of users to different network slice instances that are tailored to different use cases. The 5GC also supports a unified subscriber management, authorization and authentication solution [9].

Other fundamental 5G enabling technologies, end to end, are: *Flow based QoS*, with a much higher level of granularity than LTE, which is currently limited to the bearer service concept; *multi-connectivity*, where the 5G device can be connected simultaneously to 5G, LTE, and WiFi, offering a higher user data rate and a more reliable connection; *terminal assisted Network Slicing*, and *E2E network management and orchestration*, with in-built support for cloud implementation and edge computing [6].

The 5G flow based QoS concept is illustrated in Figure 12. The Non-Access Stratum (NAS) packet filters associate IP Flows with QoS Flows in the UE (5G terminal) and User Plane Function (UPF). The *QoS flow identifier* (QFI) uniquely identifies the QoS flow and consists of the following parameters [6]:

- *5G QoS Identifier* (5QI): *Resource Type* (GBR/N-GBR), *Priority Level*, *Packet Delay Budget* (PDB) UE ↔ User Plane Function (UPF), and *Packet Error Rate* (PER);
- Allocation Retention Priority (ARP);
- *Guaranteed* (GFBR) and *Max Flow Bitrate* (MFBR) for GBR flows (UL and DL).

In the downlink, each UPF uses policies from PCF/SMF to identify flows and adds QFI tags, enforces Session-AMBR (Aggregated) and counts packets for Charging. Each UPF sets Reflective QoS (RQI) to activate Reflective QoS in the UE. The NG-RAN uses QFI tags and policies to map flows to one or more Data Radio Bearers (DRBs), and enforces the Max Bit Rate (UE-AMBR) limit per UE for non-GBR QoS flows.

In the uplink, the UE uses either signaling or “reflective” learning approach to derive the QFI policies to map QoS flows onto DRBs, and performs UL rate limitation on a PDU Session basis for non-GBR traffic, based on Session-AMBR value. The NG-RAN and UPF police QFI usage and enforces Max Bit Rate (UE-

AMBR) and Session-AMBR, respectively. The UPF counts the packets for charging.

The transport network layer QoS treats packets based on 5QI tags (e.g. using DiffServ). The mapping is based on Software Defined Networking (SDN) policies.

IP Flows are mapped onto QoS Flows, which are mapped onto one or more data radio bearers (DRBs). DRBs are associated to one PDU Session, which is mapped onto one S-NSSAI. The S-NSSAI is mapped onto one Network Slice Instance (NSI), i.e. one Network Slice; and the NSI is mapped onto a single Data Network Name (DNN). However, it is not true the vice versa, as described in the following text. This is how 5G handles the 5G flow based QoS within a given NSI [6].

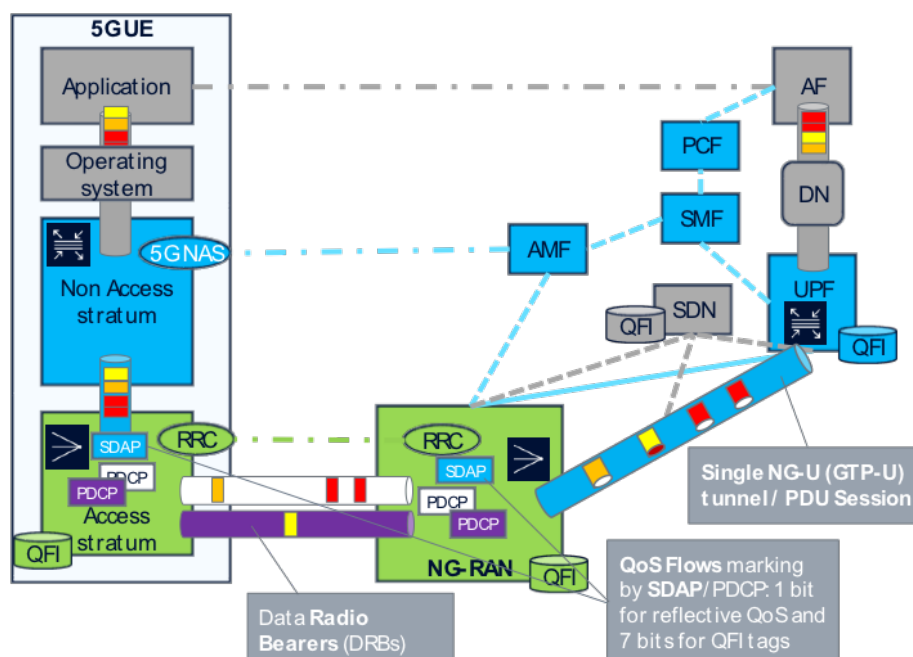


Figure 12

5G flow based QoS within a given Network Slice Instance [6]. (© Nokia 2018)

3GPP for terminal (UE) assisted network slicing defines a new parameter denoted as *Single-Network Slice Selection Assistance Information* (S-NSSAI). Each S-NSSAI assists the network in selecting a network slice instance. The S-NSSAI is composed by the following attributes:

- *Slice/Service Type* (SST): 1 (eMBB), 2 (URLLC), 3 (MIoT) are the standardized values for roaming; operator specific settings are also possible;
- A *Slice Differentiator* (SD): *Tenant ID*, optional, for further differentiation during the NSI selection.

The *Network Slice Selection Assistance Information* (NSSAI) consists of a collection of S-NSSAIs. Maximum eight S-NSSAIs may be sent in signaling messages between the UE and the Network.

The NSSAI is configured (*Configured NSSAI*) in the UE per *Public Land Mobile Network* (PLMN) by the *Home PLMN* (HPLMN). The terminal uses the *Requested NSSAI* (UE) during the *Registration Procedure* and the *Allowed NSSAI*, received from the Access and Mobility Function (AMF), within its *Registration Area* (RA). The RA allocated by AMF to UE has *homogeneous support for network slices* [10]. The 5GC supports *AMF level slicing* per UE type, and *SMF and UPF level slicing* per Service or per Tenant based on S-NSSAI and DNN. An example of three network slices for two terminal types is illustrated in Figure 13.

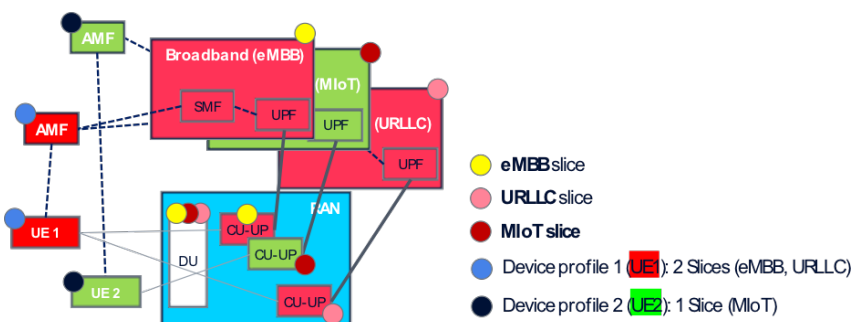


Figure 13

Example of three network slices with two different terminal types. (© Nokia 2018)

The NG-RAN is aware of the slice at PDU Session level, because the S-NSSAI is included in any signaling message containing PDU Session info. *Pre-configured slice* enabling in terms of NG-RAN functions is implementation dependent. An example of NG-RAN slicing is depicted in Figure 14. The medium access control (MAC) scheduling – based on radio resource management (RRM) policy related to the SLA in place for the supported slice and QoS differentiation within the slice – is vendor dependent [5].

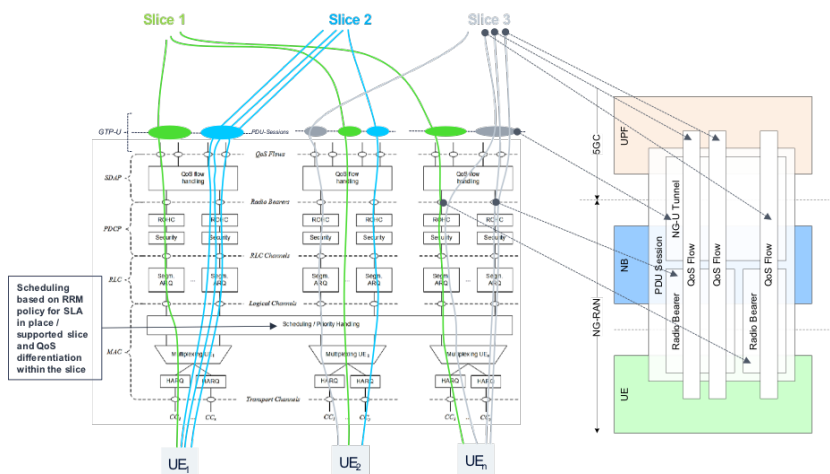


Figure 14

Example of three NG-RAN slices with different terminals and multiple PDU sessions and QoS flows. (© Nokia 2018)

The following NSSAI is exchanged between the UE and the network entities during the Registration procedure (S-NSSAI-A – DNN#A) [10], see Figure 15:

1. The configured NSSAI is provisioned in the UE based on *NS Selection Policy* (NSSP): Rules associating an App with a S-NSSAI and DNN linked to the App.
2. UE sends Registration Request message including the *Requested NSSAI*. RAN selects the AMF based on Requested NSSAI plus other criteria. 2'. Security is performed (not shown in the figure).
3. AMF fetches subscription data from UDM. UDM returns the Subscription data and Subscribed NSSAI to AMF.
4. AMF interrogates NSSF for Slice Selection. The NSSF returns set of NSI ID and Allowed S-NSSAI.
5. AMF sends Registration Accept/Complete (Allowed S-NSSAIs, 5G-GUTI, Registration Area, Mobility restrictions, etc.) to both NG-RAN and UE (NAS message, transparent to NG-RAN).
6. Subsequently, the UE uses the updated NSSAI, forwarding the list of Allowed S-NSSAIs.

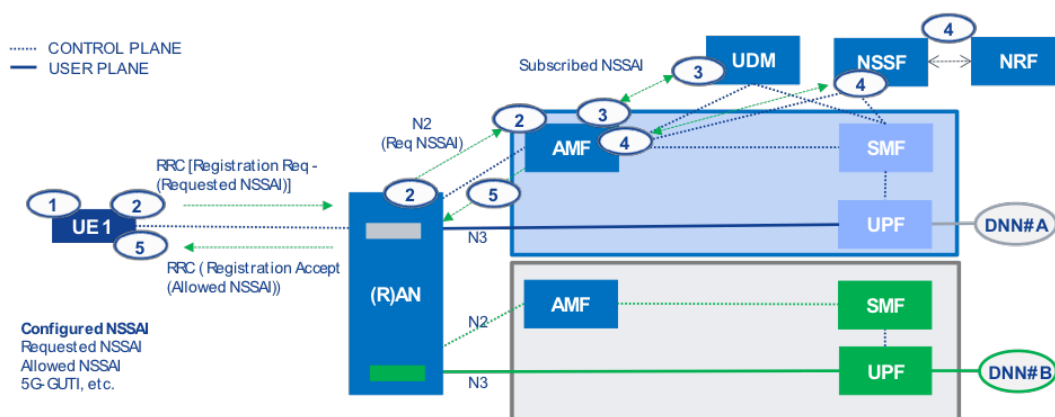


Figure 15

Network slice assistance information exchanged during the UE Registration procedure [6]. (© Nokia 2018)

The following NSSAI is exchanged between the UE and the network entities during the PDU Session establishment procedure (S-NSSAI-A – DNN#A) [10], see Figure 16:

1. UE start APP A, which is mapped onto S-NSSAI-A, and sends a PDU Session Establishment request for a given DNN and S-NSSAI.
2. NG-RAN sends a PDU session establishment to AMF (selected during the UE Registration).
3. AMF gets from Network Repository Function (NRF) the Session Management Function (SMF) for the given NSI and S-NSSAI.
4. AMF sends a Nsmf PDU Session Create Request to SMF.

- 1



0

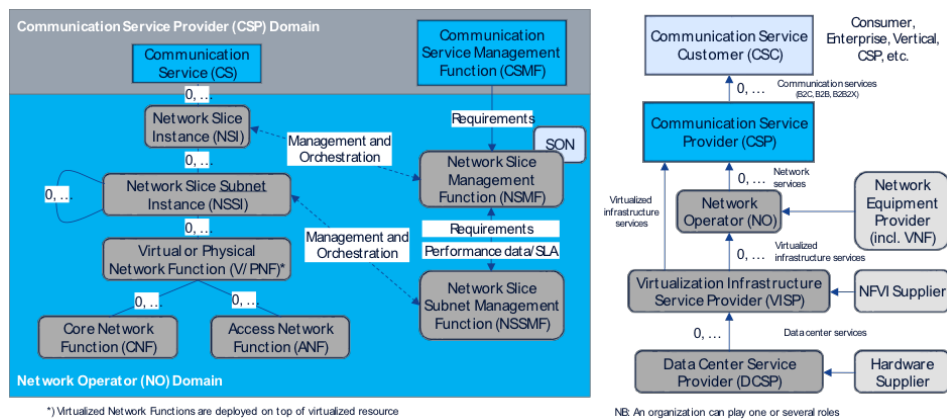


Figure 17
Network slicing: Information and role models [11]. (© Nokia 2018)

The new CSP offering, enabled by 5G Slicing, is **Network Slice as a Service (NSaaS)**. Like cloud computing SaaS, IaaS and PaaS models, the Communication Service Customer, i.e. the *Tenant*, may compose, order and pay one or more network slice instances based on its utilization; service level agreements (SLA), e.g. in terms of latency, throughput, and reliability; and value-added services (VAS), such as MMS, VoIP.

In practice, 5G will support three basic business models for network slicing, depending on the tenant's degree of slice control, which may go from **monitoring only** the KPIs related to the signed SLAs, **changing the configuration** of the deployed slice instances, to **chaining own physical/virtual network functions**, as resources and capabilities may be located in multiple domains or partners' domains or platforms.

Figure 18 shows how the partitioning model may be combined with the layering model to provide joint horizontal and vertical offerings. In slice partitioning, the *orchestration* of resources and capabilities, from an E2E service requirement perspective, must be *horizontally federated* (cooperation/collaboration), and *vertically coordinated* (hierarchy) through policies and standardized Interfaces/APIs. This combined view is also the foundation of the modeling principle of the TM Forum Platform concept [12].



Figure 18
Example of horizontal and vertical 5G slicing [12]. (© Nokia 2018)

2.1 5G solutions

The e2e 5G systems will consist of the following domains, from terminal (UE) to the Data Network (DN) and service applications, see Figure 19:

- **Terminals:** supporting Network Slice Selection Assistance Information, to request specific slice instances, based on the communication services in use.
- **Access:** eLTE/NR Radio Units with ethernet front-haul (eCPRI) or ethernet mid-haul for low latency and latency insensitive services, respectively.
- **Aggregation:** Radio Cloud with their own Software Defined Network Controller (SDN-C) and Virtual Infrastructure Managers (VIM).
- **Mobile core:** Core Cloud with own SDN-C and VIM interconnected to the Radio Clouds by IP routers and Wide Area Network (WAN) SDN-C.
- **Network Slice Management and Orchestration:** An E2E Service Orchestrator for the embedding of Network Service Descriptors (Network Connectivity topology and VNF forwarding graphs), on top of a Self-Organizing Network (SON) and VNF Manager functions.
- **Data Layer and Application Enablement:** E.g. IoT and Customer Experience Management (CEM) platforms for running applications on top of the different network slices for public safety, digital health, mobility, industry automation, smart cities, etc.

The network slice management and orchestration consists of the following steps, e.g. see [13]:

1. The CSP defines a **Network Service Descriptor** (NSD), which consists of two parts: **Network Connectivity Topology** (NCT) and **Virtual Network Function Forwarding Graph** (VNF-FG). The NCT specifies the VNF nodes that compose the network service and connection between them using Virtual Links (VLs). Each VL is connected to a VNF through a Connection Point (CP), which represents the VNF interface. The VNF Forwarding Graph (VNF-FG), which is established on top of the network connectivity topology, is composed of network Forwarding Paths (FPs), i.e. ordered lists of CPs that form a VNF chain.
2. The **end to end Service Orchestrator** (E2E SO) instantiate the network slice, end to end, based on the NSD via its southbound-interfaces: 1) It gets the optimal resources location from SON function, and asks the WAN SDN-C to instantiate WAN SDN-connections between data centers. 2) It asks VIM to reserve the virtual resources and instructs the data center SDN-C to establish the internal data center links. 3) It commands the Virtual Network Function Manager (VNFM) to a) instantiate each VNF and b) apply the pre- and post-installation scripts to complete the commissioning of the VNF. 4) Ultimately, it requests the NM/EM to configure PNFs and VNFs. (Not shown in see Figure 19).
3. The Network Slice Selection Function (NSSF) binds the CPs, NSSAI and NSI-ID for the network slice selection during the terminal Registration and PDU Session Setup, as depicted in Figure 15 and Figure 16.

Artificial Intelligence, in terms of descriptive, predictive and prescriptive analytics, will find application in three main areas:

- **SON**: Key capabilities / algorithms / architecture attributes within the different domains (RAN, Core, Transport etc.) to enable the right flexibility and tradeoffs for operators to efficiently exploit slicing; see e.g. [14];
- **Data and application layers**, i.e. big data analytics (structured data analytics, text analytics, web analytics, multimedia analytics, network analytics, mobile analytics), and
- **Data-layer platforms** for IoT and Customer Experience Management (CEM).

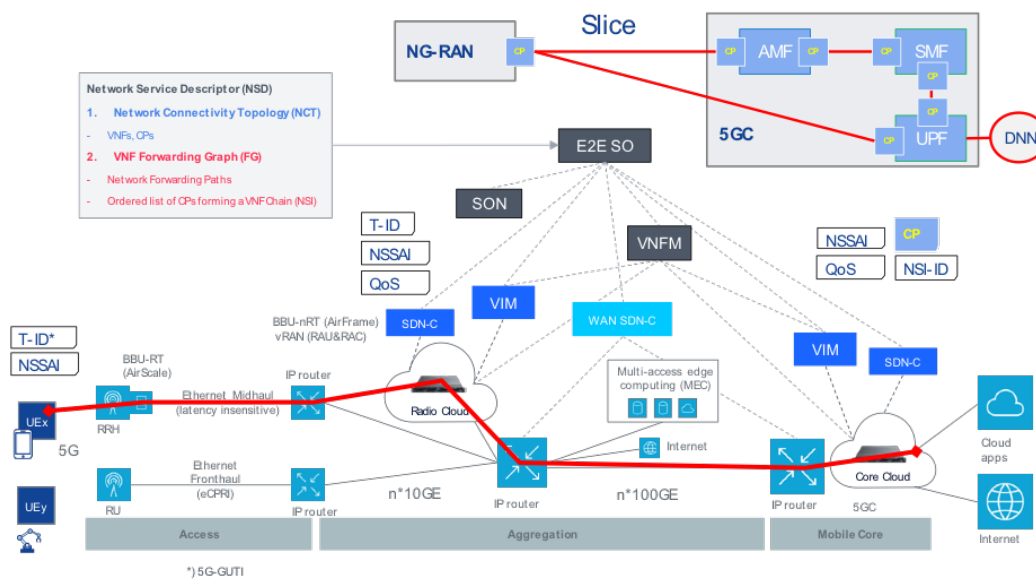
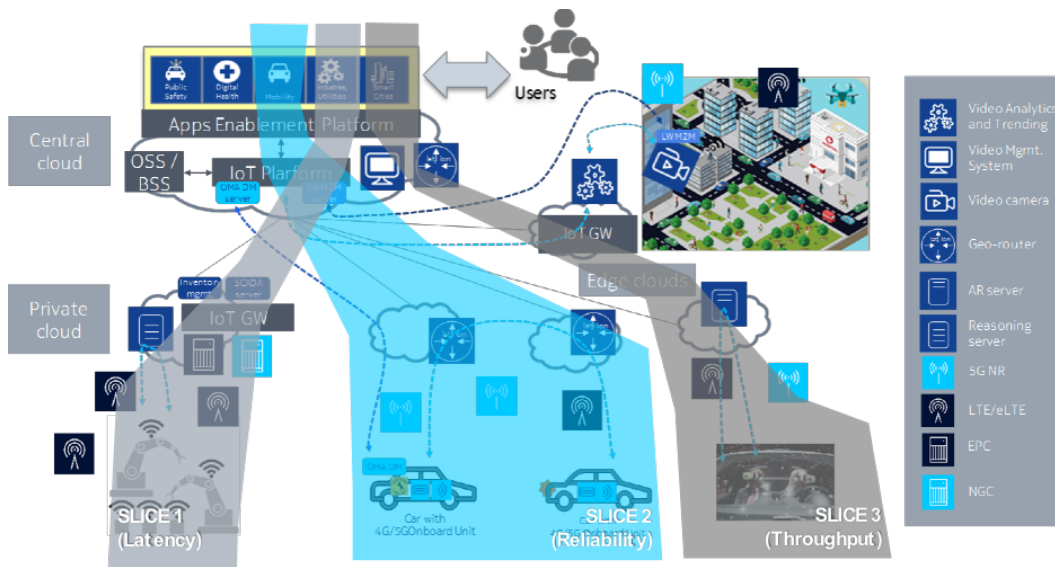


Figure 19
Example of 5G solution to end to end network slicing. (© Nokia 2018)

Figure 20 shows an example of 5G systems deployment. For other relevant scenarios see, for instance, [15]. The system consists of NG-RAN, 5GC, in a SA or NSA configuration with or without EPC; and IoT Platform and, especially, an Application Enablement Platform (AEP) on top, where verticals (tenants) may install and run own applications.

A Tenant Graphical User Interface (GUI) and an Operation GUI allow the communication service provider to flexibly offer network services to customers (tenants) and instantiate slices through automation. Moreover, the Tenant GUI allow them to place orders and monitor the key performance indicators (KPIs) linked to the signed SLAs with the corresponding CSP.

**Figure 20**

Example of NaaS for three verticals, industry process automation, automotive and smart city, with different SLA, in terms of latency (Slice 1), reliability (Slice 2) and throughput (Slice 3), respectively. (© Nokia 2018)

3. Conclusions and future developments

To realize the 5G vision presented in this paper, spectrum must be made available first, global standards (especially 3GPP technical specifications, but not limited to) next, and regulations must follow. It is also necessary a massive investment from Industry, especially from Connectivity Service Providers (Operators).

The assumptions and views reported herein are solely those of the author, and do not necessarily represent those of Nokia.

Bibliography

- [1] 3rd Generation Partnership Project, "Study on scenarios and requirements for next generation access technologies," 3GPP TR 38.913.
- [2] Nokia white paper: "5G Technology Components" <https://networks.nokia.com/innovation/5g>
- [3] Nokia white paper: "Dynamic end-to-end network slicing for 5G white paper" <https://networks.nokia.com/innovation/5g>
- [4] Nokia white paper: "Translating 5G use cases into viable business cases" <https://networks.nokia.com/innovation/5g>
- [5] 3rd Generation Partnership Project, "NR and NG-RAN overall description; Stage 2," 3GPP TS 38.300.
- [6] 3rd Generation Partnership Project, "System architecture for the 5G system; Stage 2," 3GPP TS 23.501.
- [7] 3rd Generation Partnership Project, "Study on New Radio (NR) access technology," 3GPP TR 38.912.

- [8] Nokia white paper: "Beamforming for 4.9G/5G" <https://networks.nokia.com/innovation/5g>
- [9] Nokia white paper: "Designing cloud-native 5G core networks" <https://networks.nokia.com/innovation/5g>
- [10] 3rd Generation Partnership Project, "Procedures for the 5G System; Stage 2," 3GPP TS 23.502.
- [11] 3rd Generation Partnership Project, "Study on management and orchestration of network slicing for next generation network," 3GPP TR 28.801.
- [12] TM Forum, "Dynamic Network Slices Management and Business Models", IG1152, Release 17.0.0, July 2017. <https://www.tmforum.org/resources/exploratory-report/ig1152-dynamic-network-slices-management-and-business-models-r17-0-0/>
- [13] M. Mechtri, C. Ghribi, O. Soualah, and D. Zeghlache, "NFV orchestration framework - Addressing SFC challenges," IEEE Communications Magazine, June 2017.
- [14] S. Vassilaras, et al., "The algorithmic aspects of Network Slicing," IEEE Communications Magazine, August 2017.
- [15] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," IEEE Communications Magazine, May 2017.

Biography

David Soldani received a M.Sc. degree, *Laura Vecchio Ordinamento*, in Electronic Engineering with *magna cum laude* from Università degli Studi di Firenze, Italy, in 1994; and a D.Sc. degree in Technology with *distinction* from Helsinki University of Technology (Aalto University), Finland, in 2006. In 2014 and 2016, he was appointed Visiting Professor and Industry Professor at University of Surrey, UK, and University of Technology Sydney, Australia, respectively. From 1997 to 2006 he was at Nokia in various technical and research management positions. From 2007 to 2009 he served at Nokia Siemens Networks (NSN) as Head of Customer Networks & Solutions and Solutions & Services Innovation functions, Research Technology & Platforms (RTP), Munich, Germany. Prior to joining Nokia, he was a research engineer at Rohde & Schwarz (R&S) and Sirti S.p.A., in Milan, Italy, and officer (lieutenant) at the Italian Military Navy, Livorno, Italy. From Feb 2009 to Aug 2016, he was at Huawei European Research Centre, Munich, Germany, as Head of Central Research Institute (CRI) and VP Strategic Research and Innovation, in Europe; and represented Huawei in the Board of Directors of The 5G Infrastructure Association (5G-IA) and NetWorld2020 European Technology Platform (ETP). He rejoined Nokia in 2016, and, since then, he has been serving the organization as Head of 5G Technology, End to End, Global. Areas of his responsibilities and expertise include, but not limited to: future wireless, network, big data value, computing, IoT and multimedia technologies.

Email: dr.david.soldani@gmail.com