5G Radio Access Network Architecture – Design Guidelines and Key Considerations

Patrick Marsch[1], Icaro Da Silva[2], Ömer Bulakci[3], Milos Tesanovic[4], Salah Eddine El Ayoubi[5], Thomas Rosowski[6], Alexandros Kaloxylos[3], Mauro Boldi[7]

[1]Nokia, [2]Ericsson AB, [3]Huawei European Research Center, [4]Samsung Electronics UK, [5]Orange, [6]Deutsche Telekom AG, [7]Telecom Italia

1. Introduction

After several years of research on 5th generation (5G) wireless and mobile communications, there is broad consensus on the 5G service landscape, in particular on the view that 5G will not just be a "business-as-usual" evolution of 4G networks with new spectrum bands, higher spectral efficiencies and higher peak throughput, but also target new services and business models. The main 5G service types typically considered are **Extreme Mobile BroadBand** (xMBB, a.k.a. eMBB) with data rates up to several Gbps in some areas and reliable broadband access over large coverage areas, **Massive Machine-Type Communications** (mMTC) requiring wireless connectivity for, e.g., millions of power-constrained sensors and actuators, and **Ultra-reliable MTC** (uMTC, a.k.a. URLLC) requiring end-to-end latencies of less than 5 ms and 99.999% reliability for, e.g., vehicle to anything (V2X) communication [1-3].

Investigations on the overall 5G radio access network (RAN) architecture that can efficiently support the stated diversity of services and related requirements are still ongoing. This paper describes the latest considerations from the 5G Public Private Partnership (5G PPP) project METIS-II, which includes some of the major players in the mobile industry and aims at providing a comprehensive 5G RAN design and obtaining pre-standardization consensus on key RAN design principles.

The paper is structured as follows. In Section 2, identified key 5G RAN design requirements are listed. Section 3 describes the envisioned 5G air interface (AI) landscape and latest considerations on how different air interface variants (AIVs) may be integrated into one overall 5G air interface. Section 4 captures overall system architecture considerations, such as the logical split between core network (CN) and RAN and related interfaces, while Section 5 ventures into key functional design aspects for the 5G RAN. Finally, conclusions are presented in Section 6.

2. Key 5G RAN Design Requirements

Due to the diverse and extreme requirements of the mentioned main 5G service types, it is clear that the 5G RAN must be designed to operate in a wide range of spectrum bands with diverse characteristics, such as channel bandwidths and propagation conditions [1]. It must further be able to scale to extremes in terms of throughput, number of devices, connections etc., which is likely only possible if it can handle the so-called user plane (UP), related to the transmission of actual application payload, and control plane (CP), related to control functionality and signalling, individually. To provide scalability also in the context of

various possible deployments and an evolving application landscape, it is essential that the overall 5G network (both RAN and CN) is software-configurable, meaning that, e.g., the logical and physical entities to be traversed by CP and UP packets are configurable.

A key aspect described more in Section 3 is that the 5G RAN should offer the option to integrate Long-Term Evolution Advanced (LTE-A) evolution and novel 5G radio technology on RAN level, though integration need not always take place on this level.

The 5G RAN should further support more sophisticated mechanisms for traffic differentiation than legacy systems in order to fulfill diverse and more stringent Quality of Service (QoS) requirements, and it should facilitate the Network Slicing vision from NGMN [2], enabling to operate multiple independent logical networks for different business cases on a shared physical infrastructure (see also Section 4).

Another required feature distinctive from legacy systems is the native and efficient support of communication forms like multi-connectivity (e.g. concurrent communications of a device with multiple network nodes) and network-controlled device-to-device (D2D) communication, including point-to-point, multi-cast and broadcast communication. The 5G RAN should further support a wide range of physical deployments, from distributed base stations to centralized cloud-RAN deployments or distributed edge clouds. Different types of backhaul shall also be supported with graceful performance degradation associated to the backhaul quality in terms of delay and capicity. Also self-backhauling is seen as an important feature, where also devices may act as base stations and self-establish wireless backhaul links to suitable donor base stations.

Last but not least, the 5G RAN must be highly energy efficient, e.g. by minimizing the amount of always-on signals and enabling efficient network sleeping modes, and future-proof, i.e. enabling an efficient introduction of new features and services and backward-compatibility of devices in future releases.

3. Air Interface Landscape and Integration into one 5G Air Interface

In order to handle simultaneously the extreme requirements of 5G services and use cases, an overall 5G AI is envisioned, as shown in Figure 1, which is composed of new 5G AIVs as well as evolved legacy technologies like LTE-A.

This overall 5G AI is expected to operate on a wide range of spectrum bands, where frequencies below 6 GHz are likely most suitable to support, e.g., mMTC services where coverage is most important, while spectrum above 6 GHz is essential to provide the massive capacity demanded by xMBB applications. Three authorization schemes or mixtures thereof are expected to coexist for the spectrum used by 5G: Primary user mode, Licensed Shared Access (LSA) mode and Unlicensed mode. For example, the License Assisted Access (LAA) approach already considered for LTE-A is a combination of "dedicated licensed spectrum" for primary users aggregated with unlicensed spectrum bands [1]. An exclusive use of spectrum should remain the main and preferred solution, while a shared use of spectrum may be a complement to increase spectrum availability.
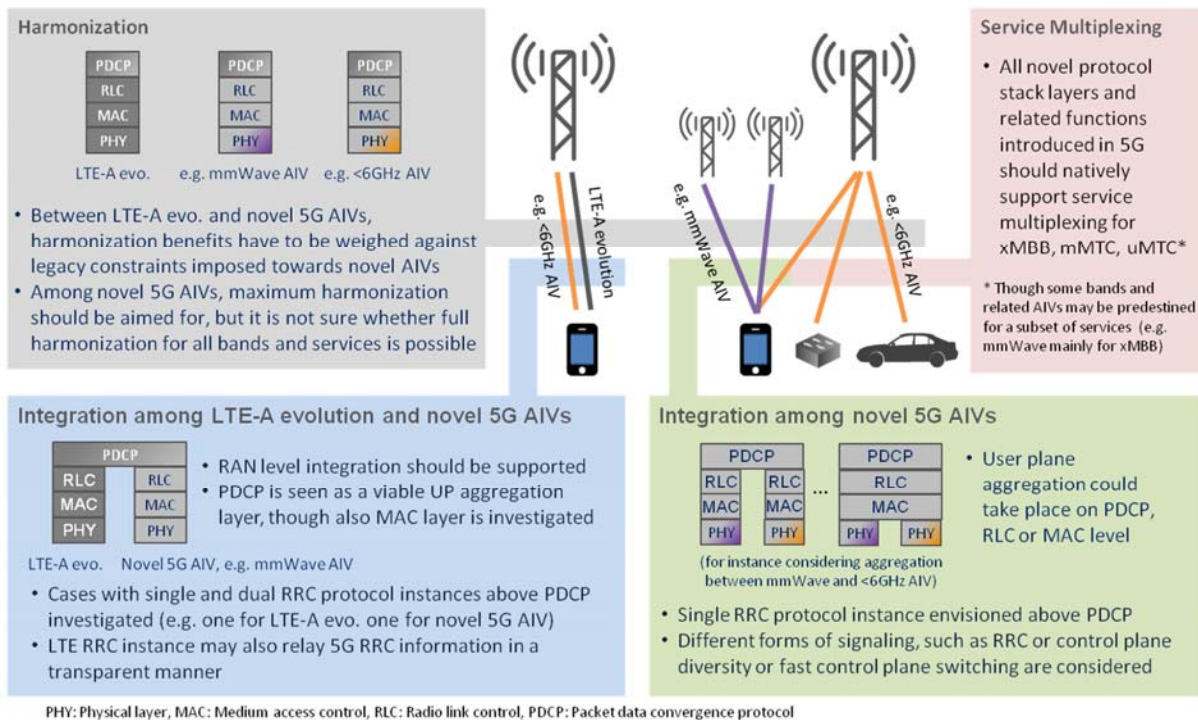
Harmonization

PDCP / RLC / MAC / PHY — LTE-A evo.

PDCP / RLC / MAC / PHY — e.g. mmWave AIV

PDCP / RLC / MAC / PHY — e.g. <6GHz AIV

- Between LTE-A evo. and novel 5G AIVs, harmonization benefits have to be weighed against legacy constraints imposed towards novel AIVs
- Among novel 5G AIVs, maximum harmonization should be aimed for, but it is not sure whether full harmonization for all bands and services is possible

Service Multiplexing

- All novel protocol stack layers and related functions introduced in 5G should natively support service multiplexing for xMBB, mMTC, uMTC*

* Though some bands and related AIVs may be predestined for a subset of services (e.g. mmWave mainly for xMBB)

e.g. <6GHz AIV / LTE-A evolution / e.g. mmWave AIV / e.g. <6GHz AIV

Integration among LTE-A evolution and novel 5G AIVs

PDCP / RLC / MAC / PHY — RLC / MAC / PHY

LTE-A evo.   Novel 5G AIV, e.g. mmWave AIV

- RAN level integration should be supported
- PDCP is seen as a viable UP aggregation layer, though also MAC layer is investigated
- Cases with single and dual RRC protocol instances above PDCP investigated (e.g. one for LTE-A evo. one for novel 5G AIV)
- LTE RRC instance may also relay 5G RRC information in a transparent manner

Integration among novel 5G AIVs

PDCP / RLC / MAC / PHY — RLC / MAC / PHY ... PDCP / RLC / MAC / PHY — PHY

(for instance considering aggregation between mmWave and <6GHz AIV)

- User plane aggregation could take place on PDCP, RLC or MAC level
- Single RRC protocol instance envisioned above PDCP
- Different forms of signaling, such as RRC or control plane diversity or fast control plane switching are considered

PHY: Physical layer, MAC: Medium access control, RLC: Radio link control, PDCP: Packet data convergence protocol

Figure 1. Overall 5G AI envisioned, and key considerations on AIV integration in 5G.

LTE-A and its evolution is likely to play a pivotal role and serve as coverage layer and potentially also as anchor layer particularly in early 5G deployments when novel 5G AIVs will not yet be able to provide the same coverage as LTE-A. Novel 5G AIVs may operate in conjunction with LTE-A or stand-alone, and may, for instance, be designed for specific frequency ranges, services, or cell types etc. For example, an AIV tailored towards lower carrier frequencies, large cell sizes and high velocity will likely have a physical layer (PHY) designed to be most robust towards delay spread and Doppler spread, whereas an AIV tailored towards mmWave frequencies and short-distance communication with limited mobility may rather require robustness towards other impairments such as phase noise. Further, in order to support applications requiring very low latencies and/or very high data rates, some new 5G AIVs are expected to use shorter transmission time intervals (TTIs) and a wider bandwidth compared to LTE-A.

The exact waveform(s) to be used for novel AIVs are still under investigation, but it appears clear that key properties will be a flexible and scalable numerology enabling user- and service-specific adaptations, flexible sub-band configurations, improved spectral efficiency, support for flexible time division duplex (TDD) and reduced out-of-band (OOB) emissions. Two approaches that are currently being compared are to have either a single waveform type parameterized to support all services and bands, or a co-existence of different waveforms such as variants of orthogonal frequency division multiplex (OFDM) and filter-bank multi-carrier (FBMC) [4].

A key question is how the different AIVs, including LTE-A evolution, can be integrated into one overall 5G AI such that standardization and implementation complexity are minimized, while the performance of

individual AIVs, e.g. tailored towards certain frequency bands and services, is not sacrificed. The following notions of integration are considered:

- **Protocol harmonization**. Two protocol stack instances related to different AIVs (e.g. serving different bands) are considered as harmonized if both can be derived from the same definition through parameterization. For example, UE and network procedures such as initial access and mobility should ideally be as similar as possible for different bands like mmWave and cmWave, bearing in mind the existence of technologies tailored to each of these bands (e.g. narrowband beamforming for mmWave). In the PHY, the individual blocks of modulation, channel coding, waveform mapping etc. would ideally be interchangeable through parameterization or activation, addition or removal of certain features via a physical or logical implementation supporting switching between different variants, and even their multiplexing in time, frequency and/or space. In practice, however, different properties of AIVs may require physically separate baseband processing chains, rendering full PHY harmonization challenging.
- **Service multiplexing** relates to the capability of a single logical instance of a protocol stack layer to handle multiple bearers or flows related to different devices and service types like xMBB, uMTC and mMTC.
- **User plane aggregation** relates to the possibility of aggregation of the UP protocol stack instances reflecting multiple instances of the same AIV (for multi-connectivity on one band), or different AIVs (e.g. in the case of different bands) on a certain protocol stack layer. Such aggregation means that on and above this layer there is only one single logical protocol stack instance, implying that the protocol stacks also have to be harmonized on and above this layer, and rendering the higher layers agnostic of the existence of multiple protocol stack instances or AIVs on the lower layers. MAC layer aggregation enables features such as cross-AIV scheduling, but may be challenging in the context of, e.g., PHY layers having different frame structure, and is typically only possible in co-located deployments or deployments with good backhaul quality. PDCP-level aggregation can enable several features similar to MAC-level aggregation (not necessarily with the same gains) except cross-carrier scheduling, with the benefit of being more suitable for distributed deployments with non-ideal backhaul and not requiring the harmonization of the lower AIV layers.
- **Control plane integration**. Similarly, one has to consider on which protocol stack layer multiple instances of the same AIV, or different AIVs, may share one single CP instance.

Clearly, the most suitable extent of harmonization, or the most suitable protocol stack layer of UP aggregation or CP integration depends on whether one is considering the integration of novel 5G AIVs, or the integration of legacy technology with novel 5G AIVs. In the former case, one can by design enable a larger extent of harmonization and tighter integration, while in the latter case this may not be easy to achieve due to legacy constraints that may pose limitations on novel 5G AIVs. The mentioned notions of integration along with the current key integration considerations in METIS-II are depicted in Figure 1. Please note that all aspects are still under research and subject to conclusion by the end of the project.

4. Overall 5G System Architecture Considerations

Considerations on CN, split between RAN and CN and Network Interfaces

In alignment with the NGMN vision [2], it is expected that the majority of CN and Service-Layer functions are deployed in 5G as virtual network functions (VNFs), thus running in virtual machines on standard servers, potentially on cloud computing infrastructures, i.e. data centers. The design of these functions will to some extent explore software-defined networking (SDN) principles such as a split of UP and CP, and allow for their flexible deployment in different sites in operators' networks depending on requirements related to latency, available transport, processing and storage capacity etc. Moreover, different services or network slices may utilize different CN and Service-Layer VNFs deployed at different network sites.

An important assumption, also considered by the 3rd Generation Partnership Project (3GPP) [5], is a logical split between RAN, CN and Service Layer functions. This is seen as beneficial because it:

- Allows for an independent evolution of RAN and CN functionality to speed up the introduction of new technology;
- Enables to make some CN functions access-independent (e.g. common UP processing);
- Facilitates mobility since some CN functions can be kept when UEs move to a new RAN node;
- Allows cross-layer optimizations when the functions are co-located;
- Facilitates multi-vendor CN/RAN interoperability.

The exact logical split between RAN and CN has not been defined yet. While the evolved packet system (EPS) provides a natural baseline for the split [6], enhancements are being investigated such as RAN-based paging for dense deployments and a new connected state optimized for inactivity periods, both resembling a shift of functions from CN to RAN (see Section 5).

In order to address future architecture requirements, a novel CN/RAN interface denoted S1* is envisioned which supports:

- E2E Network Slicing (where each slice may have its own set of CN functions);
- New 5G services with diverging requirements (where CN functions can be optimized for a specific service);
- Enhanced multi-RAT integration with common CN functions where some could be designed to be independent of the access technology;
- Potentially new UP/CP splits in the 5G CN;
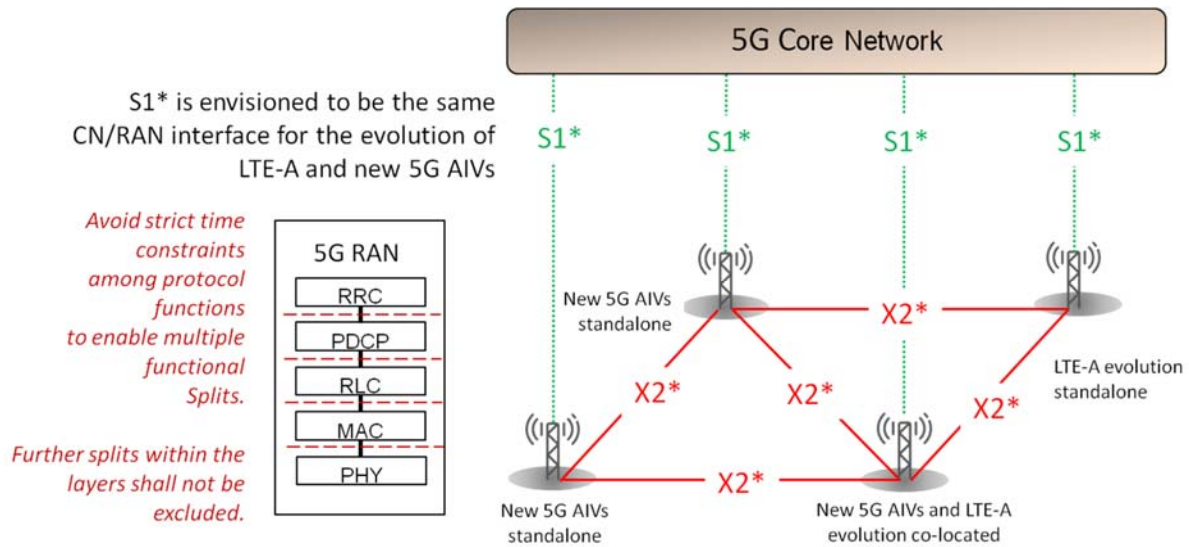- A new connected state, optimized for battery savings but enabling a fast transition to active.

Figure 2. Considered 5G network interfaces.

Also the existing X2 interface is expected to evolve (denoted herein X2*), in particular to support inter-node multi-connectivity and mobility among novel 5G AIVs (e.g. for mmWave and cmWave bands) and LTE-A evolution.

Deployment Architecture and Placement of Logical Functions

As mentioned in Section 2, the 5G RAN should support a wide range of physical deployments and be able to maximally leverage from centralization, while also supporting distributed base stations and being able to operate in non-ideal backhaul. A key enabler for this is the implementation of some radio functions as VNFs, as mentioned before in the CN context, allowing to flexibly shift these towards or away from the radio edge depending on the physical architecture and specific application requirements. Preliminary analyses [7] concluded that functions which are time-*asynchronous* to the radio interface (in LTE these are packet data convergence protocol (PDCP) and radio resource control (RRC) functions related to measurement control and reporting, handover preparation and execution, dual connectivity, random access, RRC state transition etc.) are most suitable to be implemented as VNFs and possibly centralized, as they typically require low data rates on their interfaces, scaling with the number of users and not the overall traffic. Further, these functions can typically cope with larger latency (e.g. tens of milliseconds in LTE).

Time-*synchronous* functions (in LTE these are PHY, medium access control (MAC) and radio link control (RLC) functions such as scheduling, link adaptation, power control, interference coordination etc.), however, typically require high data rates on their interfaces, scaling with the traffic, signal bandwidth and the number of antennas, and benefit from hardware acceleration. The potential for centralization is here most pronounced in deployments with low-latency and high-bandwidth backhaul due to timing and real-time processing requirements.
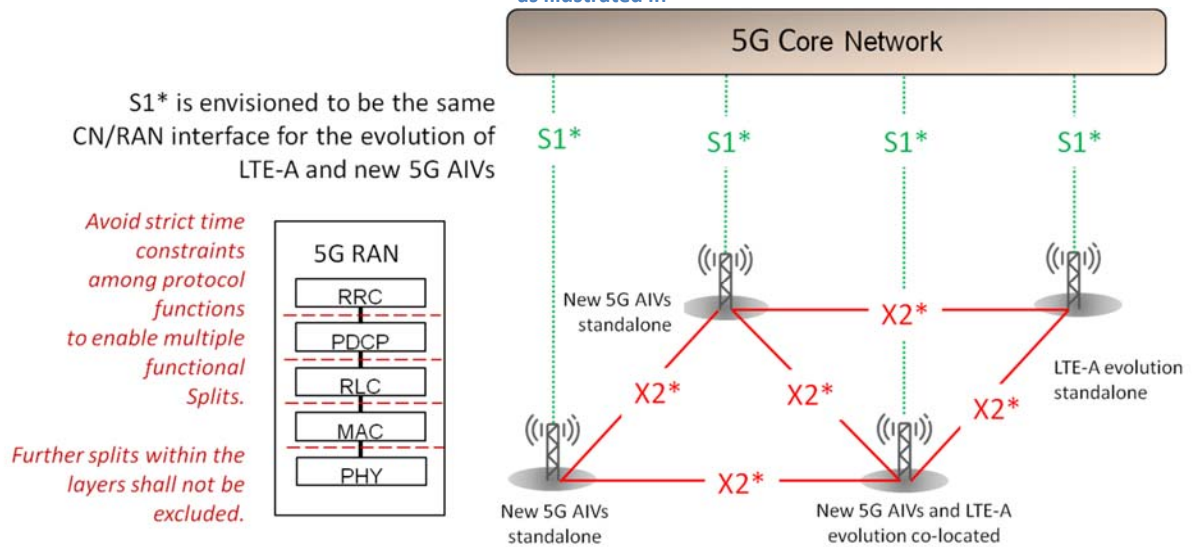
Figure 2. In this way, the 5G RAN supports various functional splits and allows catering for a diverse range of physical deployments and backhaul types.

RAN Architecture support for Diverse Services and Network Slicing

Flexibility and configurability will be key RAN characteristics to support the diverse services and related requirements mentioned before in one common network infrastructure. This may be realized by a protocol architecture supporting a service-specific selection of network functions (NFs) and service-tailored optimizations. For three specific service examples, Figure 3 shows potential service-tailored NFs, and an indication of how these could be used to, e.g., minimize control plane latency. Note that the quantitative values are based on LTE-A Release 13 figures [8][9] and solely serve to illustrate the benefit of service-specific NF selection, while further latency reduction in 5G is likely. Various mentioned functional considerations such as a novel RRC state and RACH prioritization are detailed in Section 5.

Going beyond the support of diverse services, Network Slicing [2] refers to the operation of multiple independent logical networks for different business cases on a shared physical infrastructure. For instance, a dedicated network slice could be configured for a particular V2X business case, involving service-level agreements (SLAs) among stakeholders and a slice-specific selection and configuration of CN, transport network and RAN functions. Some progress has been achieved in understanding the implications of Network Slicing on the RAN architecture, e.g. yielding the common view that [10]

- even if network slices are seen as separate logical networks, an efficient reuse of resources like radio spectrum, infrastructure and transport network among the slices is essential;
- network slices (or an abstraction thereof, such as groups of service flows) need to be visible to the RAN, such that NFs can take into account overall slice-specific metrics or constraints (for

instance, there may be the constraint that all services belonging to a slice may jointly only occupy a certain amount of radio resources);

- means for slice isolation and protection are needed (e.g. it may have to be guaranteed that events in one slice cannot negatively impact another slice);

- performance monitoring solutions (e.g. counters, traces and KPIs) need to be aggregated per slice to verify the fulfillment of SLAs and/or properly operate the different businesses associated to different slices. In the case of configuration management, some features could be tuned, turned on/off and/or possibly configured differently for different slices. This may also affect the way Self-Organizing Networks are defined. A network management framework is also needed to share infrastructure among multiple slices such as hardware platforms and provide an efficient slice setup mechanism to improve the time to market of new businesses.

	Static Temperature Sensor (mMTC)	Video Streaming (xMBB)	Smart Grid (uMTC)	Example: Control plane latency from service request until initiation depending on service-specific optimization [8, 9]:	
RRC	HO measurements ommited	State handling optim. for reduced RAN/CN signaling	State handling optim. for reduced state change latency	RACH process (11 ms)	Note that these numbers are based on LTE-A Release 13 figures [8][9] and solely serve to illustrate the benefit of service-specific NF selection, while significant further latency reduction in 5G would be possible.
PDCP	Potential omitting of ciphering and header compression	default	Potential omitting of ciphering and header compression	UE Processing (2.5ms)	
				RRC Conn. Req. (1ms)	
				H-ARQ (1.5 ms)	
RLC	Unacknowledged mode only	default	Acknowledged mode only	eNB processing (8ms)	Optimized RACH process (7 ms)
				H-ARQ (1.5 ms)	UE Processing (2.5 ms)
MAC	H-ARQ optimized for coverage	default	H-ARQ omitted for low-latency, RACH prioritization	UE processing (3 ms)	RRC Conn. Req. (1 ms)
					eNB processing (2 ms)
				H-ARQ (1.5 ms)	UE processing (0.6 ms)
PHY	Coding optimized for coverage, energy efficiency	Coding optimized for very large payloads	Coding optimized for short payloads, low latency	*Video streaming (xMBB)* 30 ms	*Smart Grid (uMTC):* 13.1 ms

Figure 3. Examples of potential service-specific selection or tailoring of NFs.

5. Key Functional Design Aspects for the 5G RAN

Beyond the overall system architecture, the envisioned range of diverse services and the multitude of frequency bands also suggests various changes in the design of network functions in 5G, as highlighted in the following.

Beam-centric, Lean and Future-proof Design

Beamforming will play an essential role in particular for higher frequency bands, where a larger path loss has to be overcome, but where shorter wavelengths also permit a larger number of antenna elements at reasonable form factor. It will likely impact the design of synchronization and reference signals (e.g. for neighbor cell measurements), since e.g. a radio entity needing to detect a narrow beam also requires a direction synchronization procedure (in addition to time and frequency). For these reasons, solutions are

being investigated – beside the stand-alone operation of higher-frequency AIVs – where lower frequency AIVs (including LTE-A evolution) are used in combination with higher frequency layers and serve as an anchor layer for synchronization and mobility aspects. The need for beamforming at higher frequencies and the aforementioned desire to harmonize NFs across AIVs suggest introducing the notion of a beam-centric design in general for all AIVs, even though a beam in a lower frequency AIV may of course be wide and correspond to an entire cell [11].

It is currently being investigated how to obtain a lean and future-proof design that maximizes energy efficiency, reduces the amount of interference generated by common signals, and better supports beam-centric communication and the possible introduction of future signals. Beside a minimization of broadcasted signals in each cell as such, as depicted in Figure 4a, one may consider new ways to distribute common signals and for instance only let some cells transmit system information, or even apply a CP/UP split among cells, see Figure 4b. In general, one may also strive to avoid transmitting, e.g., reference signals over the entire bandwidth, but instead use self-contained transmissions as shown in Figure 4c, where reference signals are transmitted jointly with the payload, minimizing the overhead and interference, being better suited for beam-based transmission and more future-proof [12].
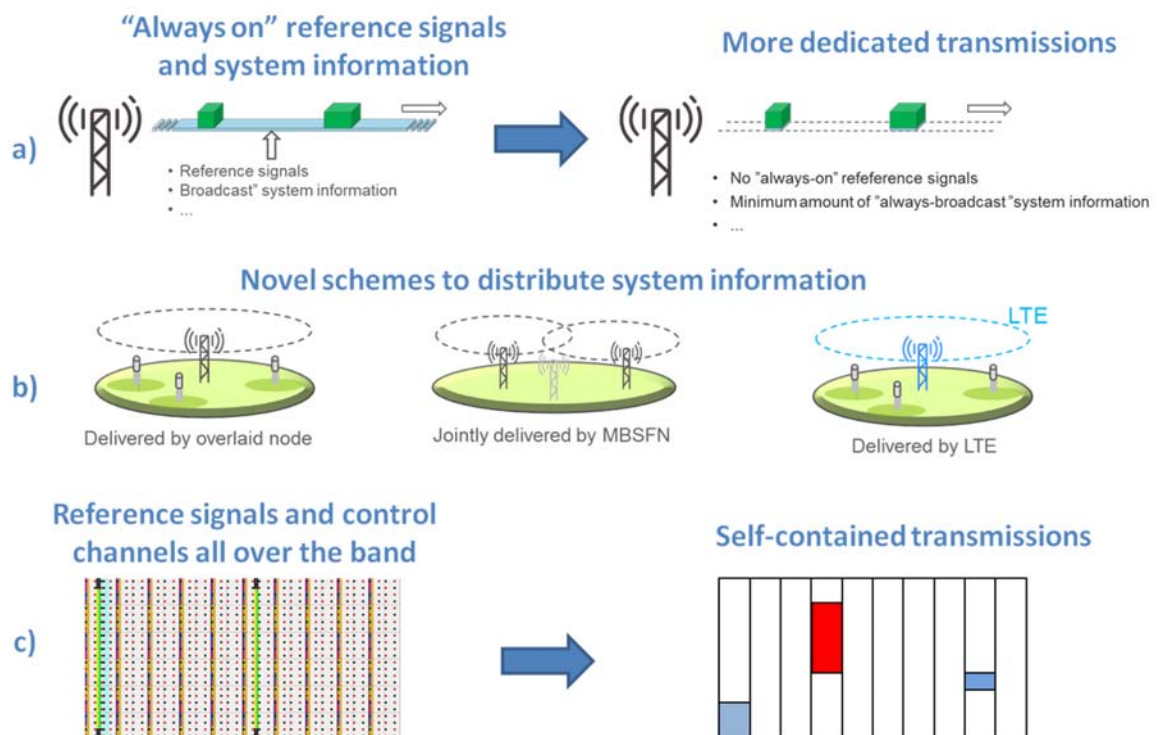

Figure 4. Beam-centric, lean and future-proof signaling design in 5G.

RACH-related considerations for service prioritization

The huge increase in the number of devices and the diversity in the accessed services will definitely pose a challenge in 5G networks. In order to provide efficient prioritization for delay-sensitive services, random

access solutions are currently investigated that provide some level of access differentiation per service, taking their accessibility requirements into account. Specifically, random access requests associated with delay sensitive services could apply a combination of preamble signatures at a given random access time slot, enabling such service requests to have a collision rate of practically zero, since they could be easily identified at the receiver side.

Resource Management and Traffic Steering in 5G

A key challenge in 5G is how to dynamically assign the foreseen wide range of services with diverse requirements to the many spectrum bands, usage types, related AIVs and the radio resources therein, also involving a more widespread usage of novel communication forms like uni-cast or multi-cast network controlled D2D. This challenge will be further pronounced through more dynamic network topologies, e.g. resulting from a flexible activation and deactivation of fixed access nodes, antennas and remote radio heads (RRHs) and the incorporation of moving cells and nomadic nodes into the RAN. Furthermore, the interference interdependencies between network nodes will increase, as for instance in very dense deployments with TDD and a flexible usage of uplink (UL) and downlink (DL), D2D or in-band self-backhauling. Consequently, it appears essential to revisit key design principles related to resource management in the context of 5G [13].

One potential paradigm change in 5G is the protocol stack layer on which traffic steering is performed, i.e. the assignment of services to AIVs. In existing systems, the assignment to radio access technologies (RATs, e.g. 3G and 4G) and cells takes place via hand-over between cells, i.e. RRC-level mechanisms. Additionally, a device may be served in multi-connectivity within a radio technology (e.g., LTE Release 12 dual connectivity), where traffic is then further steered on PDCP level to the different radio legs. In 5G, considering more stringent service requirements, more degrees of freedom and larger radio dynamics especially in higher frequency bands, it may be beneficial to perform traffic steering further down in the protocol stack, e.g. on RLC or MAC layer, and hence on a faster time scale and thus overcoming the semi-static and time-consuming bearer modification in legacy systems.

A possible architecture enabling such traffic steering and related Quality of Service (QoS) management is depicted in Figure 5. Here, the Policy and Charging Enforcement Function* (PCEF*) resembles an entity in the 5G CN, possibly similar to the one currently present in the Evolved Packet Core (EPC), which defines QoS Class Identifier (QCI) characteristics for the different services. The traffic steering framework resides in the RAN and translates these AIV-agnostic QoS metrics to AIV-specific ones, based on the real-time feedback received from the AIVs. This feedback could for instance contain AIV-specific radio conditions and radio resource utilization values. As mentioned before, a 5G network will be characterized by a dynamic activation and deactivation of access nodes, for instance for energy efficiency reasons. Since such dynamic behavior of course has an impact on the degrees of freedom available for traffic steering, a joint optimization of traffic steering and so-called RAN moderation is required.
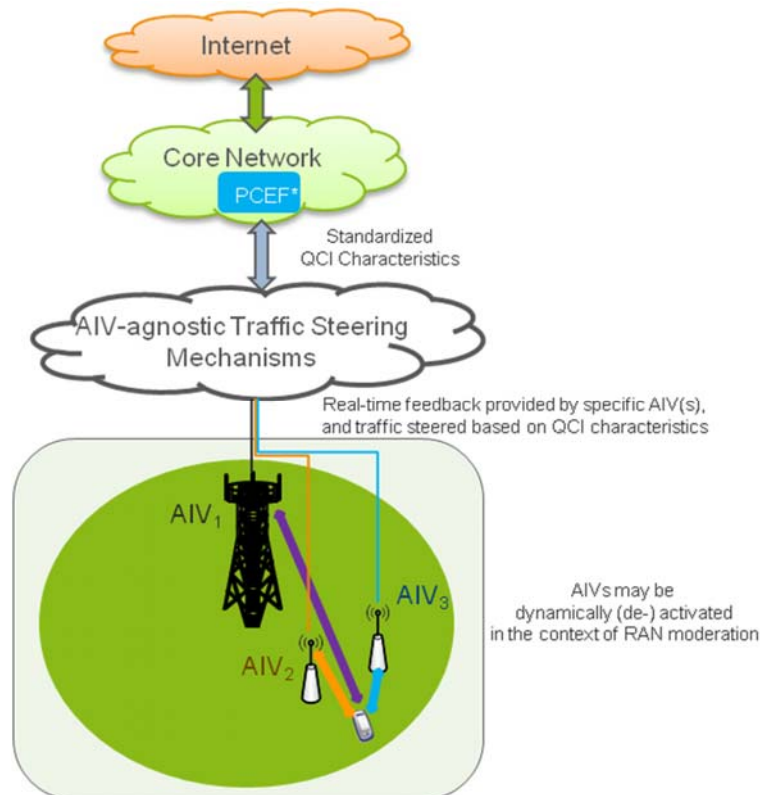
Figure 5. An example scheme for traffic steering and QoS management.

A further possible paradigm change related to resource management in 5G may be an extended notion of the term 'resource' as such: In particular in the context of network slicing and the reuse of network infrastructure among slices, it appears suitable to consider as a resource not only classical radio resources in time and frequency, but also soft network capabilities such as processing power, memory capacity and power budget.

A Novel State Model in 5G

With the trends towards the Internet of Things (IoT) it is expected that in 5G there will be even more battery-powered UEs (e.g. sensors, baggage tags, etc.). Therefore, battery efficiency and duration will be essential, especially for those devices for which accessibility is limited (e.g. remote locations, restricted areas). At the same time, the requirement for fast first packet transmission (UL or DL) may be even more stringent than in previous mobile generations. This tradeoff between device power efficiency and fast accessibility is often called "UE sleeping problem".

A novel state model is being proposed to address the problem, relying on a novel state called "RRC Connected Inactive" in addition to "RRC Connected" and "RRC Idle". This novel state explores the principle of "not discarding previously exchanged information" for inactive UEs, meaning that UEs in the new state still keep parts of the RAN context, for instance related to the security context, UE capability information,

etc. In addition to this, signaling is reduced by allowing the UE to move around within a pre-configured area without notifying the network. The state is also envisioned to be highly configurable with a wide range of discontinuous reception (DRX) cycles (from milliseconds to hours) and service-tailored optimizations related to the transition to "RRC Connected". Figure 6 shows the three proposed states and state transitions.

In the novel model, transitions from "RRC Idle" to "RRC Connected" are expected to occur mainly when a UE first attaches to the network, or as a fallback case when the devices and/or the network cannot use the previously stored RAN context. On the other hand, transitions from "RRC Connected Inactive" to "RRC Connected" are expected to occur often and are hence optimized to be fast and lightweight in terms of signaling. This is achieved by keeping the CN/RAN connection alive during inactivity periods and reducing the amount of RRC signaling needed to resume an existing inactive connection via the usage of a RAN context ID – an approach that is inspired by the suspend/resume procedure to be defined for idle state UEs in LTE [14]. The benefits of the investigated novel state model in terms of protocol overhead reduction and control plane latency are shown in Figure 6 [15].
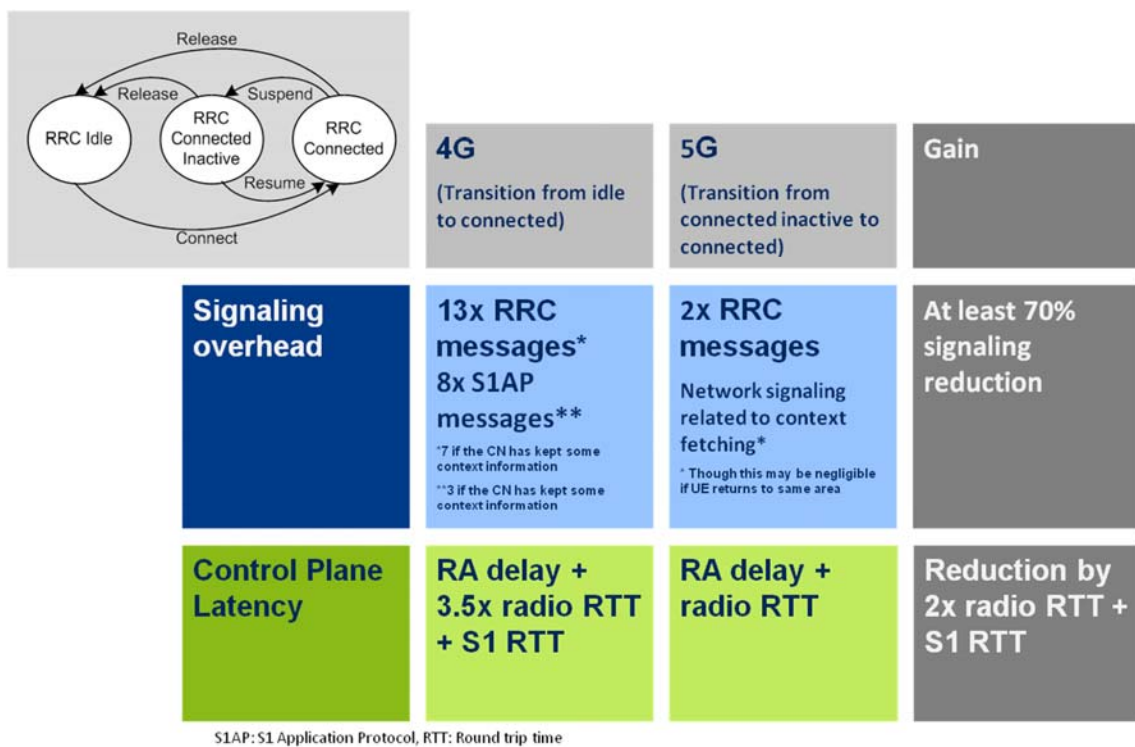

Figure 6. Benefits of the investigated novel UE state model.

RAN-based Paging

Due to the expected massive number of devices and denser deployments in 5G, paging may significantly increase the load on both the air interface and the CN/RAN interface, thus requiring new solutions for

efficient paging and UE location tracking. One potential way forward could be the introduction of such functionalities in the RAN, for instance in the form of a hierarchical location tracking where the CN tracks the registration of UEs in "RRC Connected Inactive" only on the level of groups of RAN locations, whereas the RAN tracks these on a cell-level granularity. This could involve a lightweight signaling procedure terminated in the RAN, and which uses a security handling mechanism based on retaining and updating the security context from the last attach procedure. The considered hierarchical location tracking approach would imply moving part of the paging and mobility related functionalities of the CN to the RAN, hence a change of the CN/RAN split as compared to LTE.

6. Conclusions

This paper has provided a view of the current considerations on the overall 5G RAN architecture and related key functional aspects. It has listed a set of paradigm changes expected to be introduced in 5G, for instance related to a beam-enabled lean design or the introduction of a novel UE state, as well as a set of specific functionality proposals that will be a basis for standardization in 3GPP. It has to be noted that the topics covered in this paper are still under research and subject to finalization in the next months.

ACKNOWLEDGMENT

REFERENCES

[1] ICT-671680 METIS-II, Deliverable 1.1, "Refined scenarios and requirements, consolidated test cases, and qualitative techno-economic assessment", January 2016.

[2] Next Generation Mobile Networks (NGMN) Alliance, "NGMN 5G White paper", February 2015.

[3] International Telecommunication Union – Radio (ITU-R) Working Party WP 5D: Draft New Recommendation "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond", Doc. R12-SG05-C-0199 (approved by Study Group SG5), June 2015.

[4] M. Tesanovic, M. Schellmann, P. Weitkemper, D. Calabuig, O. Aydin, C. Kilinc and V. Venkatasubramanian, "Towards a flexible harmonised 5G air interface with multi-service, multi-connectivity support ", ETSI Workshop on Future Radio Technologies: Air Interfaces, January 2016.

[5] 3GPP S2-153651, "Study on Architecture for Next Generation System", October 2015.

[6] 3GPP TS 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2", November 2015.

[7] ICT-317669 METIS, Deliverable 6.4, "Final report on architecture", January 2015.

[8] 3GPP TR 25.912, "Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN), Release 13", December 2015.

[9]   D. Szczesny, A. Showk, S. Hessel, A. Bilgic, U. Hildebrand, and V. Frascolla, "Performance analysis of LTE protocol processing on an ARM based mobile platform", International Symposium on System-on-Chip (SOC), October 2009.

[10]   I. Da Silva, G. Mildh et al., "On the impact of network slicing on 5G radio access networks", IEEE International Conference on Communications (ICC), 5G RAN Design Workshop, May 2016, submitted for publication.

[11]   F. Athley, S. Tombaz, E. Semaan, C. Tisdav and A. Furuskär, "Providing Extreme Mobile Broadband using higher frequency bands, beamforming and Carrier Aggregation", IEEE Global Conference on Communications (Globecom), December 2015.

[12]   P. Frenger, M. Olsoon, E. Eriksson, "A Clean Slate Radio Network Designed for Maximum Energy Performance", IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), September 2015.

[13]   Ömer Bulakci, Athul Prasad, Jakob Belschner; Marten Ericson, Ingolf Karls, Haris Celik, Milos Tesanovic, Roberto Fantini, Luis Miguel Campoy, Emmanouil Pateromichelakis, Fernando Sanchez Moya, Gerd Zimmermann and Icaro da Silva "Agile Resource Management for 5G - A METIS-II Perspective," IEEE Conference on Standards for Communications and Networking (CSCN), October 2015.

[14]   3GPP TR 23.720, Technical Specification Group Services and Systems Aspects, "Architecture enhancements for Cellular Internet of Things (Release 13)", November 2015.

[15]   I. Da silva, G. Mildh, M. Säily and S. Hailu, "A novel state model for 5G radio access networks", IEEE International Conference on Communications (ICC), 5G RAN Design Workshop, May 2016, submitted for publication.