White Paper

Preliminary Views and Initial Considerations on

5G RAN Architecture and Functional Design

Published on March 4th, 2016

Editors: Patrick Marsch, Nokia Networks
Icaro Da Silva, Ericsson

Authors: Salah Eddine El Ayoubi, Orange
Mauro Boldi, Telecom Italia
Ömer Bulakci, Panagiotis Spapis, Malte Schellmann, Huawei ERC
Patrick Marsch, Mikko Säily, Nokia Networks
Jose F. Monserrat, Universitat Politècnica de València
Thomas Rosowski, Gerd Zimmermann, Deutsche Telekom
Icaro Da Silva, Ericsson
Milos Tesanovic, Mehrdad Shariat, Samsung
Ahmed M. Ibrahim, Intel

METIS-II is a project funded within the framework of the 5th Generation Public Private Partnership (5G PPP), see further information under http://5g-ppp.eu.

Executive Summary

After several years of research on the 5th Generation (5G) of cellular communications, there is now a wide consensus on the main services that 5G should be designed for. A comprehensive set of key technology components that will form part of the 5G system has been identified in response. What is yet to be clarified is the detailed overall design of the 5G system. Since standardization is expected to start with first 5G radio access network (RAN) study items in the 3rd Generation Partnership Project (3GPP) already in 2016, it is important to now obtain early alignment on key RAN design aspects between key industry players and operators. The 5G Infrastructure Public Private Partnership (5G PPP) [5GPPP] project METIS-II [METIS] aims to foster exactly this pre-standardization consensus before and during the early days of standardization by providing an overall 5G RAN design corresponding to "technology readiness level 2" [EC15-AG].

This paper summarizes the initial views and considerations of METIS-II on the 5G RAN architecture and functional design. It starts by listing the main service types that are considered for 5G, namely **extreme mobile broadband (xMBB), massive machine-type communications (mMTC) and ultra-reliable machine-type communications (uMTC)**, as well as the five specific use cases towards which METIS-II is performing the 5G RAN design, and which typical represent a mixture of services. It further describes the **key requirements on the 5G RAN architecture** that have been identified and derived from the diverse service and use case needs, and explicitly elaborates on the **requirements posed by the notion of Network Slicing in 5G**.

METIS-II envisions the overall 5G RAN to operate in a **wide range of spectrum bands** to address the diverse services, for instance considering frequencies below 6 GHz as likely most suitable to support mMTC services with high coverage requirements, and spectrum above 6 GHz as essential to provide the massive capacity demanded by xMBB applications. Studies have shown that **large contiguous spectrum bands** are preferable for various reasons, in particular related to device complexity. In general, the 5G system will build upon a set of spectrum usage forms such as the **use of dedicated licensed spectrum, horizontal sharing of bands** with differentiation according to **limited spectrum pools, mutual renting and unlicensed use,** as well as **vertical sharing of bands**.

Due to this wide range of bands and the stated service diversity, METIS-II envisions **the overall 5G air interface (AI) to comprise multiple so-called air interface variants** (AIVs), including the evolution of existing radio such as Long Term Evolution Advanced (LTE-A) and novel AIVs introduced in 5G, which may be tailored towards specific bands, cell types or services. For example, two AIVs designed for bands below 3 GHz and above 60 GHz, respectively, may be distinct in terms of frame structure, the importance of beamforming and related handling of control signals etc. The precise waveform(s) and physical layer (PHY) numerologies to be used for novel AIVs, which may be derived from various waveform families such as orthogonal frequency division multiplex (OFDM) or filterbank multi-carrier (FBMC) based solutions, are still under investigation. Multiple hypotheses are currently being pursued related to the overall AIV landscape: for example, multiple waveform families may jointly cover the space of bands and services, or a single waveform family may be tailored to cover all bands and services.

A key question w.r.t. the overall RAN design is **to which extent different AIVs can be harmonized towards a single AI protocol stack specification** in order to reduce implementation and standards complexity and improve cost-efficiency for devices having to implement multiple AIVs. Regarding novel AIVs introduced in 5G, METIS-II is currently investigating the following three different kinds of harmonization (listed here in no particular order): 1) PHY harmonization of novel 5G AIVs that are potentially based on different waveform families (e.g. OFDM and FBMC based solutions), 2) Medium access control (MAC) layer (or higher) harmonisation towards a single specification that supports the

usage of different waveform families on PHY layer, and 3) scaling of a single waveform (or waveform family) across all 5G frequency bands to support all 5G use cases. Among LTE-A evolution and novel 5G AIVs, on the other hand, the benefits of harmonization have to be weighed against the potential legacy constraints imposed towards novel air interface technology.

Beyond harmonization, METIS-II investigates to which extent user plane (UP) instances related to different bands can be logically aggregated on certain layers, and beyond which layer there would be a single control plane (CP) instance. In this respect, the preliminary assumption is that for the integration **between multiple novel 5G AIVs, UP aggregation could take place on layer 2**, i.e. MAC, radio link control (RLC) or packet data convergence protocol (PDCP) level, likely dependent on the physical network architecture, and likely with **one common radio resource control (RRC) instance** w.r.t. the CP. It is further assumed that the 5G RAN should allow to **integrate LTE-A evolution and novel 5G radio technology on RAN level**, even though this may not be done in all scenarios. Among various options which are being investigated in this context, a **UP aggregation among LTE-A evolution and novel 5G AIVs on PDCP level so far appears to be the most viable option**.

Regarding the overall 5G architecture, METIS-II envisions a **logical split between core network (CN) and RAN**, taking initial orientation in the 3GPP Evolved Packet System (EPS), though it is considered to move some functionalities from CN to RAN, for instance related to paging. It is further assumed that **LTE-A evolution and novel 5G radio share common CN functions, and hence also share a common interface between CN and RAN**.

In the context of a wide range of services and bands, and novel communication scenarios such as flexible time division duplex (TDD), device-to-device (D2D) communications, and moving cells, METIS-II is considering various **paradigm changes related to resource management in 5G**, as for instance the **extension of the notion of a resource beyond conventional radio resources** towards different types of access nodes along with their extensions and soft capabilities of network entities. Furthermore, **context-aware interference management schemes** are envisioned that **minimize the dependency on inter-node interfaces**, as well as **more sophisticated Quality-of-Service (QoS) management** schemes than in legacy systems, for instance with a mechanism in the RAN that **translates air-interface-agnostic to air-interface-specific QoS metrics**.

Finally, METIS-II is considering various changes in 5G compared to legacy systems w.r.t. system access and mobility management procedures. For instance, **the introduction of a novel RRC "Connected Inactive" state** is being discussed, allowing for faster state transition, lower RRC protocol overhead, better device dormancy, and reduced CN/RAN signalling. Moreover, a **lean system design is being evaluated, through a minimization of "always-on" signals such as reference signals and system information and potentially self-contained transmission**. Various further concepts are being investigated, related to **paging, random access channel (RACH) for differentiated access, and a beam-centric system design**, as elaborated in this paper.

Abbreviations

3GPP	3rd Generation Partnership Project
5G	5th Generation of cellular communications
5G PPP	5G Public Private Partnership
AI	Air Interface
AIV	Air Interface Variant
AS	Access Stratum
CN	Core Network
CoMP	Coordinated Multi-Point
CP	Control Plane
D2D	Device-to-Device
DFT	Discrete Fourier Transform
DL	Downlink
DRX	Discontinuous Reception
DTX	Discontinuous Transmission
DUSTM	Differential Unitary Space-Time Modulation
E2E	End-to-End
eMBB	Enhanced Mobile Broadband
eNB	Enhanced Node B
EPS	Evolved Packet System
FBMC	Filterbank Multi-Carrier
FDMA	Frequency Division Multiple Access
F-OFDM	Filtered OFDM
FQAM	Freq. and Quadrature Amplitude Modulation
FSK	Frequency Shift Keying
H-ARQ	Hybrid Automatic Repeat Request
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IP	Internet Protocol
ISI	Inter-Symbol Interference
ITU-R	International Telecommunications Union – Radiocommunication Sector
JT	Joint Transmission
KPI	Key Performance Indicator
LAA	Licensed Assisted Access
LDPC	Low Density Parity Check
LSA	Licensed Shared Access
LTE(-A)	Long Term Evolution (- Advanced)
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MeNB	Master enhanced Node B
METIS	Mobile Enablers for Twenty-Twenty Information Society
MIMO	Multiple Input Multiple Output
mMTC	Massive Machine-Type Communications
mmWave	Bands with carrier frequency beyond 6..30 GHz (different definitions exist)
MTC	Machine-Type Communications

NF	Network Function
NFV	Network Function Virtualization
NGMN	Next Generation Mobile Networks
OFDM	Orthogonal Frequency Division Multiplex
OOB	Out of Band
OQAM	Offset Quadrature Amplitude Modulation
PAPR	Peak-to-Average-Power Ratio
PCEF	Policy and Charging Enforcement Function
PDCP	Packet Data Convergence Protocol
PHY	Physical Layer
PRACH	Physical Random Access Channel
QAM	Quadrature Amplitude Modulation
QCI	Quality of Service (QoS) Class Identifier
QoS	Quality of Service
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RLC	Radio Link Control
RRC	Radio Resource Control
RRH	Remote Radio Head
RRM	Radio Resource Management
RTT	Round Trip Time
S1AP	S1 Application Protocol
SA2	System Architecture 2
SC-FDMA	Single-Carrier Freq. Division Multiple Access
SDN	Software Defined Networking
SINR	Signal-to-Interference-and-Noise Ratio
TAL	Tracking Area List
TDD	Time Division Duplex
TTI	Transmit Time Interval
UE	User Equipment
UF-OFDM	Universal Filtered OFDM
UL	Uplink
uMTC	Ultra-reliable Machine-Type Communications
UP	User Plane
URLLC	Ultra-Reliable and Low-Latency Communications
V2V	Vehicular to Vehicular
V2X	Vehicular to Anything
VNF	Virtual Network Function
WF	Waveform
WRC	World Radio Conference
xMBB	Extreme Mobile Broadband

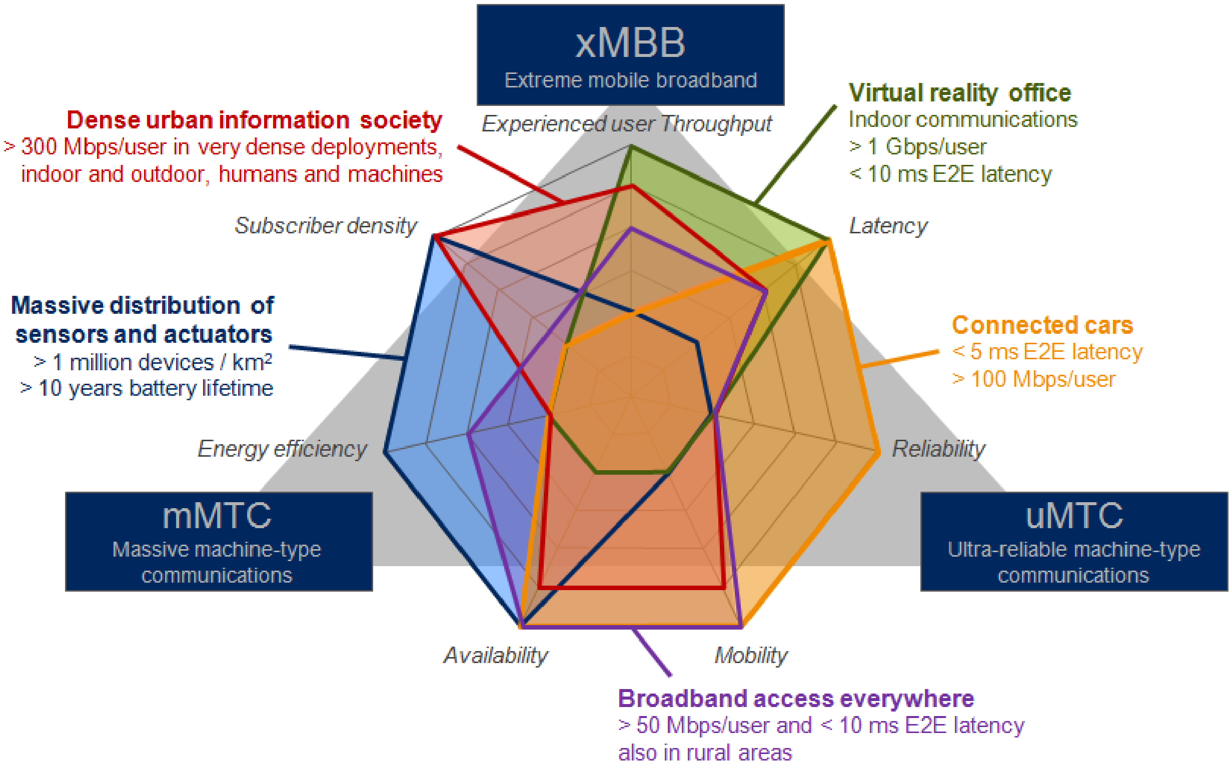


Figure 1. Considered main 5G service types and representative use cases [MET16-D11].

1. Main 5G Service Types, Use Cases and Requirements

There is already a wide consensus on the 5G service landscape, and in particular on the view that 5G will not only be a "business-as-usual" evolution of 4G mobile networks, with new spectrum bands, higher spectral efficiencies and higher peak throughputs, but will also target new services and new business models. These latter are to be developed in close collaboration with vertical industries and imply new requirements and new ways of thinking, building and managing the network. The analysis of the needs and requirements of these verticals has lead the METIS project [MET15-D15], and forums such as NGMN [NGM15] and ITU-R [ITU15] to consider the following three main 5G service types:

- **Extreme Mobile BroadBand (xMBB)**, often also referred to as enhanced mobile broadband (eMBB), requiring both extremely high data rates and low-latency communication in some areas, and reliable broadband access over large coverage areas.
- **Massive Machine-Type Communications (mMTC)**, requiring wireless connectivity for up to tens of billions of network-enabled devices worldwide. Here, scalable connectivity for an increasing number of devices per cell, wide area coverage and deep indoor penetration are key priorities.
- **Ultra-reliable Machine-Type Communications (uMTC)**, often referred to as ultra-reliable and low-latency communications (URLLC), for instance related to vehicle to anything (V2X) communication, industrial control applications, Smart Grid etc.

It goes without saying that considering each service type separately and building a 5G network accordingly, we would likely end up with very different radio access network (RAN) designs and architectures. However, only a common RAN that accommodates all three service types will likely be an economically and environmentally sustainable solution. For this reason, the METIS-II RAN design is performed specifically towards a set of 5G use cases that typically combine multiple service types. More precisely, the project has performed an analysis of the 5G use cases considered by various stakeholders, classified them into families considering the special characteristics of these (e.g., services covered, mobility, and/or number of users, infrastructure, etc.), and has chosen five use cases that are seen as most representative of these different families. These use cases, along with their key requirements, are depicted in Figure 1.

As the use cases considered by different 5G PPP projects are currently being discussed and partially harmonized, as one outcome of the 5G PPP cross-project workshop in September 2015 [MET15-WS], the particular use case selection from Figure 1 may be further updated in the next phases of METIS-II.

2. Key 5G RAN Design Requirements

Based on the previously stated 5G service types and related requirements, METIS-II has derived the following key 5G RAN design requirements:

- The 5G RAN should be able to **scale to extremes** in terms of throughput, the number of devices, the number of connections etc., as denoted in Figure 1. To enable this, it should be able to handle and scale user plane (UP) and control plane (CP) individually. Further considerations on the support of diverse service requirements are provided later in this section.
- The 5G RAN should support the **Network Slicing** vision from NGMN [NGM15], aiming to address the deployment of multiple logical networks as independent business operations on a common physical infrastructure. The implication of Network Slicing on the RAN design is a METIS-II research topic by itself and is also elaborated in more detail later in this section.
- One enabler for the system to handle the diverse service requirements stated before is that the overall network (both RAN and CN) should be **software-configurable**. This means, for instance, that the logical and physical entities to be traversed by CP and UP packets are configurable.
- The 5G RAN should natively and efficiently support **multi-connectivity** (inter-node, inter-air-interface) and **network-controlled D2D** (point-to-point, multi-cast and broadcast).
- The 5G RAN should be designed such that it can maximally **leverage from centralized processing** (e.g. in baseband hosteling scenarios), but also **operate well in the case of distributed base stations** with imperfect backhaul/fronthaul infrastructure, with soft degradation of performance as a function of backhaul quality. More precisely, METIS-II has defined four physical architecture scenarios [MET15-R21], including also a wireless self-backhauling scenario, which should all be supported by any 5G RAN design concepts.
- Some **5G devices should be able to flexibly act as a network node** as well, one example being self-backhauled, possibly nomadic access nodes.
- The 5G RAN design must be **future proof**, i.e. it should enable an efficient introduction of new features and services (e.g. by minimizing the spreading of signals over radio resources and facilitating the introduction of new physical channels) and guarantee backward-compatibility of devices in future releases.

- The 5G RAN design must be **energy efficient** (e.g. by minimizing the amount of always-on signals), enabling efficient network sleeping modes and flexible deployments (where not all the nodes need to send system control signals).
- The 5G RAN must be designed to **operate in a wide spectrum range** with a diverse range of characteristics such as bandwidths and propagation conditions. For higher frequency bands such as mmWave bands, beamforming will become essential, for instance in the form of multiple input multiple output (MIMO).

In addition, it is assumed that the 5G RAN should offer the option to **integrate LTE-A evolution and novel 5G radio technology on RAN level** (though integration needs not always to take place on this level), as motivated and detailed in Section 4.

Design Requirements specifically related to diverse Services and Network Slicing

Beside the aforementioned design requirements, the envisioned **set of diverse services and their requirements** will likely pose the following further requirements on the 5G RAN design:

- **Traffic differentiation**: The RAN should support more sophisticated mechanisms for traffic differentiation than legacy systems in order to be able to treat different services differently and fulfill more stringent QoS requirements. Potential solutions are discussed in Section 8.
- **Resource reuse**: It is expected that many 5G services can be economically supported only if infrastructure resources (e.g. radio resources, but also hardware and software platforms) are extensively reused among different services.

The previous requirements are also applicable to support **Network slicing** in order to efficiently utilize infrastructure resources and provide a differentiated treatment for different slices (i.e. logical networks) regardless of whether they are running the same or different services. The following further requirements have been identified that are more specifically related to Network Slicing:

- **Slice-aware RAN**: Slices (or some abstraction thereof, such as particular groups of flows or bearers) should be visible to the RAN to enable a treatment related to joint key performance indicators (KPIs) concerning all services within a slice or across slices. As an example, all services within one slice may jointly occupy no more than a certain extent of infrastructure resources, but common resources should generally be sharable between different slices.
- **Slice protection**: The RAN should support slice isolation e.g. by providing related slice protection mechanisms so that events within one slice, such as congestion, do not have a negative impact on another slice.
- **Slice management and setup:** The RAN should support efficient management mechanisms, e.g., to efficiently setup and operate slices.
- **Slice-specific network management**: The RAN should allow offering slice-specific network management functions as a service.

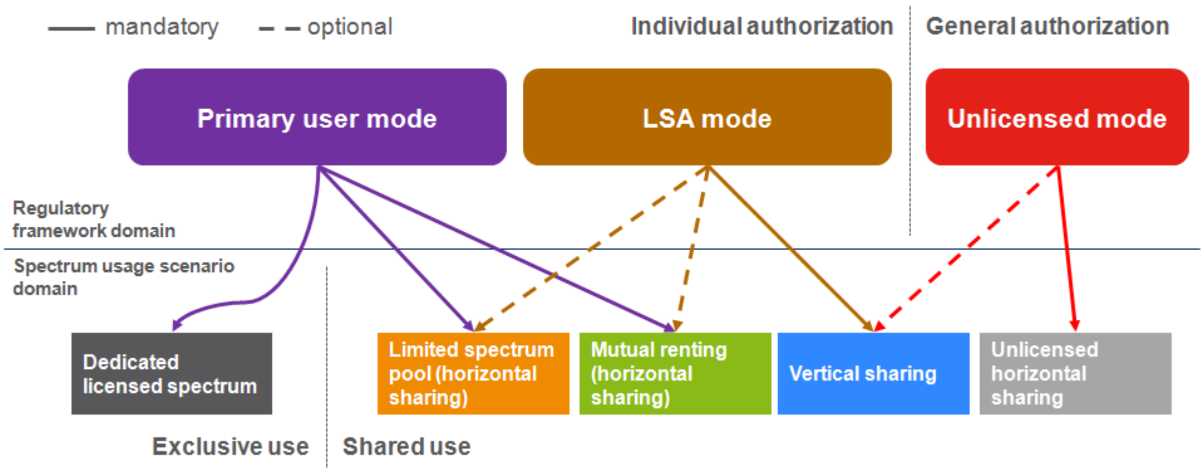


Figure 2. Considered spectrum usage and sharing scenarios [MET15-D54].

3. Air Interface Landscape envisioned for 5G

METIS-II envisions an overall 5G air interface, including evolved LTE-A, designed to operate in a wide range of spectrum bands in order to be capable of addressing the stated diverse services with diverging requirements. For instance, frequencies below 6 GHz are likely most suitable to support mMTC services where coverage is most important, whereas spectrum above 6 GHz is essential to provide the massive capacity demanded by xMBB applications. Analyses also show that contiguous spectrum bands offer advantages with regard to device complexity, signaling, guard bands and interference [MET15-R31]. WRC-15 has agreed that ITU-R will conduct and complete sharing and compatibility studies in a number of frequency bands between 24 GHz and 86 GHz in time for WRC-19.

Radio spectrum can be authorized in two ways: Individual authorization (licensed) and general authorization (license-exempt / unlicensed) within three authorization modes: Primary user mode, Licensed Shared Access (LSA) mode and unlicensed mode. Five basic spectrum usage scenarios can be identified for these authorization modes: dedicated licensed spectrum, limited spectrum pool, mutual renting, vertical sharing and unlicensed horizontal sharing. These modes and the relations between the regulatory and the usage domain which are mandatory (continuous lines) or optional (dotted lines) are depicted in Figure 2. The Licensed Assisted Access (LAA) approach considered for LTE-A is a combination of “dedicated licensed spectrum” and “unlicensed horizontal sharing” by using carrier aggregation. 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.

METIS-II considers the overall 5G air interface[1] to be comprised of multiple so-called AIVs[2], which may for instance be characterized by tailored numerology and/or features for certain frequency ranges, services, or cell types etc. As an example, an AIV tailored towards lower carrier frequencies, large cell sizes and high velocity will likely be based on a physical layer (PHY) designed to be most robust towards

---

[1] An air interface (AI) is here defined as the RAN protocol stack (i.e. PHY / MAC / RLC / PDCP / RRC or 5G equivalents, or subset thereof) and all related functionalities describing the interaction between infrastructure and device and covering **all** services, bands, cell types etc. that are expected to characterize the overall 5G system.

[2] An air interface variant (AIV) is defined in the same way as an air interface, but covers only a **subset** of services, bands, cell types expected to characterize the overall system.

delay spread and Doppler spread, whereas an AIV tailored towards mmWave frequencies and used for 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 (in the order of 1 ms) and/or very high data rates, some new 5G AIVs are expected to use new time-domain structure(s) based on shorter transmit time intervals (TTIs) and a wider bandwidth for radio resource blocks compared to the one specified for LTE-A [MET15-D24]. As another example, an AIV tailored towards a specific service may foresee a specific flavor of hybrid automatic repeat request (H-ARQ), or a specific form of packet data convergence protocol (PDCP) functionality, as shown also in Figure 5 in Section 5. LTE-A and its evolution is likely to play a pivotal role in the overall 5G system, allowing this to maximally leverage the installed base, which will serve as a coverage layer and potentially also as an anchor layer. This will be particularly important in early 5G deployments when novel 5G AIVs may not yet be able to provide full coverage. The exact mechanics of integration of LTE-A evolution into the overall 5G system are an important research topic in METIS-II, with initial views provided in the following section. This being said, METIS-II is also studying 5G standalone operation, as also covered in Section 10.

The number of distinct AIVs that are needed to jointly cover all services, bands and cell types that are expected to characterize the overall 5G system, and the extent of difference between these AIVs, is still under investigation. In this respect, different hypotheses for the overall AIV landscape are being pursued. On one hand, one could consider to use one waveform family such as orthogonal frequency division multiplex (OFDM), and use different numerologies and slight modifications and enhancements of the waveform to cater to different services, bands, cell types. On the other hand, one could consider a co-existence of different waveforms (e.g. OFDM and filterbank multi-carrier, FBMC) that jointly address the overall space of services, bands and cell types.

The exact waveform(s) and PHY numerologies to be used for the novel 5G AIVs are still under investigation, although a set of suitable PHY design considerations that meet the METIS-II design requirements and one or more 5G KPIs have been identified and are summarized in Table 1 on the following page. Key points to be concluded from Table 1 are [MET15-R41]:

- Using the same waveform with different numerologies, or a co-existence of different waveforms (e.g. OFDM / FBMC based solutions), is a key element of many PHY design considerations;
- It is further noted that in some cases certain aspects (like flexible numerology) of proposed PHY technologies could work with both OFDM and FBMC based solutions;
- The trade-off between implementation complexity and performance plays an important role in the selection of PHY technologies, and will be made more challenging by the desire to harmonise functionalities, as discussed in detailed in Section 4;
- Not all PHY design considerations (such as very short TTIs geared towards very high frequencies, e.g., above 6 GHz) are applicable to all bands to the same extent;
- A widespread use of quadrature amplitude modulation (QAM) is noted, except in certain special cases, such as the concepts related to communication with relaxed synchronism (CRS) and communication with non-coherent reception (CNCR) listed in Table 1;
- Most PHY design considerations make use of an LTE-A-like resource grid over time and frequency, but likely with heterogeneous numerology.

It has to be noted that other 5G PPP projects may be investigating further PHY design considerations which are not yet captured in Table 1, but which may be considered in later phases of METIS-II.

Table 1. Overview on PHY design considerations in METIS-II. (for abbreviations please see page 4)

Name	Motivation	Waveform Details	Frame structure	Main features	Frequency bands	Other PHY details
OQAM/FBMC [FB11]	Low OOB emissions, flex. sub-band configurations, higher spectral efficiency, higher robustn. towards time/freq. distortions.	Filtering per subcarrier, time/freq. localized filter design, no cyclic prefix, OQAM, i.e. real-field orthogonality.	Scalable frame design, enabling user- and service-specific adaptations. OQAM posing constraints.	Supports asynchronous transmission, efficient spectrum sharing.	Original design for < 6 GHz. Applicability for above 6 GHz.	Due to OQAM modulation, adaptations are necessary for some MIMO schemes.
QAM/FBMC [YKK+15]	Low OOB emissions, flexible sub-band configurations, higher spectral efficiency, OFDM-compatible.	Separate filters for even- and odd-numbered sub-carrier symbols, no cyclic prefix, QAM, i.e. complex-field orthogon.	Scalable frame design, supporting multiple numerology sets.	Supports asynchronous FDMA transmission, efficient spectrum sharing.		All MIMO schemes supported. QAM modulation, LDPC coding preferred over turbo coding.
P-OFDM (pulse shaped OFDM) [ZSW+15]	Low OOB emissions, flexible sub-band configurations, higher robustness to time/freq. distortions, OFDM-comp.	Filtering per subcarrier, time/freq. localized filter design, QAM, i.e. complex-field orthogon., windowing special case.	Scalable frame design, enabling user- and service-specific adaptations.	Supports asynchronous transmission, efficient spectrum sharing, robust towards phase noise.		All MIMO schemes supported. Modulation and coding like in LTE-A.
F-OFDM [AJM15] **/ UF-OFDM** [WSC14] **based user-centric multi-service AI**	Low OOB emissions, flexible sub-band configurations, OFDM-compatible.	Filtering per sub-band (aggregation of M subcarriers) with steep roll-off, windowing optional.	Scalable frame design, enabling user- and service-specific adaptations.	Supports asynchronous FDMA transmission, efficient spectrum sharing.		
Cyclic prefix OFDM for xMBB in mmWave bands	Support mmWave transmission and adaptive beamforming for hotspots with high rate and low latency.	Cyclic prefix OFDM for ease of implement. and backward compatibility with LTE/LTE-A, windowing optional.	Scalable frame design with LTE-A grid as special case, frame length and symbol duration significantly shortened.	Beam scheduling.	Above 6 GHz with focus on mmWave bands.	Both short and long cyclic prefix supported. QAM modulation and LDPC (preferred over turbo), MIMO support.
Cyclic prefix OFDM for cell-edge and energy efficient application	Increase of cell-edge rate, reduction of PAPR.	FQAM based on OFDM [HSL+14] (other waveforms also possible), windowing optional.	Scalable frame design with LTE-A resource grid as special case.	Tailored for cell edge users and energy constrained services.	Mainly for below 6 GHz.	QAM and LDPC (preferred over turbo), MIMO support.
Harmonized OFDM enhancements	Harmonized cyclic prefix OFDM with scalable numerology for different operating frequencies, low OOB emissions.	Cyclic prefix OFDM for DL/UL/D2D, SC-FDMA for UL, zero-tail SC-FDMA [BTS+13] and OFDM for D2D, windowing and F-OFDM optional.	Support for flexible TDD with scalable and flexible numerology, dynamic TTI sizes, short subframes (~ 0.2 ms).	Multiple numerology sets for scaling in time and freq., multiplex. of different services using flexible spectr. sharing.	Both above and below 6 GHz. Multiple carrier frequencies with target bandwidths of 5 MHz to 2 GHz.	LTE-like modulation up to 256-QAM. New DL and UL control channels embedded within a subframe, MIMO supp.
Communication with relaxed synchronism (CRS)	D2D with relaxed synchronization requirements and MTC with low power budget.	FBMC, UFMC or F-OFDM, windowing optional.	Any frame structure with a low quantity of synchronization signals.	Tailored for D2D and MTC with high data rates.	Any, scalable bandwidth.	MCS-agnostic, MIMO support.
Communication with non-coherent reception (CNCR)	Pilot signal overhead can be drastically reduced for non-coherent reception.	Any saveform that provides negligible ISI.	Any frame structure with a low quantity of pilot signals suitable.	Tailored for V2V and massive MIMO in high-mobility scenarios.	Any, scalable bandwidth.	Modulation: DUSTM and Grassmannian constellations, MIMO support.

4. Air Interface Integration and Protocol Architecture in 5G

A key question related to the 5G system is how the different AIVs, including LTE-A evolution, can be integrated into one overall 5G AI, such that this maximally benefits from the wide landscape of bands, cell types etc., and such that both the complexity of the standard and that of the implementation are minimized, while the performance of individual AIVs is not sacrificed.

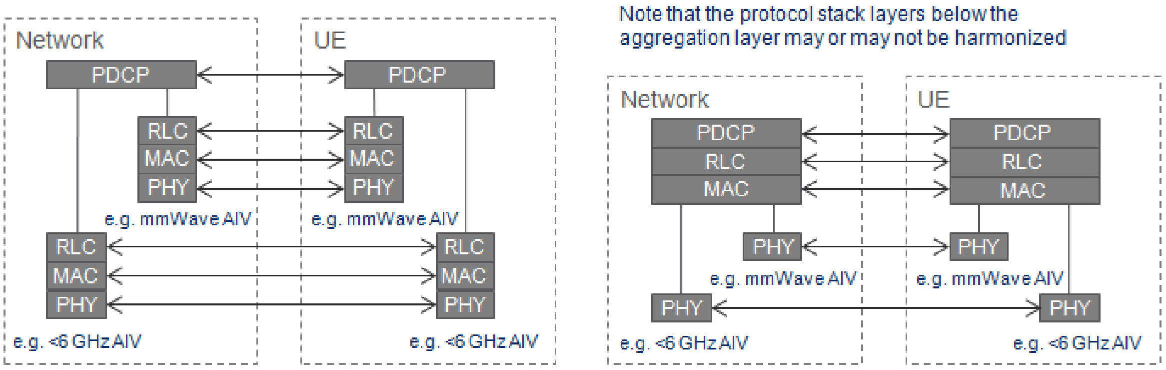
METIS-II envisions that the overall 5G AI should ideally be characterized by a large extent of **protocol harmonization across the AIVs used for different bands, services and cell types.** As an example, user equipment (UE) and network procedures (such as initial access and mobility) should ideally be as similar as possible in the different carriers such as mmWave and lower frequency bands, bearing in mind the existence of technology tailored to each of these bands (such as e.g. narrow band beamforming for the mmWave bands); such specific technologies may in fact be quite disparate, making harmonization (and finding an optimal compromise between potentially highly specialized solutions for specific services and/or frequency bands, and the broader goal to only have one AIV supporting multiple services and bands) highly challenging. A harmonized PHY, as an example, could mean the choice of the same waveform family or reference frame structure for different bands, such that different PHY variants for different bands can be derived from the same framework simply through parameterization (e.g. through adjusting the PHY numerology) or through (de-)activation, addition or removal of certain functionalities, such as, e.g., an additional Discrete Fourier Transform (DFT) in the processing chain for some PHY variant. Note that while a large extent of lower-layer harmonization among novel 5G AIVs may already be considered in their design phase, the lower-layer harmonization of novel AIVs with evolved legacy technology may be challenging or not even desirable: Here, the benefits of harmonization have to be weighed against the potential legacy constraints imposed towards novel air interface technology.

Ultimately, it is clear that the final choice of a particular 5G AIV landscape has to be based on a careful trade-off between the potential benefits of a large extent of harmonization (e.g. from standards and implementation complexity point of view) vs. the potential performance benefits of AIVs that are comparatively more highly specialized for certain services, bands and cell types.

A further key element of integration is the potential capability of one logical instance of a protocol stack layer to **handle multiple bearers or flows related to different devices and service types**. For example, it would likely be desired that a 5G AIV used for lower bands (e.g. < 6 GHz) has PHY and medium access control (MAC) layers that are capable to multiplex all main service types xMBB, uMTC and mMTC.

Another important aspect related to air interface integration is the possibility of **logical aggregation** of the UP protocol stack instances related to multiple instances of the same AIV, as in the case of intra-frequency multi-connectivity, or multiple different AIVs, as for instance in the case of multiple bands. Such aggregation on a certain protocol stack layer means that on and above that layer there is only one single logical protocol stack instance, and hence the higher layers are agnostic with regard to the existence of multiple protocol stack instances or AIVs on the lower layers. Note that aggregation on a certain layer does not require harmonization of the protocol stack layers below (though some constraints may apply, e.g. it may be difficult to aggregate on MAC layer if the different PHY instances underneath have very different frame timing), but it requires harmonization of the AIVs on the layer of aggregation, to the extent that these can actually share one single logical protocol instance. Similarly, one has to consider the protocol stack layer on or above which multiple instances of the same AIV, or multiple different AIVs, may share one **common and single CP instance.**

In this respect, METIS-II has reached the common understanding that is listed in the following pages.



User plane aggregation on PDCP layer
(similar as LTE Rel. 12 dual connectivity)

User plane aggregation on MAC layer
(similar as LTE Rel. 10 carrier aggregation)

Figure 3. Exemplary UP aggregation options for novel 5G AIVs.

Integration among novel 5G Air Interface Variants

Regarding the extent of **harmonization** among novel 5G AIVs, the individual PHY blocks of modulation, channel coding, waveform mapping etc. would ideally be interchangeable through parameterization or (de-) activation of certain features. In other words, there would be a physical or logical implementation that supports easy switching between different variants, and even their multiplexing in time, frequency and/or space. In practice, however, physically separate baseband processing chains (with potentially different / parallel transmission circuitries) may be required for different services, bands or cell types. METIS-II is currently investigating the following three different kinds of harmonization:

1) PHY harmonization of novel 5G AIVs that are potentially based on different waveform families (e.g. OFDM and FBMC based solutions)[3];
2) MAC layer (or higher) harmonisation towards a single specification that supports the usage of different waveform families on PHY layer;
3) Scaling of a single waveform (or waveform family) across all the 5G frequency bands to support all 5G use cases (e.g. variants of OFDM meeting all the requirements of 5G).

The **UP aggregation** among novel 5G AIVs (e.g. for multi-connectivity, and considering the aggregation of links from the same or different AIVs) could take place on MAC, radio link control (RLC) or PDCP level (or their 5G equivalents), given that the AIVs are harmonized on and above the protocol layer of aggregation. MAC layer aggregation has the potential to enable tighter integration features like cross-carrier scheduling, but may be challenging in the context of, e.g., PHY layers with very different frame structure. Also, UP aggregation on MAC or RLC layer would typically only be possible in co-located deployments and/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 benefits of being likely more suitable for distributed deployments with non-ideal backhaul and not requiring the harmonization of the lower layers of the AIVs. Figure 3 shows UP aggregation on PDCP and MAC layer as two potential aggregation options for novel AIVs. Note that the figure applies both to different AIVs or multiple instances of the same AIV (e.g. multi-cell connectivity).

---

[3] As mentioned earlier, full harmonization on PHY layer of different waveform families could prove challenging.

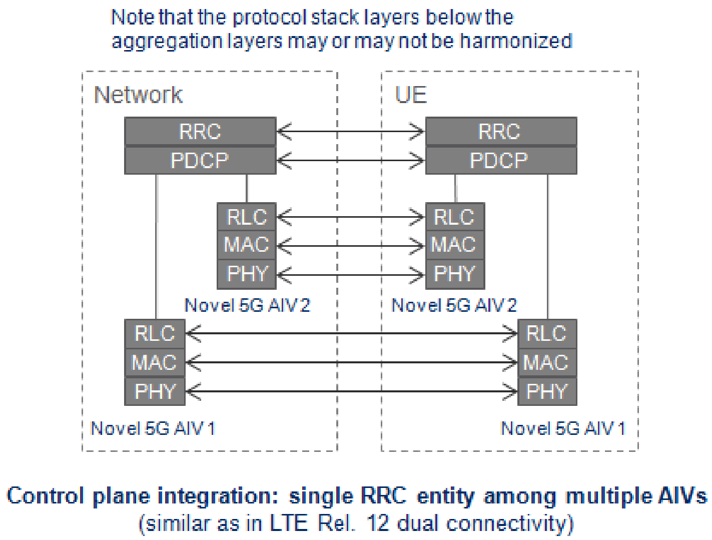


Figure 4. Common CP for novel 5G AIVs realized via a single RRC instance.

Regarding the **CP** support for multi-connectivity (and using LTE parlance for illustration purposes only)**,** a single radio resource control (RRC) protocol instance is envisioned to support the UP aggregation previously described. A single CP protocol instance refers to the case where a single RRC entity at the RAN communicates with a single RRC entity at the UE side, where the single RRC entities are associated to the two (or multiple) links related to the same or different AIVs. This would be equivalent to the LTE Release 12 solution where the single RRC entity at the network side resides at the Master enhanced Node B (MeNB), while the Secondary eNB (SeNB) communicates with the MeNB over the X2 interface to support the configurations of lower layer parameters. This is shown in Figure 4, which again applies equally to different AIVs or multiple instances of the same AIV. Alongside the assumption of a single RRC protocol instance, METIS-II is investigating various forms of CP signaling, such as RRC or CP diversity (where the signaling is provided redundantly over multiple radio legs), or fast CP switching [MET15-R61].

Integration among Novel 5G Air Interface Variants and LTE-A Evolution

In the case of an integration of novel 5G AIVs with LTE-A evolution, additional constraints exist. Despite potential efforts to **harmonize** the lower layers of the new 5G AIVs with LTE-A evolution (e.g. via multiple numerology and/or PHY/MAC structure or waveform choice), limitations may still exist due to the fact that the choices for the new 5G AI are driven by new requirements which may require non-backward compatible protocol stack layers and related functions.

For this reason, the PDCP layer is currently seen as a viable option of **UP aggregation** between LTE-A evolution and novel 5G AIVs [SMR+15], though also aggregation on MAC layer is being investigated. Regarding the **CP integration** between novel 5G AIVs and LTE-A evolution**,** two alternatives are being investigated [MET15-R61]:

- **Single RRC**: a single RRC instance exists at the UE and at the RAN either for the new 5G AI or for the evolved LTE-A;
- **Dual RRC**: two RRC instances exist at the UE and at the RAN, one for the new 5G AI and another for the evolved LTE-A.

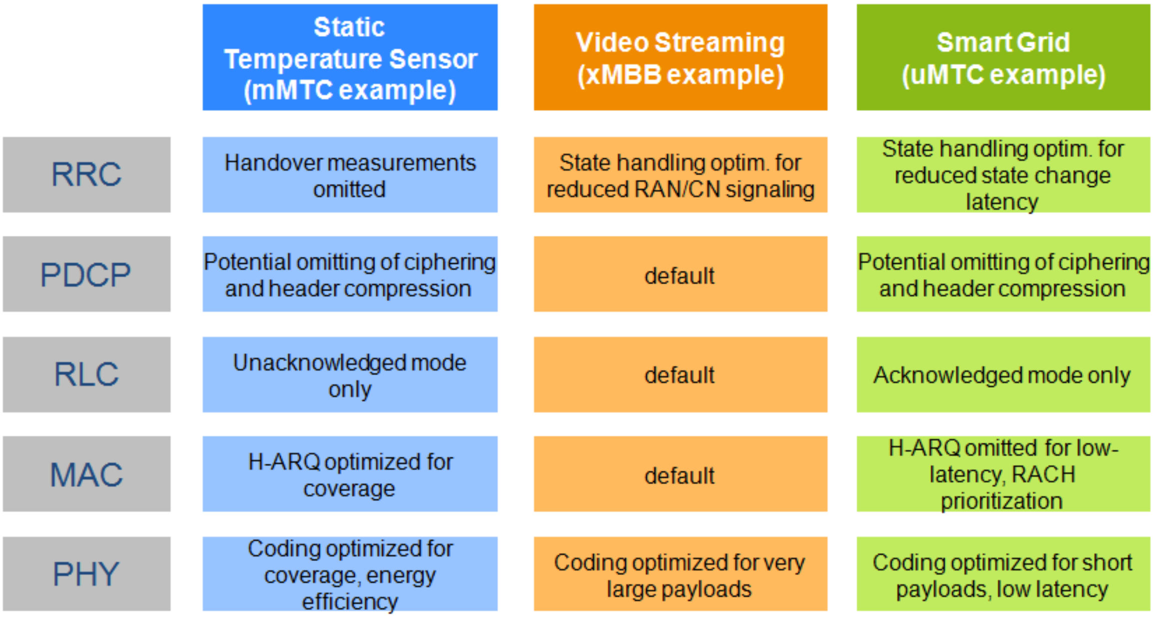
It can be envisioned that these two alternatives may lead to different interaction processes between standardization groups in 3GPP and different ways to capture specifications. A single RRC, for example, could either be specified as an extension of current LTE-A RRC specifications or by creating a new specification having as a starting point the LTE-A RRC. In the case of a dual RRC, two specifications would co-exist and interworking mechanisms should be defined. Regardless of the chosen alternative, a step forward would involve co-locating the RRC protocols of LTE-A and 5G, implying a common Access Stratum (AS) context for the two systems, and also possibly a common RRC connection per two systems. At this stage it should be noted that it is also possible that the LTE-A RRC instance may relay the 5G RRC info in a transparent manner. In other words, the LTE-A RRC instance (residing, e.g., in a MeNB) may not be aware of the 5G RRC details (i.e. the actual content or meaning of the 5G RRC messages) intended for the 5G RRC instance (residing, e.g., in a SeNB). Therefore, while this may be considered as dual-RRC according to the above definition (i.e. two independent RRC entities in the network talking to two RRC entities in the UE), only one RRC entity (MeNB RRC) is providing all RRC signalling to the UE. This effectively means that the difference between single-RRC and dual-RRC operation may be small, with impact on standardization still to be examined.

5. Implication of diverse Services and Network Slicing on the RAN Architecture

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, as shown for the example of three specific services in Figure 5. Note that some functional considerations such as RRC state handling or RACH prioritization will be explained in more detail in Sections 9 and 10, respectively.

Going beyond the support of diverse services, Network Slicing was initially envisioned for the 5G CN and some work has already started in 3GPP to address this area. However, 3GPP does not exclude that also the RAN may need specific functionality to support multiple slices [3GPP15-22891]. METIS-II has achieved some progress in understanding the implications of Network Slicing on the RAN architecture, e.g. yielding the common views that [MET15-R61]:

- 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 e.g. by protection mechanisms 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 fulfilment 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 example)	Video Streaming (xMBB example)	Smart Grid (uMTC example)
RRC	Handover measurements omitted	State handling optim. for reduced RAN/CN signaling	State handling optim. for reduced state change latency
PDCP	Potential omitting of ciphering and header compression	default	Potential omitting of ciphering and header compression
RLC	Unacknowledged mode only	default	Acknowledged mode only
MAC	H-ARQ optimized for coverage	default	H-ARQ omitted for low-latency, RACH prioritization
PHY	Coding optimized for coverage, energy efficiency	Coding optimized for very large payloads	Coding optimized for short payloads, low latency

Figure 5. Examples of potential service-specific selection or tailoring of NFs.

6. Considerations on CN and split between RAN and CN

As a step to fulfill NGMN's vision [NGM15], METIS-II foresees the following for **the majority of the CN and Service Layer functions**:

- CN functions are expected to be deployed as virtual network functions (VNFs) in the 5G timeframe thus running in virtual machines over standard servers, potentially on cloud computing infrastructures, i.e. data centers;
- The design of these CN functions will to some extent explore software defined networking (SDN) principles such as UP/CP split, partially fulfilling the envisioned native architecture of SDN and network function virtualization (NFV);
- VNFs (CP and UP) can be flexibly deployed in different sites in an operator's network depending on requirements related to latency, available transport, processing and storage capacity, etc.;
- Different services or network slices can utilize different CN and Service Layer VNFs which can be deployed at different network sites.

Another important assumption taken in METIS-II is the **logical split between the RAN and CN (and Service Layer) functions**. This is seen as beneficial for the reasons that it:

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

The **METIS-II baseline for the functional split between the RAN and CN is the same as in the EPS** [3GPP15-36300, 3GPP15-23401]. However, alternative CN/RAN splits and/or cross layer optimizations are currently being studied throughout METIS-II. To give one example, the project is investigating the design of RAN-based paging solutions to address densified deployments and a new connected state optimized for inactivity periods, described in later sections, both of which can be seen as a move of functions from the CN to the RAN. It is worth mentioning that the highlighted assumptions have also been assumed in the recently approved SA2 study item about the 5G architecture [3GPP15-153651].

7. Considerations on Network Interfaces

As mentioned before, METIS-II considers the option of integrating LTE-A evolution with novel 5G AIVs on RAN level. While current interworking (for instance between 3G and 4G) relies on inter-CN interfaces basically allowing Internet Protocol (IP) continuity, coverage continuity and load balancing via hard handovers (i.e. always involving CN signaling) and semi-independent resource management for the different access technologies [3GPP15-36300], a RAN level integration would go far beyond the existing interworking between access technologies, fulfilling the vision of what NGMN calls a "5G radio access technology (RAT) family" [NGM15]. An initial step towards an enhanced integration is the assumption that different accesses could benefit from common 5G CN functions and a common CN / RAN interface.

It is envisioned that this **new CN / RAN interface, called herein S1*** will be designed to address the future requirements on the 5G architecture such as the support for:

- End-to-end (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;
- Potentially new UP/CP splits in the 5G CN (designed to follow an SDN/NFV-native architecture);
- A new "Connected Inactive" state, optimized for battery savings but enabling a fast transition to connected state, see also Section 9.

To facilitate the integration of novel 5G AIVs with evolved LTE-A, METIS-II assumes that this new interface will have the S1 interface defined in TS 36.300 [3GPP15-36300] as its baseline.

Within the RAN, METIS-II initially envisions a **new inter-node RAN interface, called herein X2*,** designed to provide new features such as:

- Inter-node multi-connectivity and mobility among the multiple carriers of the new 5G AIVs (such as mmWave and lower frequency variants);
- Inter-node multi-connectivity and mobility among the new 5G AIVs and LTE-A evolution.

Another novel aspect of the X2* interface is that beside supporting distributed deployments it should also leverage from centralized ones, e.g. by enabling the control of multiple synchronous functions (RLC/MAC/PHY) by centralized asynchronous functions (PDPC and/or RRC) that could possibly be implemented in a centralized cloud. To facilitate the envisioned multi-RAT integration with the evolution of existing technologies such as LTE-A, METIS-II assumes that this new interface will have X2, defined in TS 36.300, as its baseline. The METIS-II considerations on network interfaces are depicted in Figure 6.

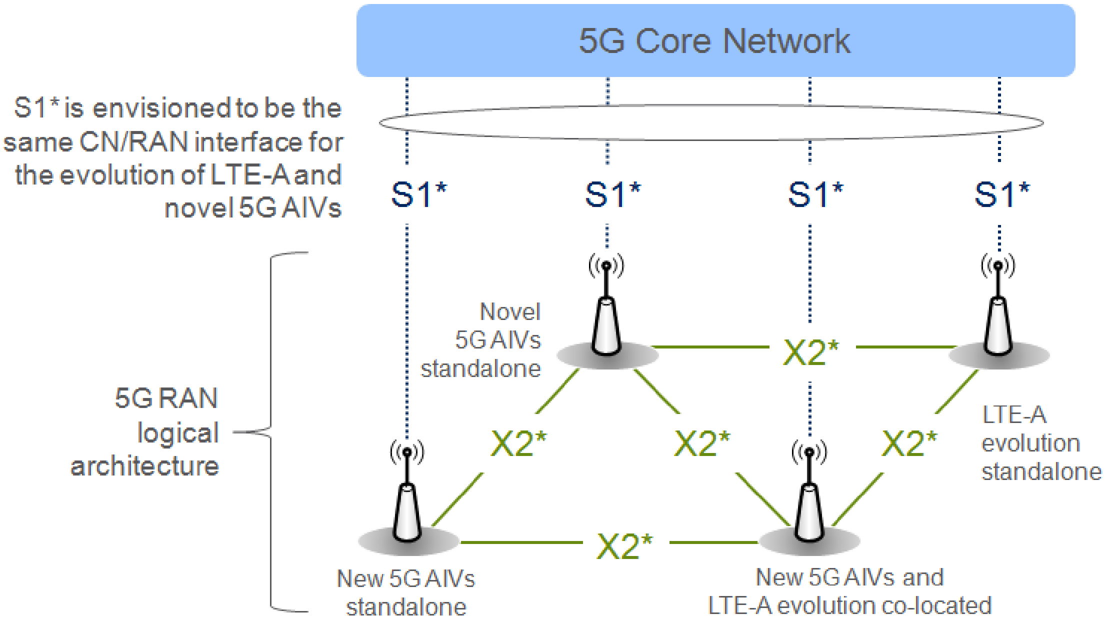


Figure 6. Considerations on network interfaces.

8. Resource Management in 5G

As mentioned earlier, 5G will be characterized by a wide range of services and related requirements, as well as a variety of spectrum usage types and spectrum bands, and will hence likely see a co-existence of different AIVs, such as evolved LTE-A or novel 5G AIVs. Further, it is expected that the network topology will become more dynamic, with a flexible activation and deactivation of static access nodes and their extensions, such as antennas and remote radio heads (RRHs), the incorporation of moving cells such as nomadic nodes into the RAN, and a more widespread and dynamic usage of novel communication modes, such as uni-cast or multi-cast D2D. Finally, the interdependencies between network nodes may increase, as for instance in the context of very dense deployments with time division duplex (TDD) and a flexible usage of uplink (UL) and downlink (DL), or in the context of D2D and in-band self-backhauling. Thus, the forms of interference generated among entities will be different and more demanding to be managed in comparison to legacy networks.

Consequently, a key challenge in 5G is how to dynamically assign services to the most suitable bands, AIVs, communication modes and the radio resources therein to fulfil the diverse requirements of the individual services, as well as constraints applied to groups of services that may be needed in the context of Network Slicing. Further, each network slice can have specific requirements for the radio part (e.g., guaranteed service levels in terms of delay) and restrictions (e.g., wireless transmission only in a dedicated frequency band). In this respect, METIS-II **extends the notion of a resource beyond conventional radio resources**, i.e. strictly related to radio resource management (RRM), towards different types of access nodes along with their extensions, and soft capabilities of network entities. The latter capabilities, such as processing power, memory capacity and energy budget, become an essential set of resources particularly considering the network slices which are operating on the same infrastructure. The conventional use of resources can also span new dimensions, for example unlicensed bands, which can enable additional degrees of freedom provided that proper interference coordination is achieved. This extended notion of resource is visualized in Figure 7.

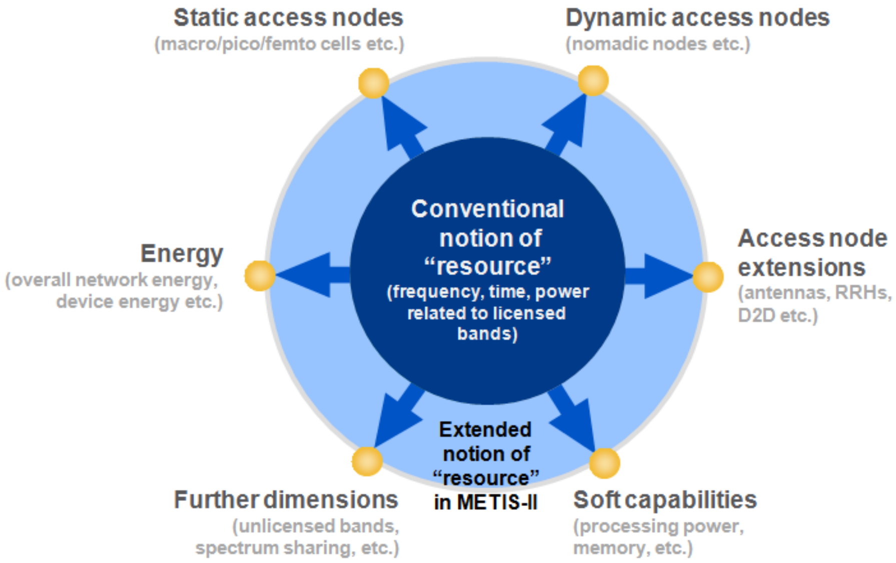


Figure 7. Extended notion of a resource as considered in METIS-II.

In light of the challenges stated before, it appears important to revisit key design principles in current communications systems, and investigate possibly fundamental changes that can be introduced in the context of resource management in 5G. In particular, the following design questions shall be answered:

- Which changes to the X2* interface will be required?
- Which particular PHY/MAC features may be needed to support resource management in 5G?
- Which mechanisms will be required to support QoS management in 5G with relation to bearer, flow management and related granularity?
- How will the Network Slicing concept impact the overall resource management of the system comprising inter-slice and intra-slice resource management schemes?
- What will be the role of different protocol stack layers for resource management in 5G?

An example set of essential resource management related mechanisms is highlighted in the following.

Interference Management

In the envisioned 5G landscape, interference mitigation will be essential for ensuring high capacity and wide coverage of high end-user data rates, as well as for uMTC services, and will have to cope with very dense deployments, flexible TDD and a wide range of device capabilities (e.g. also including devices with sophisticated interference mitigation capabilities). The existing interference management schemes are coupled with varying levels of network coordination. For instance, sophisticated coordination schemes like Coordinated Multi-Point (CoMP) rely heavily on access node synchronization. Therefore, to materialize the theoretical gains in practice, the dependency and impact of the interference management schemes on the X2* shall be minimized. To attain this goal, principles from context-aware management can, e.g., be incorporated to predict users' demands and requirements to decide on appropriate coordination and cooperation mechanisms with lower signaling.

In addition, the interference management schemes may benefit from changes made on the PHY layer. For example, by combining QAM with Frequency-Shift Keying (FSK) referred to as Frequency and QAM (FQAM) [HSL+14] in interfering cells, inter-cell interference can be made non-Gaussian in a victim cell, and the performance of users in low signal-to-interference-and-noise (SINR) regime can be improved provided that the interfering base stations coordinate the assignment of the new modulation type with the target victim cell. Another approach is to apply CP interference management based on the introduction of a spreading and scrambling pre-coding to be applied over the complex baseband symbols that will be transmitted using the assigned time-frequency resources. The coordination of clustered access points can then be achieved by the simple interchange of the appropriate coordination patterns to be used in the spreading and scrambling procedures.

From the RAN architecture perspective, these interference management schemes will mainly imply possible modifications on the signaling procedures, for instance related to inter-node interfaces, rather than necessitating substantial changes in the protocol architecture.

Traffic Steering and RAN Moderation

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 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.

A possible architecture enabling such traffic steering and related Quality of Service (QoS) management is depicted in Figure 8. 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 (de-)activation 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.

A vital criterion for the RAN moderation is the network energy consumption. That is, when the active traffic is not at its peak, certain cells could be dynamically switched off or signals transmitted by the cells could be turned off so that nodes can efficiently apply discontinuous transmission (DTX) mechanisms to reduce the network energy consumption. Please note that even if cells or certain signals are not completely turned off, a lean signaling design as such can likely provide substantial energy savings, as elaborated in Section 10. The exact energy savings obtainable through these different means are still to be evaluated. If cells or certain signals are turned off, it is possible that certain users that are under the coverage of such cell could still be served from the remaining cells, e.g. exploiting joint transmission (JT) in order to improve the quality of the offered service. Subsequently, a joint optimization of the RAN moderation and traffic steering is needed.

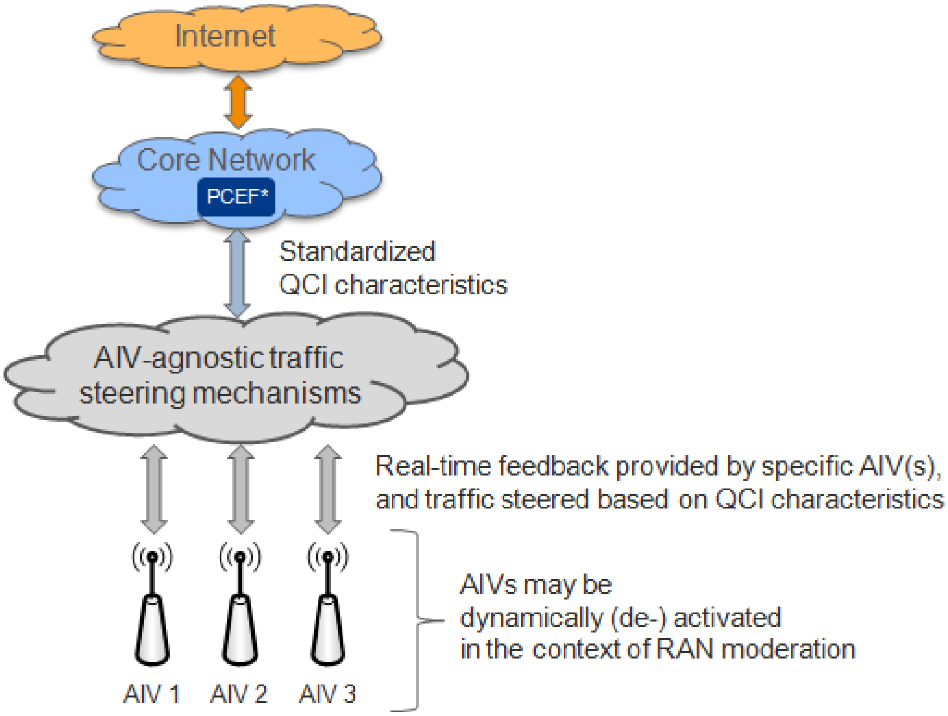


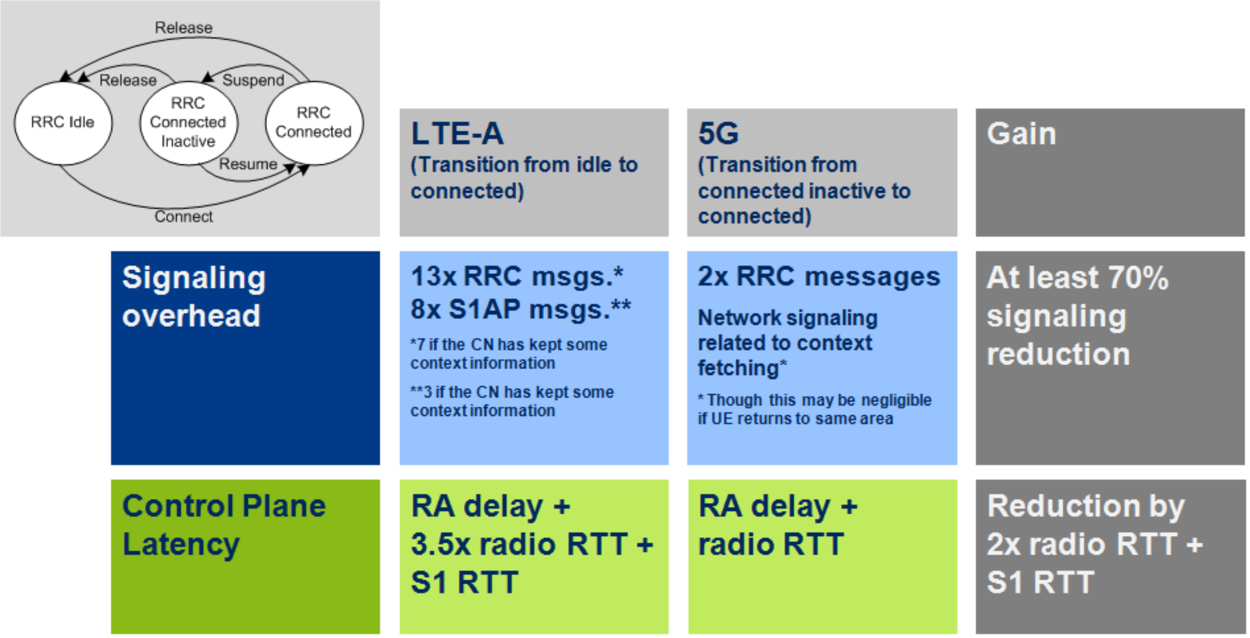
Figure 8. An example for an investigated traffic steering and RAN moderation concept.

UE-Centric Context-Aware Resource Management

Novel fast resource management for LTE-A evolution and novel 5G AIVs will take into account data from services, applications down to PHY layer and will have an impact on network and UE measurements gathering UE associated parameters for controlling resources efficiently (i.e. so-called UE context data). Therefore, the context awareness has clear impact on the UE measurements and the associated implementation complexity. The amount of data to be gathered and the complexity of resource management algorithms need to be treated very carefully between the network performance enhancements they will make available and the load they will impose on both, the radio nodes and the UE in terms of data gathering, signaling, processing and storage. Based on the aforementioned trade-off, possible implications on the related specifications in ITU-R (e.g. context definition) and 3GPP RAN (e.g. measurement definition and measurement gaps) are being investigated.

9. State Handling in 5G

With the trends towards the Internet of Things (IoT) and mMTC it is expected that in 5G there will be even more battery powered UEs (e.g. sensors, baggage tags, etc.). Therefore, battery efficiency and lifetime 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 (either DL or UL) is expected to be more stringent in 5G than the previous mobile generations. This tradeoff between device power efficiency and fast accessibility is often called “UE sleeping problem”.



	LTE-A **(Transition from idle to connected)**	**5G** **(Transition from connected inactive to connected)**	**Gain**
Signaling overhead	**13x RRC msgs.*** **8x S1AP msgs.**** *7 if the CN has kept some context information **3 if the CN has kept some context information	**2x RRC messages** **Network signaling related to context fetching*** * Though this may be negligible if UE returns to same area	**At least 70% signaling reduction**
Control Plane Latency	**RA delay + 3.5x radio RTT + S1 RTT**	**RA delay + radio RTT**	**Reduction by 2x radio RTT + S1 RTT**

S1AP: S1 Application Protocol, RTT: Round Trip Time

Figure 9. Potential RRC state model for the 5G RAN and estimated benefits.

METIS-II is investigating a potential novel state model to address the problem, relying on a novel RRC state called “Connected Inactive” in addition to “Connected” and “Idle”. This novel state explores the principle of not discarding previously exchanged information for inactive UEs, so that UEs in RRC “Connected Inactive” state keep parts of the RAN context. This is valid at least for the semi-static information such as security context, UE capability information, etc. In addition to storing the RAN context, signalling 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 regarding the method to perform the transition to RRC “Connected” [SMS+16]. Figure 9 shows the three proposed states and state transitions as well as estimated benefits.

In the novel model, the state transitions from “Idle” to “Connected” are expected to occur mainly during the first initial access (e.g. when the UE attaches to the network) or as a fallback case (e.g. when the devices and/or network cannot use the previously stored RAN context). Consequently, this transition is not expected to occur as often as in LTE-A. On the other hand, transitions from “Connected Inactive” to “Connected” are expected to occur often and should be optimized as a lightweight and fast transition. 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 reporting of a RAN context ID that can be fetched at the network side (being inspired by the suspend / resume procedure to be defined in LTE-A [3GPP15-23720]). The definition of a novel state provides additional flexibility and further reduction of the CP latency since the transition does neither require the setup of the CN/RAN interface as in [3GPP15-23720] nor the whole RRC connection setup procedure.

The performance benefits are a reduced protocol overhead (in terms of number of RRC messages exchanged during the state transitions) and reduced CP latency. Further details of the proposed three states, state transitions and detailed performance evaluations are currently under investigation.

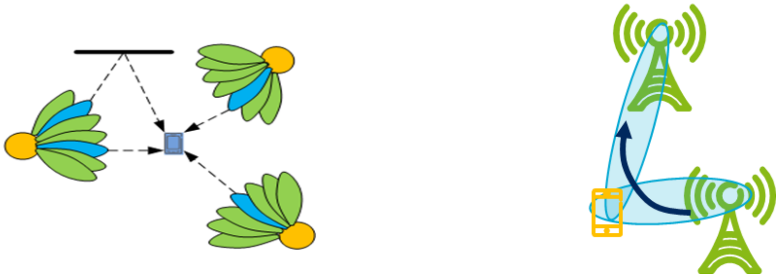


Figure 10. Beamforming will likely impact the access and mobility procedures.

10. Access and Mobility Management in 5G

The wide diversity of services in 5G and the expected number of frequency bands and AIVs involved also suggest rethinking the design of access and mobility management in 5G. Some key design paradigms already identified in METIS-II are explored in the sequel [MET15-R61].

Beam-centric Design

Beamforming, as shown in Figure 10, is an essential aspect of communication in higher frequency bands. While this aspect is of lesser relevance for lower carrier frequencies, it may make sense for the purpose of AI harmonization to generally introduce a beam-based design. That basically means that from a logical perspective, coverage is provided by beams instead of cells. In the case of low frequency AIVs, these beams will likely be as wide as a cell sector, whereas for high frequency variants these beams will be narrow, which will likely impact the way mobility and system access procedures will be designed.

In the case synchronization and reference signals (e.g. for neighbor measurements) are beamformed, the fact that the UE would need to detect a beam that could have a narrow coverage will lead to the need for a direction synchronization procedure in addition to time and frequency synchronization, either for data transmission or acquisition of system information. The same issues are similarly valid for other common channels in the case they would need to be beamformed, such as the Physical Random Access Channel (PRACH), etc. Some work is reported in literature on improving cell search procedures and minimizing scanning delay in this high-gain directional access environment [TN15]. METIS-II is investigating these and other relevant topics in order to incorporate mobility support into the beam-centric design.

METIS-II is also considering solutions where lower frequency layers, possibly including LTE-A evolution, are used in combination with higher frequency layers for coverage detection and synchronization, and for serving as mobility and access anchor layers. However, it is also important to consider in parallel the possibility of UE access on the higher frequency carrier operating in standalone mode.

Lean and Future-Proof Design

In order to maximize energy efficiency, reduce the amount of interference generated by common signals and have a system which is better prepared for beam-based transmissions, METIS-II is investigating different options of how to design a lean system where the amount of broadcasted signals is minimized, as shown in Figure 11.

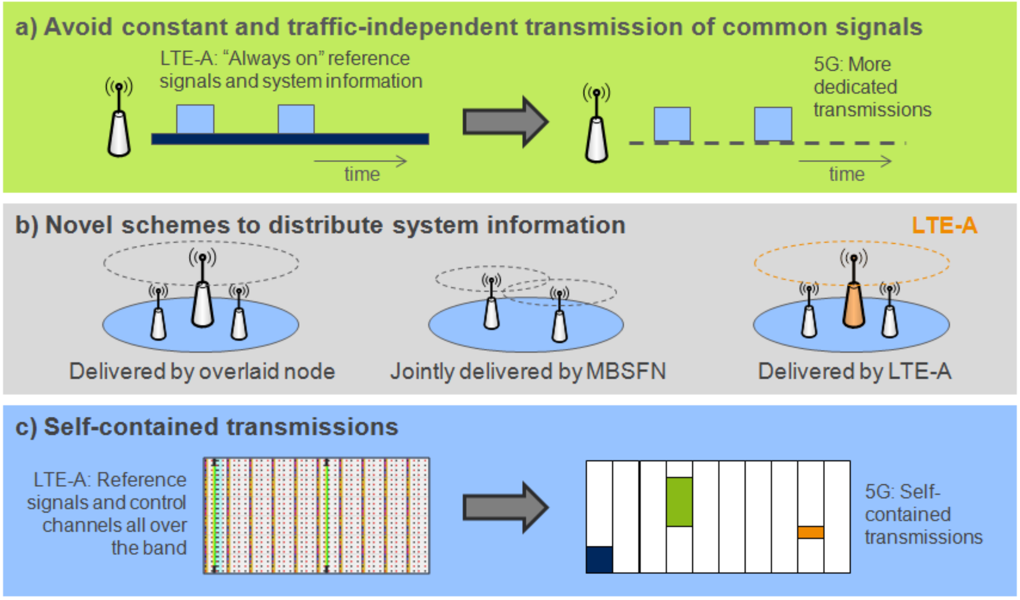


Figure 11. Different means to obtain a lean and future-proof 5G design.

One option, as depicted in Figure 11, a) is to already by design **avoid the constant and traffic-independent transmission of cell-specific reference signals and system information** as it is the case in LTE-A. As a further option, shown in Figure 11, b), METIS-II is investigating **new ways to distribute system information**. One of the alternatives considers a split between UP and system CP allowing some nodes to be dedicated for UP data, i.e. without the need to transmit system information. In another alternative, LTE-A evolution coverage may be used to distribute system information related to multiple AIVs. Finally, lessons learnt from LTE-A have shown that a future-proof air interface design needs to allow the introduction of new physical channels. **Self-contained transmissions** could provide both future proofness and be more adapted to a beam-based system. In the case of self-contained transmissions, reference signals are not spread all over the bandwidth such as control channels and/or reference signals used for neighbor cell measurements. Instead, reference signals are transmitted jointly with the payload, minimizing the overhead and interference as illustrated in Figure 11, c).

In current systems, broadcasted signals are essential for system access (such as synchronization sequences and system information) and mobility (such as reference signals for connection control and neighbor link measurements). METIS-II is currently investigating mechanisms to allow these signals to be most of the time transmitted on demand (i.e. only when needed) or transmitted via dedicated signaling (i.e. avoiding the transmission when there is no active traffic). This may also lead to improvements in terms of the amount of interference generated due to common signals.

RACH-related considerations for differentiated access

Different 5G services may have different requirements in terms of system access, e.g., when a device needs to transit from "Connected Inactive" to "Connected" state in a very fast fashion, for instance in some V2X use cases where a latency on the order of milliseconds may be required. In order to provide an efficient prioritization mechanism for delay-sensitive services (not relying on the assignment of dedicated preambles), METIS-II is currently investigating random access solutions to provide some level of access differentiation per service, taking their accessibility requirements into account.

In the currently proposed solution, random access requests associated with delay sensitive services could be configured to apply a combination of preamble signatures at a given random access time slot. Specifically, a higher priority device (i.e. UE2) sends two preambles PA1 and PA2, and even though a collision may occur with the preamble transmitted by a lower priority device (i.e. UE1), the network will detect the high priority device and grand transmission resources to it. The mentioned approach would enable requests with more stringent delay requirements to have higher priority, since combinations of preambles can always be identified by the receiver. In this way, requests with higher priority could possibly be immune from the collisions with and the re-transmissions related to other requests.

RAN-based Paging

Due to the expected massive number of devices and more small cells (e.g. due to higher frequencies and denser deployments) in 5G, paging may significantly increase the load on both the air interface and the CN/RAN interface, thus making it important to have new solutions for efficient paging and UE location tracking. One potential way forward could be the introduction of such functionalities in the RAN.

One approach METIS-II is currently investigating is to use a hierarchical location tracking where the CN tracks the registration of a UE on a group of RAN locations, and the RAN tracks a higher granularity of the UE location. The RAN location of the UE could be possibly updated to the RAN for “Connected Inactive” UEs. This update could be done via a light signalling procedure which is terminated in the RAN. The details of such a signalling procedure are being investigated along security handling mechanisms for the messages exchanged in the signalling procedure. Using the considered hierarchical location tracking approach, the paging and part of the mobility functionalities of the CN would be moved to the RAN.

One of the advantages of such an approach, as depicted in Figure 12, is the potential to significantly reduce the paging load on the AI using a higher granularity paging, especially in the case of semi-static UEs where these RAN updates would not be so frequent (and basically the network gets the UE location for free). The RAN level location of UEs can be used by the network in different beneficial ways such as:

- The RAN can provide the UE with a dedicated RACH preamble to minimize the delay to reach UEs with stringent delay requirements;
- The CN can proactively forward data to the RAN location where a UE is known to be camping for a fast first packet.

For high mobility UEs (especially when assuming small cell deployments), this approach may create some drawbacks in terms of the UL signalling generated by the UEs towards the network and power consumption in the “Connected Inactive” state. In this respect, METIS-II is investigating the impact of different mobility patterns for that approach.

Another advantage of the proposed scheme is the potential to significantly reduce the paging load on the CN/RAN interface. Assuming that in “Connected Inactive” state the RAN keeps the S1* connection alive, the CN is not aware whether the UE state in RAN is active or inactive. The CN forwards incoming packets to a given S1* termination for that UE at a given RAN logical node (denoted herein 5G-eNB). The “Connected Inactive” state is optimized for the case that the UE is semi-static and remains in a certain location controlled by the same 5G-eNB. When the UE moves and notifies a new 5G-eNB about the new location, the CN forwards packets to the right RAN termination point. If the location is not updated or the UE moves within the tracking area, the CN simply forwards packets to the last RAN node where the UE was active. That node could also act as a mobility anchor and needs to be aware of the UE location in the RAN, or leave the mobility anchor responsibility to the new RAN node the UE has moved to.

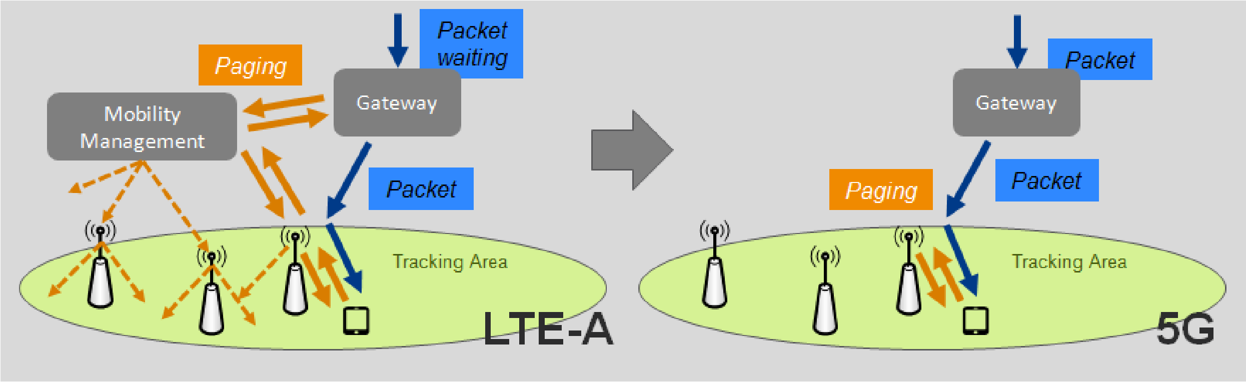


Figure 12. RAN-based paging in 5G.

Another concept for signalling optimization in mobility management is related to group based design, by exploiting grouping and D2D connections for handling certain control operations. Groups of devices will be available in the future for several purposes, e.g. device tracking, car to car clustering, etc. In such cases, only one device may perform location updates on behalf of the group. D2D communication or secondary interfaces (e.g., IEEE 802.11p, IEEE 802.15, etc.) may be used to form groups and to exchange information within these. The group based approach may reduce the number of location area updates that are performed for the overall group, thus enabling performing updates more often, and consequently reducing the size of the paging areas. For concise groups, this approach will facilitate performing location updates even on a per cell basis and thus targeted paging in the cells where the UE is located, thus significantly reducing the latency for locating a user.

11. Outlook

This White Paper has captured the preliminary views and initial considerations on 5G RAN architecture and functional design of the METIS-II partners after the first 8 months of the project. The 5G RAN design in METIS-II will naturally be further progressed in the coming months, based also on the discussion with other 5G PPP projects through, e.g., jointly organized workshops. The next update on the 5G RAN design considerations from METIS-II will be provided in the time frame April – June 2016, when various public deliverables will be issued. Please find further information at https://metis-ii.5g-ppp.eu/.

12. Acknowledgement

This work has been performed in the framework of the Horizon 2020 project METIS-II, which is co-funded by the European Commission.

References

[5GPPP] 5th Generation Public-Private-Partnership, see http://5g-ppp.eu

[3GPP15-153651] 3GPP S2-153651, “Study on Architecture for Next Generation System”, Oct. 2015

[3GPP15-22891] 3GPP TR 22.891, “Study on New Services and Markets Technology Enablers”, Nov. 2015

[3GPP15-23401] 3GPP TS 23.401, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”, Dec. 2015

[3GPP15-36300] 3GPP TS 36.300, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2”, Nov. 2015

[3GPP15-23720] 3GPP TR 23.720, “Architecture enhancements for Cellular Internet of Things (Release 13)”, Nov. 2015

[AJM15] J. Abdoli, M. Jia and J. Ma, “Filtered OFDM: A new waveform for future wireless systems", IEEE SPAWC 2015, Stockholm, Sweden, June 2015

[BTS+13] G. Berardinelli, F. M. L. Tavares, T. B. Sorensen, P. Mogensen and K. Pajukoski, “Zero-tail DFT-spread-OFDM signals”, Globecom 2013 Workshops, Atlanta, USA, Dec. 2013

[EC15-AG] European Commision, HORIZON 2020 – WORK PROGRAMME 2014-2015, General Annex G, see http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

[FB11] B. Farhang-Boroujeny, “OFDM Versus Filter Bank Multicarrier”, IEEE Signal Processing Magazine, vol.28, no.3, pp.92-112, May 2011

[HSL+14] S. Hong, M. Sagong, C. Lim, S. Cho, K. Cheun and K. Yang, “Frequency and Quadrature-Amplitude Modulation for Downlink Cellular OFDMA Networks”, IEEE Journal on Selected Areas in Comms., vol. 32, no. 6, pp. 1256–1267, June 2014

[ITU15] 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

[METIS] METIS-II Project (“Mobile and wireless communications Enablers for the Twenty-twenty Information Society II”), see https://metis-ii.5g-ppp.eu/

[MET15-D15] ICT-317669 METIS, Deliverable 1.5, “Updated scenarios, requirements and KPIs for 5G mobile and wireless system with recommendations for future investigations”, April 2015

[MET15-D24] ICT-317669 METIS, Deliverable 2.4, “Proposed solutions for new radio access”, Feb. 2015

[MET15-D54] ICT-317669 METIS, Deliverable 5.4, “Future spectrum system concept”, April 2015

[MET15-R21] ICT-671680 METIS-II, Report 2.1, “RAN Design Guidelines and Methodology”, Sept. 2015

[MET15-R31] ICT-671680 METIS-II, Report 3.1, “Preliminary spectrum scenarios and justification for WRC Agenda Item for 5G bands above 6 GHz”, Sept. 2015

[MET15-R41] ICT-671680 METIS-II, Report 4.1, “Preliminary air interface analysis and user plane design considerations”, Nov. 2015

[MET15-R61] ICT-671680 METIS-II, Report 6.1, “Preliminary asynchronous control functions and overall control plane considerations”, Nov. 2015

[MET15-WS] ICT-671680 METIS-II, Press Release, “EU project METIS-II hosts workshop on 5G system with 5G PPP projects”, see https://metis-ii.5g-ppp.eu/wp-content/uploads/METIS-II-Press-release-Sept-2015.pdf

[MET16-D11] ICT-671680 METIS-II, Deliverable D 1.1, “Refined scenarios and requirements, consolidated use cases, and qualitative techno-economic feasibility assessment”, Jan. 2016

[NGM15] NGMN Alliance, “NGMN 5G White paper”, Feb. 2015

[TN15] M. Tesanovic and M. Nekovee, “mmWave-based mobile access for 5G: key challenges and projected standards and regulatory roadmap”, IEEE GLOBECOM 2015, San Diego, December 2015

[SMR+15] I. Da Silva, G. Mildh, J. Rune et al., “Tight Integration of New 5G Air Interface and LTE to Fulfill 5G Requirements”, IEEE VTC Spring 2015, Glasgow, May 2015

[SMS+16] I. Da Silva, G. Mildh, M. Säily and S. Hailu, “A novel state model for 5G radio access networks“, IEEE ICC 2016, 5G RAN Design Workshop, Kuala Lumpur, May 2016

[WSC14] T. Wild, F. Schaich and Y. Chen, “5G Air Interface Design based on Universal Filtered (UF-)OFDM”, International Conference on Digital Signal Processing (DSP), Hong Kong, Aug. 2014

[YKK+15] Y. H. Yun, C. Kim, K. Kim, Z. Ho, B. Lee and J.-Y. Seol, "A new waveform enabling enhanced QAM-FBMC systems", IEEE SPAWC 2015, Stockholm, Sweden, June 2015

[ZSW+15] Z. Zhao, M. Schellmann, Q. Wang, X. Gong et al., “Pulse shaped OFDM for asynchronous uplink access“, Asilomar Conference on Signals, Systems and Computers, Monterey, USA, Nov. 2015