5G PPP mmMAGIC

Architectural aspects of mm-wave radio access integration with 5G ecosystem

Editor:

Krystian Safjan - *Nokia Networks*

Authors:

Hardy Halbauer - *Alcatel-Lucent*

Patrik Rugeland, - *Ericsson*

Yilin Li - *Huawei*

Joerg Widmer - *IMDEA*

Marcin Rybakowski, Krystian Safjan, Arnesh Vijay - *Nokia Networks*

Isabelle Siaud, Anne-Marie Ulmer-Moll - *Orange*

David Gutierrez Estevez, Mehrdad Shariat - *Samsung*

Maciej Soszka - *Technische Universität Dresden*

Date: 2016-04-14 Version: 1.0

Executive Summary

The following white paper discusses a range of architectural aspects that are crucial for the integration of mm-wave Radio Access Technology (RAT) in 5G mobile networks. While radio spectrum above 6 GHz offers contiguous high bandwidth resources for high capacity radio links; it gives rise to new challenges in terms of user mobility and reliability due to the harshness of the mobile radio environment, necessitating the deployment of highly directional links. To overcome these challenges, we propose architectural solutions based on tight integration of mm-wave RAT with RAT(s) operating in frequencies below 6 GHz with the intention to diversify the control plane and data split, and the duplication of the user plane.

The paper outlines initial architectural concepts envisioned for 5G mm-wave systems designed to operate in frequencies between 6-100 GHz, and discuss some crucial aspects of its integration within the 5G networks. The proposed concepts include: i) network slicing to address needs of 5G use cases with highly divergent requirements, ii) multi-connectivity with multiple mm-wave base stations and as a way to integrate multiple RATs; iii) mobility related aspects such as mm-wave cell clustering to make small scale mobility of the UE transparent to the CN, and the introduction of a novel inactive RRC state, iv) multi-RAT multi-layer management scheme which introduces abstraction at different layers of the protocol stack for flexible and scalable multi-RAT management, vi) control and user plane aspects for low-band integration, and finally vii) architectural aspects of self-backhauling.

Table of Contents

List of Acronyms and Abbreviations

4G	Fourth generation
5G-NB	5G NodeB
5G PPP	The 5G Infrastructure Public-Private Partnership
ACK	Acknowledgement
AP	Access Point
BS	Base Station
CH	Cluster Head
CN	Core Network
CoMP	Coordinated Multi-Point
CP	Cyclic Prefix
CQI	Channel Quality Indication
DL	Downlink
HARQ	Hybrid Automatic Repeat request
LDPC	Low Density Parity Check
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
MAC	Medium Access Control
MBB	Mobile Broadband
MC	Multi-Connectivity
MCS	Modulation and Coding Scheme
MRC	Maximal Ratio Combining
MTC	Machine Type Communication
NACK	Negative Acknowledgement
NFV	Network Function Virtualization
NGMN	Next Generation Mobile Networks
PDCP	Packet Data Convergence Protocol
PHY	Physical layer
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
sBH	Self-backhaul
SDN	Software Defined Network
SFN	Single Frequency Network
TA	Tracking Area
TAU	Tracking Area Update
TTI	Transmission Time Interval
UDN	Ultra Dense Network
UE	User Equipment
UL	Uplink
UP	User Plane
WLAN	Wireless Local Area Network

1 Introduction

It has been widely accepted by the technical community, that the radio spectrum above 6 GHz offers diverse opportunities for contiguous high bandwidth resources, promising high capacity radio links for access and backhaul communication. However, the expansion into higher frequency bands gives rise to new challenges in terms of user mobility and interference, requiring novel intra- and inter-RAT cooperation schemes for network integration. Furthermore, it is anticipated that mm-wave technology can support other advanced technologies in the future 5G systems.

The main scope of this white paper is to outline some of the initial concepts envisioned for 5G mm-wave architecture in 6–100 GHz frequency band, and discuss some crucial aspects of its integration within the 5G networks. Further ahead, we look into the identified challenges for 5G systems when compared to LTE-A technology, and propose tightly integrated solutions to address challenges in three key areas, namely: performance, ultra-dense networks (network management and backhaul provision), and architecture flexibility. The proposed concepts include:

- Network slicing to address needs of 5G use cases with highly divergent requirements by defining multiple logical network within the same physical infrastructure.
- Multi-connectivity with multiple mm-wave base stations and as a way to integrate multiple RATs. Multi-connectivity between mm-wave and <6 GHz RATs addressing challenges related to performance requirements in terms of capacity, coverage and reliability. One out of the possible practical realizations might be an evolution of LTE-A Dual Connectivity.
- Mobility related aspects such as mm-wave cell clustering to make small scale mobility of the UE transparent to the CN. And the introduction of a novel RRC connected inactive state to minimize heavy signalling procedures, when the user's mobile broadband data traffic consists several infrequent small data bursts interspersed by relatively long waiting periods, e.g. during web browsing or short video or music streaming.
- Multi-RAT multi-layer management scheme which introduces abstraction at different layers of the protocol stack for flexible and scalable multi-RAT management that can be jointly deployed with multi-connectivity.
- User and control plane aspects for mm-wave and low-band integration, like: efficient control signalling which minimizes control overhead or RRC diversity to improve reliability.
- Selection of initial access control information which in multi-connectivity case should be sent over low-band (<6 GHz) system.
- Architectural aspects of single- and multi-hop self-backhauling e.g. resource allocation for access and backhaul, and multi-hop routing.

2 Addressed Use Cases

To derive the necessary network architecture and identify the requirements in various 5G applications, we define a number of use cases. Each use case comes with a set of challenges that needs to be analysed and solutions proposed. The selected use cases are as shown in Figure 1.

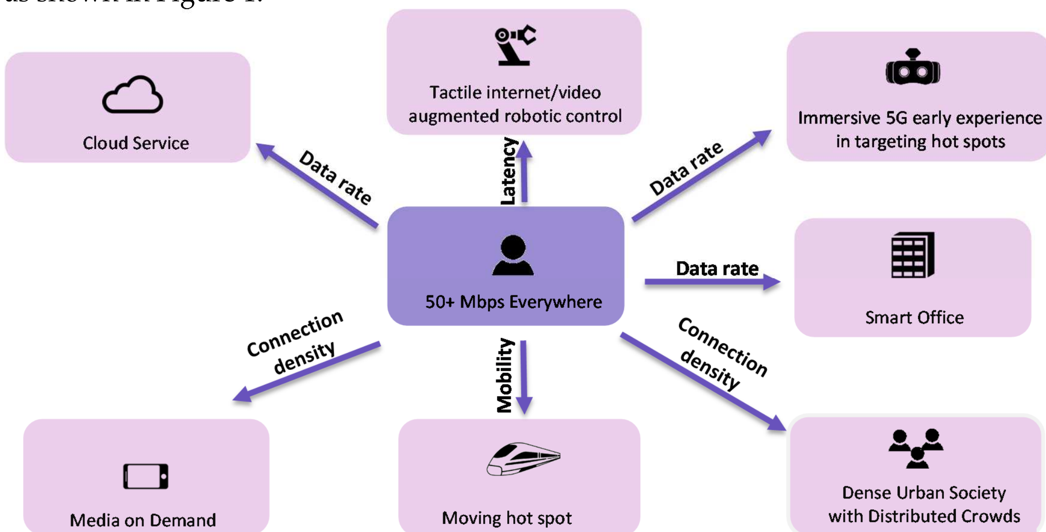

Figure 1 Use cases

The baseline use case in 5G system is defined as **"50+ Mbps Everywhere".** This use case imposes a stringent requirement of above 50 Mbps in every scenario, including cell edges, whenever required. This requirement is considered crucial and forms a basis for other use cases which are extended in various dimensions as depicted in Figure 1. Consequently, the high data rate in both DL and UL is also expected for the "**Cloud Service**" use case. This case aims at enhancing the customization for users and guarantees high availability of "anytime and anywhere" services. Next, to meet high data rate demands in crowded cities, the "**Dense Urban Society with Distributed Crowds**" use case is selected. The goal here is to provide high capacity, robustness, low latency and high data rate within distributed crowd. In the "**Smart Office**" use case, the network is expected to provide high data rate in DL and UL transmissions. Therein, it is assumed to cover large number of devices with limited mobility. Moreover, the network should serve high cell edge throughput and low delays. To satisfy the 5G subscribers, the other use case of **"Immersive 5G early experience in targeting hot spots"** is considered. This use case stresses upon the high data rate in outdoor scenario with stationary hot spots. For this, beamforming, small-cell deployment and low band support can be deployed. The support of low-band frequencies can be beneficial to the **"Moving hot spot"** use case**.** In this case, the UEs are expected to move with high speed up to 500 km/h; whereby, relay type network architecture can be considered. Some cases, where the baseline use case of **"50+ Mbps Everywhere"** is somewhat relaxed is for "**Media on Demand**" and "**Tactile internet**". The "**Media on Demand**" case imposes desired throughput demands on a large number of simultaneous users, which entails the requirement for overall high capacity networks. Here, the data transmission is expected to be robust for long periods of connection. While on the other hand, the anywhere remote control of virtual and real objects in real-time is enabled by the "**Tactile internet**" use case. This comes together with high reliability and low latency requirements. In some applications, which are known as mission-critical applications, all requirements have to be met, whereas other applications are more relaxed in at least one of the requirements.

Hence, to meet the requirements of various use cases, a set of novel solutions is proposed. These solutions are expected to enable the 5G system in the near future.

3 Network Slicing

One important aspect of the 5G ecosystem will be the concept of network slicing. This was initially proposed for the 5G core network (CN) [E14] and was extended by NGMN to include the radio access network (RAN), defined as "An end-to-end (E2E) network slice" [NGMN15]. It is envisioned that network slicing will primarily be business driven, where each slice will support one or more communication service, as shown in Figure 2; possibly with a specific way of handling control plane and user plane for these services. Each network slice must be capable of being managed by the end customer or slice owner as an independent network. Instead of deploying separate network infrastructure for each slice, the slices are realized on a common physical infrastructure such as hardware, processing, storage, backhaul, spectrum resources, etc. using a "virtual network". Even though the mm-wave RAT initially targets primarily mobile broadband (MBB) services for enhanced coverage and throughput, this does not preclude that some slices may support e.g. Machine Type Communication (MTC) services and the architecture must be capable of supporting this.

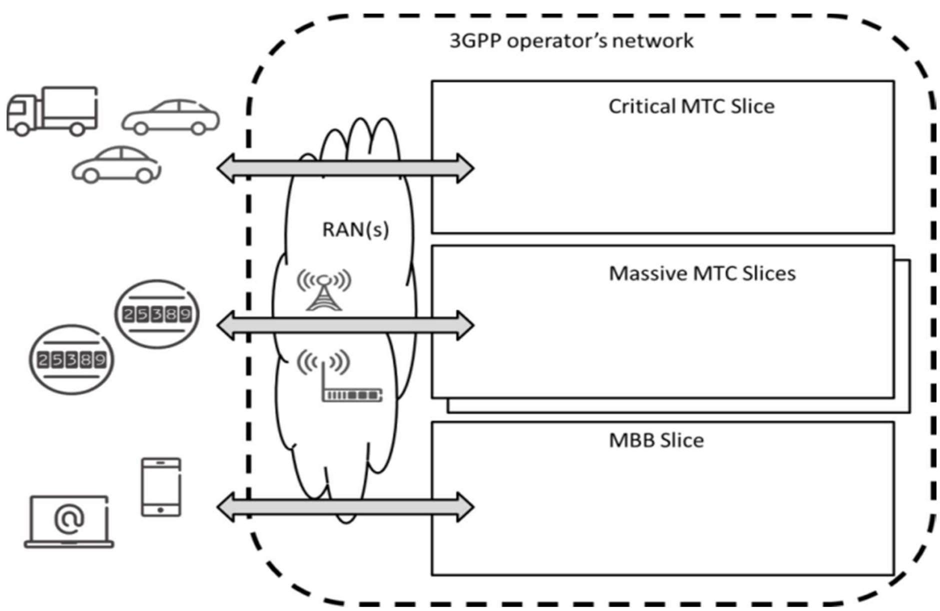

Figure 2 Concept of network slicing

As different parallel network slices may target a variety of use cases with very diverging Quality-of-Service (QoS) requirements there are few baseline assumption that can be made on the 5G architecture and RAN design.

- **Sharing** of most RAN resources between multiple slices is assumed as default
- **Differentiation** of traffic between slices is enabled by 5G architecture mechanisms
- **Visibility** of slices to the 5G RAN is required to be able to apply slice differentiation
- **Protection** of slices by the 5G RAN is required to minimize inter-slice effects
- **Management** of independent slices should be supported by the 5G architecture

The alternative to network slicing would be to deploy independent physical networks which would drive costs. To optimize the resource utilization, it is important that the resources are shared as much as possible; but, in order to assure slice protection, it may be necessary to temporarily provide dedicated physical resources to certain slices e.g. critical communication to assure the fulfilment of the QoS requirements. However, as the traffic demand will invariably fluctuate over time; the dedicated resources should be released from the slice as soon as they are no longer needed.
To summarize, network slicing is about addressing needs of 5G use cases with highly divergent requirements. By operating on logical instead of physical network elements, slicing provides architecture flexibility and future proofness; and since 5G RAN should be aware of the traffic flow to slice association, it is important to design RAN part of mm-wave RAT to be open for network slicing and slice-specific architecture optimization instances.

4 Multi-Connectivity

The propagation properties of the mm-waves are expected to be more challenging in terms of propagation loss, reduced diffraction, and increased outdoor-to-indoor penetration loss. It will be necessary to employ multi-connectivity (MC) to ensure ubiquitous and persistent coverage for seamless end-user experience. In general multi-connectivity refers to UEs connecting simultaneously to multiple links. These links can be provided by a single or multiple network nodes (collocated/non-collocated), on a single or multiple carrier frequencies (intra-/inter-frequency), using a single or multiple radio access technologies (intra-/inter-RAT).

Designing MC solutions for 5G requires taking several decisions on RAN architecture level that are affecting both Control- (e.g. enabling RRC diversity or control information split between <6 GHz and mm-wave RATs) and User-Plane design (data duplication or split and if split, then what are envisioned splits of service flows)

The intra-frequency MC solutions comprise e.g. single frequency networks, joint transmission (JT) Coordinated Multi-Point (CoMP) or coordinated scheduling/beamforming CoMP. JT CoMP requires strict time synchronization between the nodes which implies an ideal fronthaul connection to the central unit in order to enable coherent reception. Remaining two solutions have more relaxed requirements for synchronization: SFN on CP level and CS/CB CoMP on DL fame timing level.

The inter-frequency MC considers several options comprising e.g. carrier aggregation (CA), dual connectivity (DC), as well as MC in-between the mm-wave RAT and low band RATs, e.g. LTE-A or a new 5G RAT (further discussed in Section 7).

CA as defined for LTE Rel-10, integrates multiple carriers from a single node at the MAC layer and could be extended to be used for the mm-wave RAT. If CA should be possible between the mm-wave RAT and low frequency RATs, e.g. LTE-A, then the numerology and signal format on PHY layer must be aligned in an appropriate way and cannot be designed independently.

DC as defined by LTE Rel-12 [3GPP TS 36.300] could also be extended to the mm-wave RAT, where the data split is done in the PDCP layer. As the different links maintain separate PHY and MAC layers, it will be possible to use DC between the mm-wave RAT and e.g. LTE-A, even without aligning the different numerologies. Additionally, the DC can accommodate large latencies between the nodes as there is no strict timing relation between the links. To extend the Rel-12 DC, we propose an enhanced DC (eDC), which can support data split and duplication to more than two sites, and potential diversity of RRC signalling. Various considered service flows (SFs), i.e. Master Cell Group SF and Secondary Cell Group SF as well as SF Splits at Master 5G-NB (M5G-NB) and Secondary 5G-NB (S5G-NB) are shown in Figure 3. To enhance the flexibility, it is considered that the service flows (SFs) could be split in both the Master eNB (MeNB) and the Secondary eNB (SeNB). An alternate solution would be the data split and duplication of fast switching of User and/or Control Plane between the eNBs which may provide sufficient performance with lower complexity.

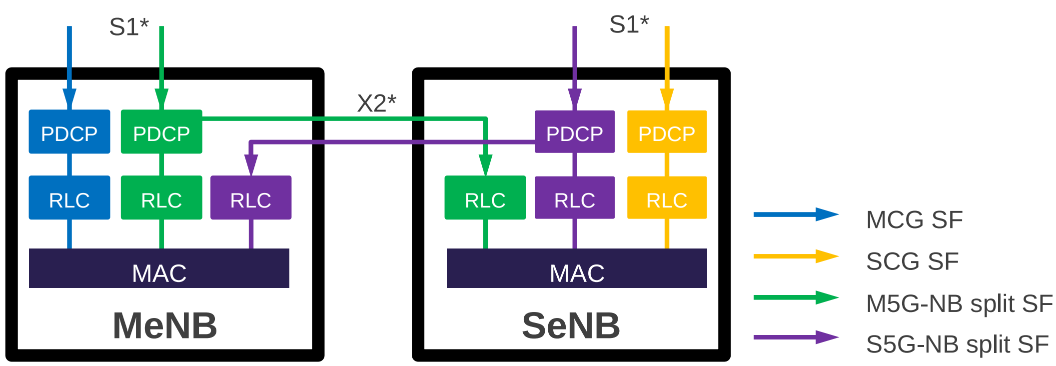

Figure 3 Potential radio flow splits for enhanced Dual Connectivity

To ensure robust connectivity, it will be important to enable proper link selection and timely rapid handovers between the links. Given the wide range of envisioned use cases and deployments, the variability and frequency of handovers will differ significantly. For instance, if the mm-wave RAT is supported by e.g. LTE-A, it can be possible to preload coverage maps, either in the UE or the network and select links based on heuristics and the geographical location of the UE to enable faster handovers at the cost of increased overhead and processing. Additionally, to improve the reliability of the connection, it can be possible to use redundant links if the backhaul capacity allows for tight synchronization, whereby the reliability can be increased by orders of magnitude at the cost of additional resource usage. Some link combination schemes are proposed in [MMM16-D31]: maximum-ratio-combining (MRC) and Low-Density Parity-Check (LDPC) coding schemes. The idea behind is to transmit a few signal copies (in MRC) or redundant bits (in LDPC) through various channels to enable full diversity between links i.e. the data is recovered even if only one link is successfully received.

5 Mobility

The applications expected to be supported by next generation mobile systems will introduce more stringent requirements on performance regarding capacity, latency and energy efficiency; which will necessitate the review of existing operational state of devices [3GPP TS.36.304] [3GPP TS.36.331]. For 5G mm-wave systems, the beam-based antenna patterns will result in dynamic variations in coverage, signal quality, and channel quality with slight movements in the UE, and rotations will change the directionality of the beams. At the same time, signal blockage from obstacles can greatly reduce the beam coverage. This leads to frequent handovers between different beams in order to provide sufficient coverage and connectivity. The future mobile systems apart from offering seamless mobility, will also introduce efficient schemes to minimize signalling load and delays, provide low energy consumption during inactivity periods, whilst maintaining the network management simple.

5.1 Active Mode Mobility: mm-wave Access Point Clustering

When the UE has ongoing traffic, the challenges of beam based mobility could be solved by using access point clusters within which the mobility of the UE is transparent to the CN. Within the cluster, the handover of the UE between different nodes or beams can be performed without any CN signalling.

The layout and architecture of the cluster will depend on the deployment: the quality of backhaul and coverage of the different nodes. If the backhaul is ideal with very low latency, the cluster can be coordinated by a central node, handling all scheduling between the nodes. A more common scenario is likely to be a deployment with a non-ideal backhaul, which may preclude a central scheduler. Instead, in such cases each node is responsible for the lower layers (MAC and PHY), and can relay packets through an evolved RLC layer to other nodes when a UE needs to switch APs as can be seen in Figure 4.

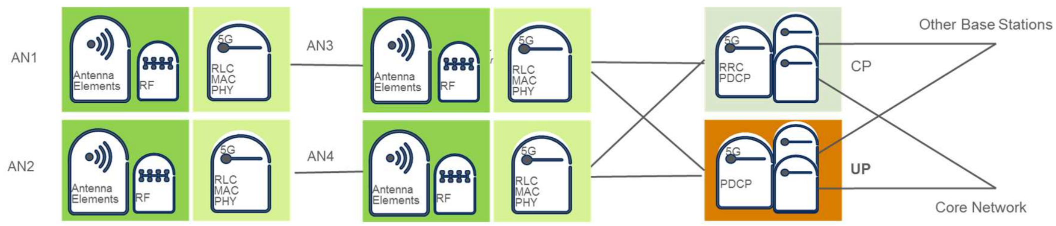

Figure 4 Architecture of base station clustering

As the complexity of the APs may vary, the coordination of the cluster will differ. If each node has a full protocol stack implemented, either of them can serve as Cluster Head (CH) responsible for deciding which node or beam should transmit to the UE and when to switch to another node. The CH can then be relocated to another AP e.g. due to UE mobility. The location of the CH is UE-centric and an AP serving as CH for one UE may be a secondary AP for another UE. In some deployments, relocation of the CH may not be feasible, e.g. if only one of the APs has the processing power or CN connection quality to support the cluster, then this node will always be the CH and coordinate the mobility within the cluster.

To ensure connectivity within the cluster, it may be necessary to rely on the wide area coverage of low-frequency RATs, e.g. LTE-A, when the mm-wave RAT has limited reliability e.g. due to signal blockage. The lower frequency can then relay traffic and control signals from the CH to the UE and assist in intra-cluster mobility.

Additionally, the mm-wave access clustering is expected to work even with wireless self-backhauling (described in more detail in Section 8), where the nodes may relay traffic using the mm-wave air interface. However, this may introduce additional latencies in the system which needs to be considered.

The main advantage of the access point cluster is to enable local area mobility without any CN signalling. In order to ensure connectivity when a UE moves, it will need to switch beams either within the same AP or to a beam from a different AP. If the beam switch is within a single AP, that AP can decide by itself; however if it needs to switch to another AP, the CH need to coordinate the switch.

In addition, if the UE moves beyond the current coverage of some of the APs in the cluster, it will be necessary to update the cluster constituents and possibly the location of the CH. If cluster members need to be added or removed from the cluster, this will be handled by the CH with possible assistance from the CN if certain addresses need to be added. If the CH need to be relocated within the cluster, the CH can decide this by itself and notify the CN only once completed. If the target AP is beyond the reach of the CH, e.g. due to use of self-backhauling, the CN will need to coordinate the CH relocation. To reduce the latencies associated with inter-cluster CH relocation, the CN can prepare potential target AP with the UE context and activate transmission of reference signals. The UE would then measure these reference signals and trigger the switching once predefined thresholds are fulfilled.

To optimize the cluster management, it could be beneficial to consider the extent of UE mobility and implement location information and heuristics to predict when and where a UE should perform handover; considering link quality and the overhead associated with the handover. Additionally, as the mm-wave RAT will be heavily reliant on beam-based transmission, the beam training and beam width adaptation strategies need to be evaluated to optimize the handover procedure for various mobility scenarios. As some of the APs within a cluster may be serving multiple UEs using overlapping beams will be beneficial to coordinate the beam steering between the APs based on network statistics, to minimize the beam interference.

5.2 Inactive mode mobility – novel inactive RRC state

It has been observed in the existing technology, that inactivity timers are typically configured to be very short (i.e. down to 10-20 seconds), resulting in frequent transitions between RRC_Idle and RRC_Connected state. Although, the state transition time in LTE between RRC_Idle and RRC_Connected is required to be less than 100 ms and much lower than 50 ms for LTE-A; the transition time may be too large for some applications. In such instances, the option to keep the UE in RRC_Connected during inactivity periods is not viable, especially in situations when the network will require frequent handovers in beam-based mm-wave propagation, resulting in significant overhead. To address this state handling in 5G systems, a novel state has been proposed. In addition to the RRC_Idle and RRC_Connected states as in traditional systems, a new RRC_Connected_Inactive [S2- 161323] state is introduced as shown in Figure 5.

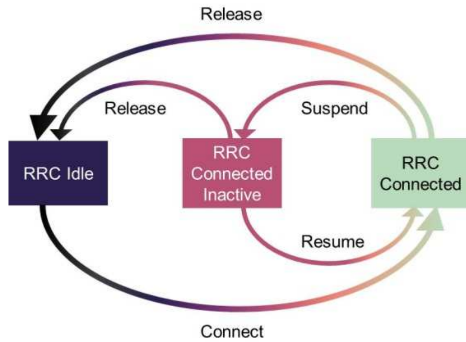

Figure 5 State transition diagram showing three RRC states including the novel state RRC_Connected_Inactive

To briefly explain, the novel state is flexible— in the sense that it can be configured by the network based on the services requested by the UE. The UE mobility and cell reselection is based on UE measurements, where the configurations are received from the network. Importantly, in this state, the UE context is maintained by both the RAN and the UE to facilitate seamless, lightweight state transitions for transmitting small data packets, with active user plane and control plane connections between the RAN and core network (CN). If the UE should be optimized for fast system access, it can be configured to constantly monitor paging and dedicated signalling. Another characteristic feature of this new state is the distributed tracking area (TA) management scheme. The RAN configures a TA within which the UE can move without notifying the CN, while the RAN tracks the UE mobility and reachability. If the UE moves beyond the TA, the CN will be informed through a tracking area update.

6 Multi-RAT Multi-Layer Management

The generic approach to multi-RAT management is to seek scalable management at several layers, depending on the logical and control functions required to switch between different RATs. The proposed generic model considers several layers for link adaptation metric integration.

Link adaptation metrics make the decision on selected RATs to transmit data, following several targeted criteria (power, cost, spectral efficiency…).

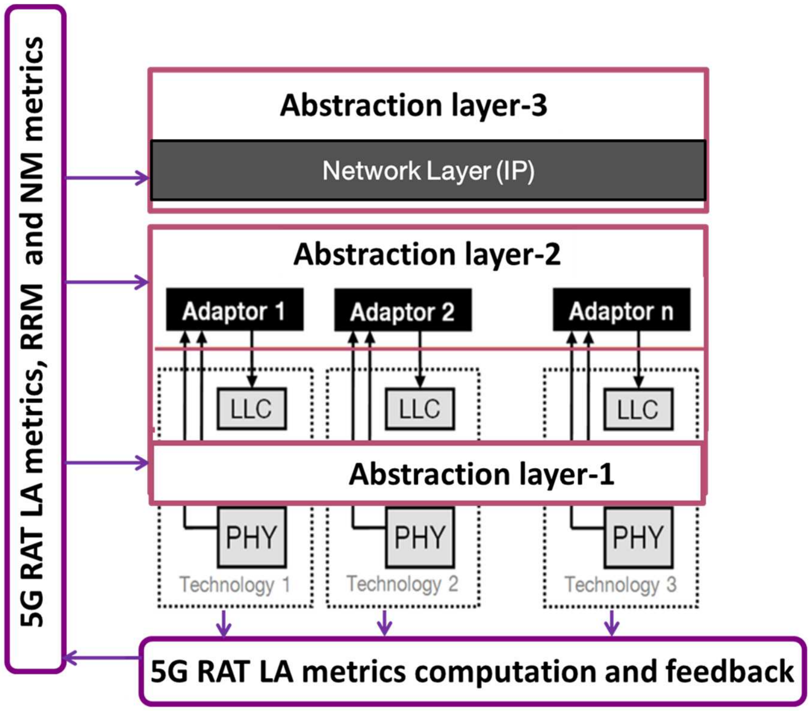

Figure 6: Multi-layer multi-RAT management architecture

Figure 6 illustrates the multi-layer multi-RAT concept, where three layers are envisioned to perform air interface switching or MCS switching in a single air interface. This multi-layer multi-RAT architecture is also denoted multi-layer abstraction layer linked to NGMN 5G concept.

The abstraction Layer-1 performs transmission mode switching (transmission mode is assimilated to baseband processing and MCS) by exploiting MAC mechanisms. These protocols integrate new link adaptation metrics using the same communication channel to forward the decision.

The abstraction Layer-2 considers a layer, named L2.5 layer, able to manage several air interfaces using common control functions. These functions set up a dedicated interface in accordance with link adaptation metric decision to transmit data. It should be noted that this solution is an extension of the I-MAC concept [KBN12].

The abstraction Layer-3 exploits IP network protocols to switch from one interface to another, or slice from one network to another. C/U plane splitting application for WLAN-LTE aggregation utilizes the abstraction Layer-3, in order to perform Wi-Fi offloading in small cells.

Multi-RAT link adaptation metrics [SUMP16] are designed to perform multi-layer multi-RAT management. These metrics are usually evaluated at the PHY layer and forwarded to one of these levels of abstraction layers. The choice between these architectures trends is low complexity and latency, depending on the concerned interfaces in the multi-RAT scenario.

7 Low-Band Integration

The mm-wave RAT will in many cases be deployed in areas already covered by low frequency RATs below 6 GHz, e.g. LTE-A which can provide much wider coverage and reliability due to the different propagation properties at these frequencies. In order to fulfil the stringent requirements of the 5G use cases, it may be beneficial to rely upon the low frequency bands, either for the control plane or user plane.

7.1 Control Plane Aspects

As the low frequency bands can offer a more reliable connectivity compared to the high frequencies, it can be beneficial to transmit some of the control signals using the low frequency RATs. For this, it will be necessary to integrate the mm-wave RAT with the low frequency RAT in order to allow a flexible distribution of the control signals.

Some of the control signals are directly related to the physical properties of the transmission channel, e.g. synchronization and channel estimation signals, both for initial access and data transmission (see Figure 7).

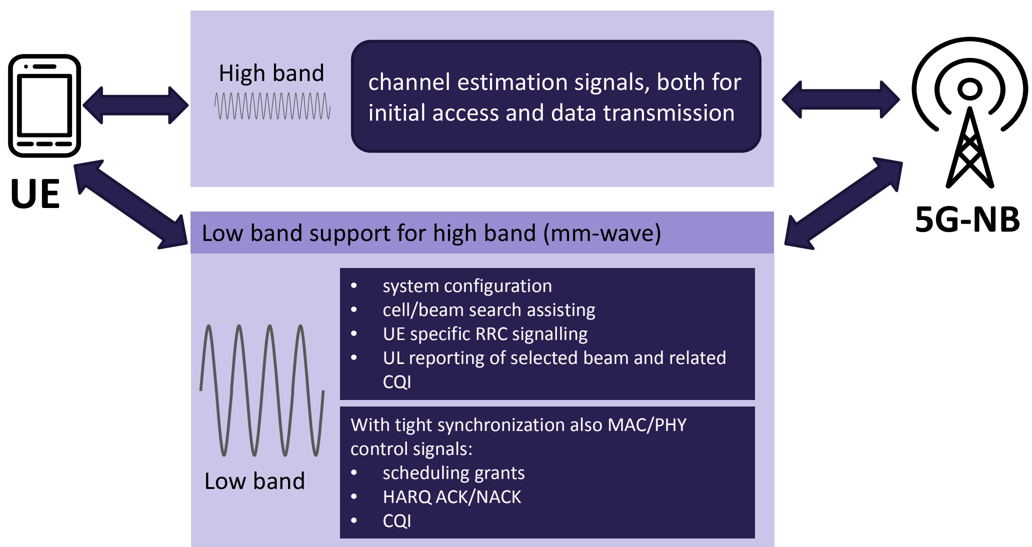

Figure 7: Low-band can support mm-wave system by carrying selected control or system information. Control signals directly related to the physical properties of the transmission channel cannot be offloaded to low-band system

These signals will have to be transmitted through the mm-wave channels and cannot be offloaded to the low frequency bands. Additionally, some of the MAC/PHY control signals are delay sensitive, e.g. scheduling grants, HARQ ACK/NACK or CQI, and the timing would have to be tightly synchronized between the mm-wave RAT and the low frequency RAT in order to offload them to the lower frequencies. Especially, if the TTI lengths differ between the RATs, the tight synchronization will be more difficult.

However, some of the control signals are less delay sensitive, e.g. system configuration information, cell/beam search assisting information, UE specific RRC signalling, UL reporting of selected beam and related CQI. These signals could be sent with greater reliability over the lower frequency RAT at a stage where the mm-wave link setup has not fully been completed.

7.2 User plane aspects

There are several options considered for the tight integration of the mm-wave RAT with low frequency RATs as described in Section 4. As a baseline, the mm-wave RAT is assumed to have at least a common enhanced packet data convergence protocol (PDCP) layer with LTE-A, where the PDCP instance can be placed in either an mm-wave AP or an LTE-A eNB interchangeably. This will allow the network to setup DC within the mm-wave RAT or within LTE-A separately, or jointly between them. One of the enhancements that are proposed for the mm-wave RAT is to combine the 3GPP dual connectivity options 1a (bearer level split at S-GW) and 3c (packet level split as PDCP PDUs in MeNB) to allow both options simultaneously. In addition, a variation of option 3c will be explored where the traffic is split in the SeNB instead of the MeNB to allow full flexibility to the network. The benefit of splitting the traffic at or above the PDCP layer is that the lower layers can remain independent and not require strict time synchronizations. This also enables the use of high latency backhaul between the nodes. It may also be possible to perform the tight integration at lower layers using a common MAC, if the scheduling and synchronization can be aligned.

Another aspect that needs to be taken into account in is in situations when the high frequency and low frequency links differ greatly in capacity, where switching from a high-capacity to a low-capacity link may be detrimental to the quality of experience (QoE) of the application. One way to solve this could be for the lower layers to inform the application layer of impending handovers and predicted future throughput, so that the application layer can adjust its traffic pattern accordingly. For instance, a video streaming application could fill up its buffer more rapidly, if it expects the high capacity link to be lost.

8 Self-Backhauling

Wireless self-backhauling is a promising solution to support emerging networks via autonomously establishing backhaul connectivity to existing and emerging network structures; in particular, where dedicated backhaul becomes expensive and difficult to deploy. In self-backhauling, the backhaul and access links share the same radio resources as depicted in Figure 8.

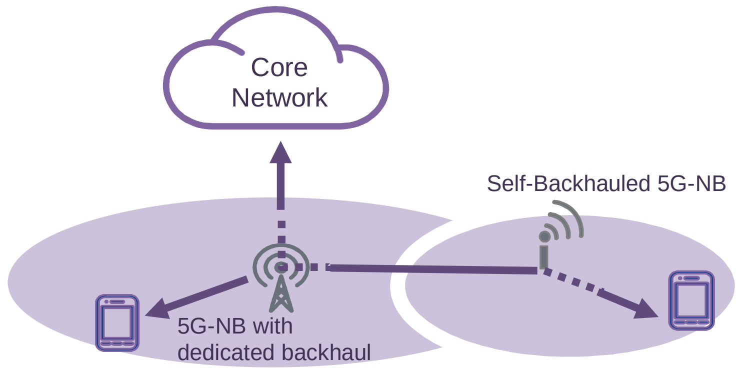

Figure 8 Concept of self-backhauling: backhaul and access links share the same radio resources

Self-backhauling can be seen as a means for coverage extension and can be alternatively used for capacity extension (in UDNs), by employing multi-connectivity towards more than one access points (i.e. with dedicated backhaul links). One amongst the diverse requirements of wireless self-backhaul node, is to provide the ability to automatically attach itself to the surrounding donor access/base stations in a plug-and-play fashion. Therefore, the node should be flexible to support handovers between the donor NBs, multi-connectivity to donor NBs, out-band backhauling and combination of all of them.

The dynamics and self-autonomy of self-backhauling solutions can gradually evolve into SDN-based solutions; where one logical controller is supposed to monitor topology changes, node-to-node radio channel status and all the traffic needs in a real-time manner.

In this case, backhaul networking for dense deployed small cells could be characterized by a ringed-tree topology with multiple backhaul links per node, and different levels of backhaul links where vertical links would have higher priorities in route selections than horizontal ones. An example of ringed-tree backhaul networking is illustrated in Figure 9.

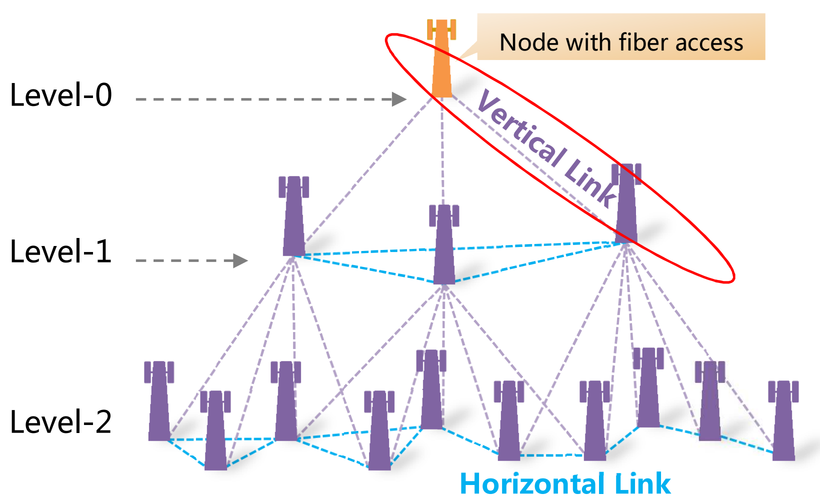

Figure 9 Ringed-tree as an example of backhaul topology

Moreover, dynamic RRM decisions will be made by the controller in terms of how much radio resource is to be allocated per link at each node. The network can also enable multi-route connection, allowing coordination and cooperation amongst network elements to be performed via management provided by the controller (i.e. in order to achieve network-level optimization). In this context, efficient multiplexing of time and frequency resources between backhaul and access links, dynamic link scheduling (to achieve the optimum distribution of non-conflicting resources) and fast congestion management and routing algorithms define attractive areas to be further explored.

9 Conclusions

This white paper has introduced key concepts for a 5G mm-wave architecture and its integration within 5G networks to provide high performance, flexibility and support to ultra-dense deployments. The concept of Network Slicing enables 5G use cases with highly divergent requirements, providing architectural flexibility by operating on logical instead of physical network elements. It is intended to be implemented not just in the core part of the network but also in the 5G RAN, allowing the RAN to be aware of traffic flow for slice association and the optimization of resource allocation accordingly. Here, several multi-connectivity solutions in the RAN architecture level have been proposed which will affect both Control- and User-Plane design (e.g. split of control information between RATs, data duplication, etc.). These solutions can be implemented as an extension of LTE Rel-12 Dual Connectivity. In this context, it may also be possible to perform the tight integration at lower layers using a common MAC, if the scheduling and synchronization can be aligned. To provide highly efficient mobility management and limit CN signalling, solutions for mm-wave cell clustering has been proposed. In case of infrequent transmission of small packets, CN signalling can be further reduced by introducing an intermediate RRC inactive connected state. The multi-layer multi-RAT is a generic approach to multi-RAT management, seeking scalable multi-RAT management at several layers, depending upon the logical functions required to switch between different RATs and interfaces. The proposed generic model considers several abstraction layers for link adaptation metric integration. Self-backhauling is yet another concept of high importance for upcoming mm-wave 5G networks. It can be used to assist coverage extension and/or allow capacity extension (in UDNs), by employing multi-connectivity towards more than one access points. Such self-backhauling solutions can gradually evolve into SDN-based solutions supervised by one logical controller in a real-time manner.

The intention of this white paper was to provide a high-level overview of mm-wave related 5G architecture aspects, and interested readers may refer to this project related deliverable [MMM16-D31] for in-depth technical details. The mmMAGIC project will continue developing the architecture concepts and a final revision will be provided by June 2017.

10 Acknowledgement

The research leading to these results received funding from the European Commission H2020 programme under grant agreement n°671650 (mmMAGIC project).

11 References

[3GPP TS 36.300] 3GPP TS 36.300 V13.2.0 E-UTRA, E-UTRAN, Overall Description; Stage 2, 2015

[3GPP TS.36.304] 3GPP TS.36.304, UE procedures in IDLE mode;

[3GPP TS.36.331] 3GPP TS.36.331, Radio Resource Control (RRC) specification.

[E14] Ericsson, "The Real-Time Cloud White Paper", February 2014, see http://www.ericsson.com/res/docs/whitepapers/wp-sdn-and-cloud.pdf

[KBN12] R. Kraemer, M. Brzozowski, S.Nowak, "Reliable Architecture for Heterogeneous home networks: the OMEGA I-MAC Approach", Elec. Energ. Vol. 25, No 1, April 2012

[MMM16-D31] mmMAGIC D3.1, "Initial concepts on 5G architecture and integration", March 2015. https://5g-mmmagic.eu/results/#deliverables

[NGMN15] NGMN Alliance, "5G White Paper", February 2015

[S2- 161323] 3GPP S2- 161323 "Solution: Mobility Framework ", February 2016

[SUMP16] I. Siaud, A.M. Ulmer-Moll, H. Peng, S. Nanba, K. Moriaki, „C/U-plane splitting architectures and Inter-RAT management for Radio Reconfigurable Systems", ETSI workshop on Future Radio Technologies focusing on Air Interfaces, Sophia-Antipolis, January 2016