Split Options for 5G Radio Access Networks

Paul Arnold, Nico Bayer, Jakob Belschner, Gerd Zimmermann
Technology Innovation
Deutsche Telekom AG
Darmstadt, Germany
{paul.arnold, nico.bayer, jakob.belschner, zimmermanng}@telekom.de

Abstract

5G networks are supposed to offer a high flexibility in a several ways. In this regard, a twofold split of the processing in the radio access network is under discussion: A control plane / user plane split to support the software defined networking principle and a radio protocol stack layer based split to allow a flexible placement of processing functions between a central and one or more distributed units. In this work, the motivation and state of the art for both splits are described including a discussion of the advantages and disadvantages. It is followed by a description of a network architecture allowing a flexible implementation of these splits. This especially focuses on the required interfaces between control and user plane.

1 Introduction

The network functions (NFs) of a wireless network are typically categorized into two groups[1]: The user plane (UP, also called data plane) is responsible for forwarding data from the source to the destination, including the corresponding processing. The control plane (CP) controls the UP, for example in terms of setting the routing path of a packet or how to map data packets onto radio resources in time and frequency domains (radio resource management). The CP also provides a set of other functionalities such as connection / mobility management and broadcasting of system information.

The separation of CP and UP according to the Software Defined Network (SDN) concept is a recent trend in the definition of the 5G architecture [1-3]. It requires to categorize all NFs as being either part of CP or UP based on functional decomposition [3] [4]. Any kind of interaction between CP and UP is supposed to happen through standardized interfaces.

The anticipated benefits of a CP/UP split ("vertical functional split") are:

- In multivendor networks, a standardized interface to the CP enables a consistent control over network elements and NFs from different vendors / manufacturers, e.g. in terms of interference management for ultra-dense networks [4] [5].
- Due to the tight coupling of CP and UP NFs in today's networks, the replacement or upgrade of a CP function often requires also the replacement of UP functions. Avoiding this might offer significant cost savings.
- The independent evolution of CP and UP by possibly modifying and adding CP functions without changing the UP (and vice versa) could make the rollout of new NFs faster thus enabling a more flexible network.

Besides, there are also disadvantages:

- CP and UP functions are often tightly coupled, especially in the lower radio protocol stack layers. It might be challenging and could affect the performance when fully separating CP and UP handling, especially if the processing is not collocated.
- Standardization is required in case the interfaces between CP and UP have to be extended to introduce new features which might slow down this process. Integrating additional interfaces in a proprietary manner in combination with standardized ones is not a suitable solution, as it would destroy the benefits of a CP/UP split. For example, a flexible change of CP NFs in logical network elements would not be possible any more if only selected UP NFs support certain proprietary interfaces.
- Additional effort in terms of testing is required to guaranty the interoperability of CP and UP functions from different sources (shifting the effort to system integrators supporting the operators instead of doing this work at a single vendor).

In parallel to the CP/UP split, also a second split is discussed, the so-called "horizontal functional split". Here NFs (CP as well as UP) can be flexibly allocated either in distributed units (DU) close to the antenna sites or in a central unit (CU). The main intention of the horizontal split is to enable gains from centralization, e.g. through coordination as anticipated in cloud-based radio access networks (C-RAN) [3] [4], but it also allows NFs to be placed in CU and DU according to performance criteria like latency as well as to adapt the placement to the characteristics of the x-haul (back-, mid-, fronthaul) transport

---

[1] A third group are management and operation (MANO) functions, but these are out of scope of considerations within this paper.

network between CU and DUs [6] [7]. NF centralization strongly increases the x-haul requirements in terms of bandwidth and latency (in the extreme case corresponding to today's CPRI implementation [8]). Also 3GPP is currently considering both functional splits as part of its study item on the 5G New Radio (NR) [9].

Section 2 of this paper provides an overview about a RAN network architecture supporting flexible CP/UP splits based on SDN principles. In Section 3 results for a functional decomposition of 5G RAN CP and UP are described. Section 4 covers aspects on multi-connectivity including common control functions required for it, followed by a summary and conclusions in Section 5.

2 CP/UP-split based network architecture

In this section a RAN design concept with a full CP/UP split in between is described. It covers the transport as well as the access network and uses a horizontal split into CUs and DUs.

Figure 1 gives an overview of the proposed architecture. The CUs in the RAN are represented by the Central Access Controllers (CACs) that centrally host CP and UP functions. They are split into an UP part (CAC-U) and a CP part (CAC-C). Typically, the lower layers of the radio protocol stack are hosted close to the antenna sites, whereas the higher layers are processed at the CAC, but in principle also a fully flexible allocation is feasible.

The transport network (aggregation) which forwards the UP data from and to the core network (CN) is implemented through SDN switches or routers. With respect to traffic routing the CN mobility management function [10] acts as the responsible SDN controller. The main role of the SDN controller is to enforce that data is forwarded to the correct antenna site, especially in case of mobile users.

Beside the already mentioned general advantages of a CP/UP split, this SDN-based approach offers additional improvements compared to legacy tunnel-based approaches as the GRPS Tunneling Protocol (GTP) based solution in LTE and UMTS, such as reduced overhead and improved integration with fixed networks [11].

To realize a scalable approach it does not make sense to implement a country-wide RAN via a single CAC (or CU, respectively), but to implement several CACs each controlling the radio processing for a certain number of antenna sites (domain). Suitable locations for CACs are e.g. the central offices of fixed or integrated network operators [3]. To support especially low latency applications, mobile edge computing (MEC) facilities [2] [3] can be integrated into the CU. Typically, the NFs running in the CU (CAC-C/U) are implemented as virtual functions (VNFs) on server platforms based on network function virtualization (NFV) principles [12].

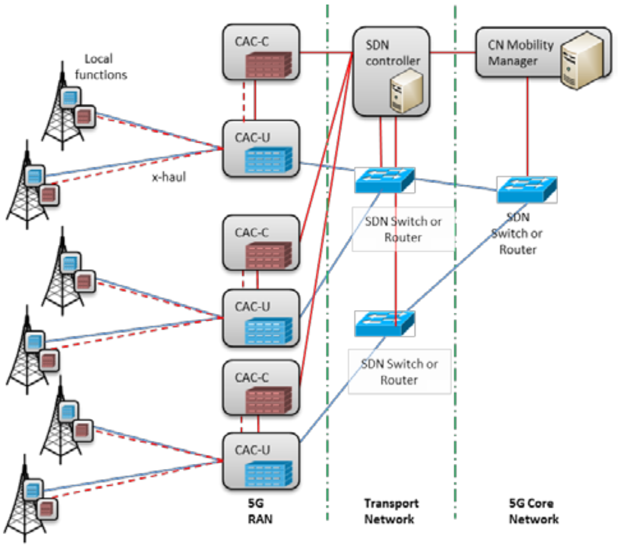

Figure 1: SDN-based 5G network architecture supporting flexible functional CP/UP splits especially in the radio access (not all CN functions are shown)

In the presented architectural approach, three cases of user mobility handling are possible:

1. Between the sites within the domain of a CAC, mobility is handled CAC-internally. This can happen through fast UP switching [5]. In that case no signaling traffic is required between RAN and CN.
2. Inter-CAC-U handover: Here the user equipment (UE) moves from one CAC domain into another one. If both CACs are connected to the same SDN switch or router the SDN controller of the transport network can simply trigger the redirection of the data flow.
3. CN-based handover: In case a path switch has to happen at the highest level (CN-based), it is under the responsibility of the CN mobility manager to send a command to the SDN switches/routers of the CN. In addition, the new route in the transport network has to be set by the corresponding SDN controller.

The cases 1 and 2 describe a RAN-based mobility, where the mobility handling happens only within the RAN. This is beneficial because of low latency between involved components and therefore a low handover interruption time (ideally zero). This advantage is especially relevant for ultra-dense radio node deployments (using e.g. millimeter Wave (mmW) bands) with a high number of mobility events [5].

3 Functional decomposition of CP and UP

With respect to the processing in the RAN a more complex interaction between CP and UP is required in case of a CP/UP split. Figure 2 shows the UP processing chain

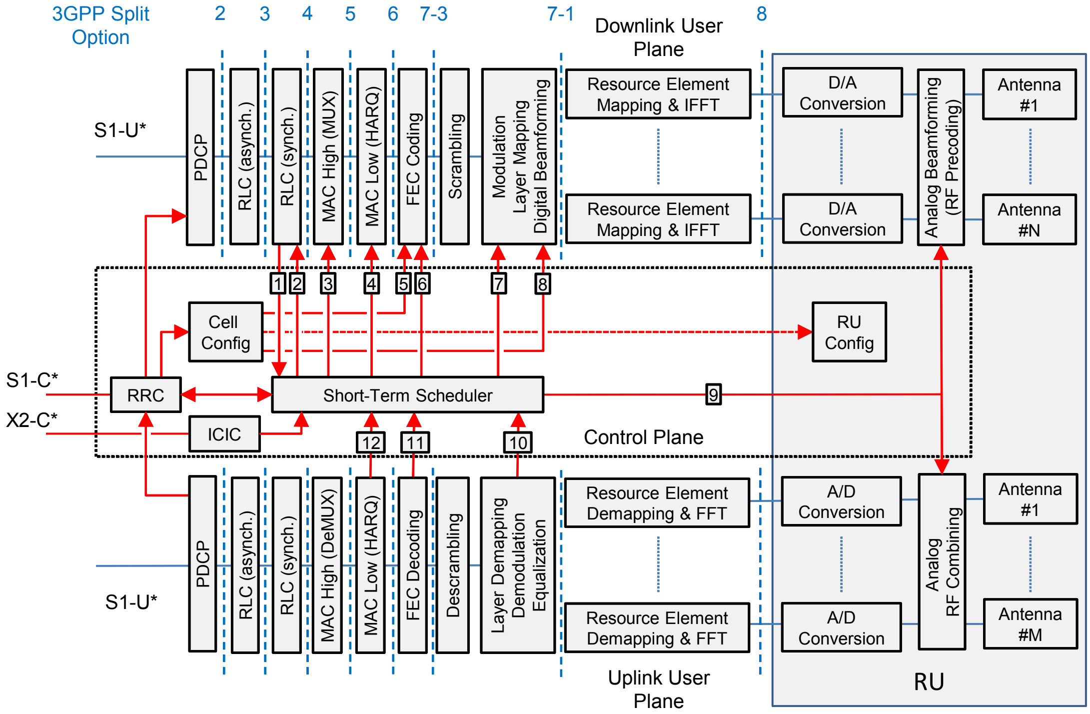

Figure 2: Control and user plane decomposition and interactions in the radio access network (network infrastructure part only; single radio protocol stack)

for downlink (DL; upper part of the figure) and uplink (UL; lower part).

The CP functions are separated in the middle of the figure. The interactions between CP and UP are indicated by arrows and described in the following (please note: only main interactions are shown to not complicating the figure).

The CP NF Radio Resource Control (RRC) implements the corresponding 3GPP protocol layer. It is mainly responsible for the establishment, maintenance and release of connections to the UEs. The required interaction with the UEs happens by generating RRC control messages, which are then forwarded to the UP. By handing over the generated messages to the Packet Data Convergence Protocol (PDCP) layer, they enter the UP processing chain and are finally transmitted through the antennas. Corresponding RRC messages generated by the UEs are processed by the UL UP chain and then forwarded to the CP NF. Thus a full communication between the CP NF RRC and the UEs is enabled through the UP.

The CP NF “Cell Configuration” is responsible for transmitting cell information (e.g. the cell identification) and setting basic cell parameters (e.g. transmit power and electrical tilt). This happens via sending broadcast information and reference symbols through the UP (interactions 5 and 8) and by configuring the radio unit (RU).

The scheduler represents the CP NF with the strongest coupling to the UP. The following interactions with the UP have been identified and are indicated with corresponding numbers in Figure 2:

1. DL buffer status: DL data arrives from the CN through the S1-U* interface (via the transport network). It is processed by PDCP and Radio Link Control (RLC) layer which then reports to the scheduler that data for DL transmission is available.

2. Payload selection: The scheduler selects data to be forwarded to the Medium Access Control (MAC) layer.

3. DL resource assignment and generation of UL transmission grants: In the DL, this enables the MAC layer to generate corresponding transport blocks. For the UL transmission grants are generated and transported by the UP to the UEs.

4. Retransmission control: Retransmissions by means of Hybrid Automatic Repeat Request (HARQ) are also controlled by the scheduler, who sends the corresponding commands to the UP.

6. Coding scheme: The scheduler sets the coding rate to be applied (per UE) and configures the UP accordingly.

7. Antenna mapping, precoder, modulation scheme: Similar to coding scheme, the scheduler also configures the modulation scheme to be applied. For Multiple Input Multiple Output (MIMO) operation, also antenna mappings and precoder settings are required at the UP.

9. In case of analog beamforming (e.g. for Massive MIMO), the scheduler sets the corresponding antenna weights used in the UP.
10. Channel State Information (CSI) from UL sounding: In UL, after demodulation, CSI can be generated based on sounding sequences that the UEs sent.
11. CSI from reporting, UL scheduling request: After the demodulation the CSI information from reporting is available. Also scheduling requests for future UL transmissions have to be forwarded to the scheduler.
12. HARQ status: The scheduler receives the status of UL and DL HARQ processes, e.g. acknowledgements.

The inter-cell interference coordination (ICIC) also acts as CP NF, but works in contrast to the scheduler on a long-term basis, i.e. not on transmission time interval (TTI) level.

Figure 2 also shows some selected options for horizontal functional splits in the UP currently discussed in 3GPP for 5G NR [9]. Options 2 and 3 represent higher radio layer splits. In Option 2 the UP processing of PDCP takes place at the central unit (the CAC-U). All other UP functions remain in the DUs at the antenna sites. Option 3 is similar to this with the difference that also asynchronous RLC processing takes places at the CAC-U. Synchronous RLC NFs are performed in the DUs. The applicability of Option 3 is related to proposed changes in the 5G protocol structure separating NFs with strict timing requirements from those with loose ones [4] [9].

Options 7 and 8 represent lower layer splits (within physical layer) with Option 8 known as conventional CPRI. Also for Option 7 (here variants 7-1 and 7-3 are considered) most of the UP processing happens in the CAC-U. This would imply that also the scheduler is centralized, i.e. hosted at the CAC-C. Output of the scheduler that is required at the CAC-U (e.g. interface number 7) would have to be signaled from the CAC-C to the DU in this case.

The required data rates for the interface between CAC-U and a DU scales increasingly from higher layer splitting options to lower ones, notably dramatically at Option 8. This is depicted by an example in Figure 3, where cases with 2 to 8 antennas are studied under typical LTE assumptions listed in [13] (Annex C). With respect to Full Dimension (FD) or Massive MIMO approaches to be considered for evolved LTE-A Pro and 5G NR this situation tightens. Calculations for different multi-antenna cases taken care also of expected channel bandwidth increase for 5G NR are shown in Figure 4, where the most challenging case for LTE from Figure 3 is compared with 5G Air Interface Variants (AIVs) with 32 and 64 transmit antenna ports. For the 32 port case a channel bandwidth of 100 MHz is assumed which may be typical for a carrier frequency of e.g. 3.5 GHz, whereas for the 64 port case a bandwidth of 400 MHz was taken as example for future mmW signals. The nearly unachievable extreme high data rates make Option 8 infeasible especially in many deployment scenarios.

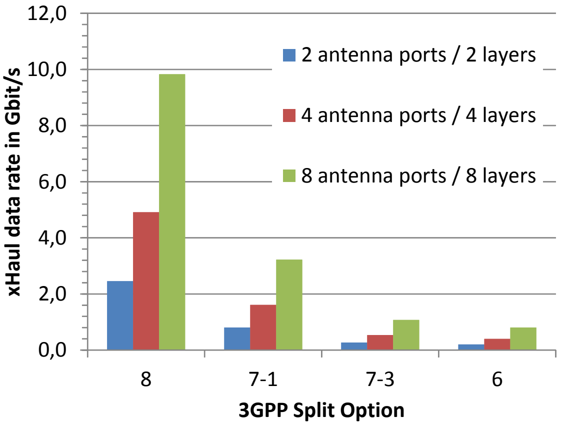

Figure 4: Downlink data rate for horizontal functional split options 8 – 6 considering LTE-like parameters (e.g. 20 MHz channel bandwidth)

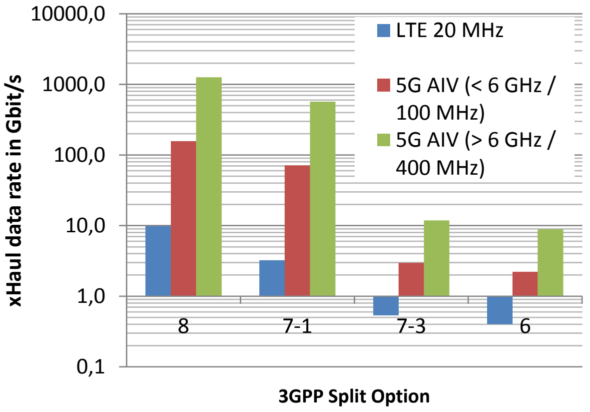

Figure 3: Downlink data rate for horizontal functional split options 8 - 6 considering typical LTE and 5G NR parameters w.r.t. channel bandwidth and antenna port number

4 Multi-Connectivity and Common Control Functions

Multi-connectivity (MC) will be an important feature in 5G to achieve higher reliability than existing systems required for ultra-reliable services (e.g. for industry automation or vehicular communications). MC may be realized through radio links from collocated or non-collocated antenna sites, applying the same or even different AIVs in varying frequency bands (5G NR, LTE-A Pro, WLAN, etc.) [4] [5]. MC can be seen as an extension of the LTE dual-connectivity (DC) approach [4].

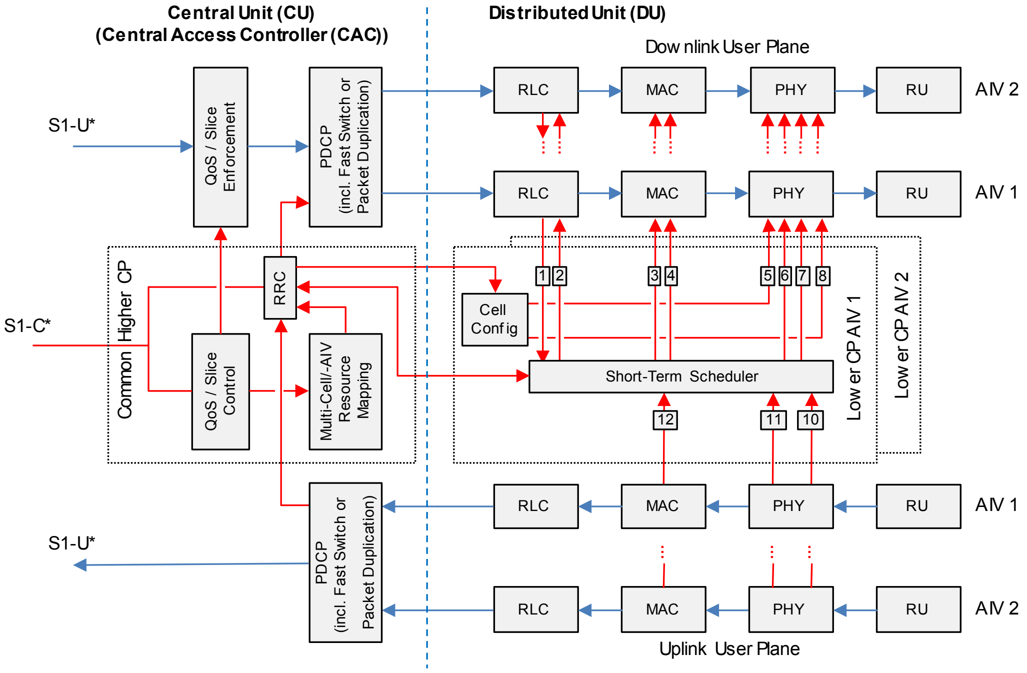

Figure 5: **Partially centralized control and user plane deployment in multi-cell/AIV environment based on horizontal split Option 2**

A multi-AIV deployment based on horizontal split Option 2 in combination with the related CP/UP split, demonstrating also the needed CP split between CU and DU, is depicted in Figure 5. It also shows additional CP NFs hosted at the CU (CAC-C) for e.g. quality of service and network slice control and corresponding UP enforcement above the PDCP layer [6].

Due to increased opportunity range for AIV handling in a centralized environment, the ICIC CP NF is evolved to a so-called Multi-cell/-AIV Resource Mapping which operates on an extended resource framework (antenna sites, frequency bands, AIV-related time-frequency grids, etc.). This NF also controls (via RRC) UP NFs in the PDCP layer, resulting in e.g. a duplication of data packets to be transmitted on one or more AIVs or also a allowing fast switching of data streams between AIVs in one or more DUs. Also horizontal split Option 3 can be applied if only novel 5G NR AIVs are used. For a combination of 5G NR with LTE-A Pro Option 2 has the positive aspect that it is already applied for LTE DC, thus no changes in LTE-A Pro specifications are required. Introducing Option 3 also in LTE-A Pro would result in more efforts for realization.

The approach shown in Fig. 4 and 5 for a high layer split strongly relaxes the x-haul requirements for 5G deployments and allows at least partial central coordination of data transmissions and receptions. The applicability is especially relevant for Massive MIMO usage where the x-haul data rates using a lower layer split scale with the antenna numbers and therefore prevent the implementation of fully centralized CP/UP via the classical C-RAN approach.

5 Summary and Conclusions

In this paper a 5G RAN architecture has been presented which on the one hand allows flexible placement of CP/UP NFs to cope with diverging requirements of 5G services and on the other hand supports scalability. A full CP/UP split in combination with a centralization of CP NFs in a controller according to the SDN principles seems complex to realize. If a C-RAN implementation based on a fully centralized CP and UP cannot be realized due to limitations on x-haul interfaces (bandwidth, latency), the partially centralized approach based on horizontal split Options 2 and 3 can lower the requirements, but keep a sufficient degree of centralization gains.

Acknowledgements

This work has been performed in the framework of the H2020 project METIS-II co-funded by the EU. The views expressed are those of the authors and do not necessarily represent the project. The consortium is not liable for any use that may be made of any of the information contained therein.

References

[1] R. Trivisonno, R. Guerzoni, I. Vaishnavi, D. Soldani: “SDN-based 5G mobile networks: architecture, functions, procedures and backward compatibility”, Transactions on Emerging Telecommunications Technologies (ETT), November 2014

[2] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, et al.: “Mobile Network Architecture

Evolution toward 5G", IEEE Commun. Mag., May 2016

[3] 5G PPP Architecture Working Group: White Paper "View on 5G Architecture", Version 1.0, July 2016

[4] 5G PPP project METIS-II: Deliverable 2.2 "Draft Overall 5G RAN Design", June 2016

[5] 5G PPP project METIS-II: Deliverable 5.2 "Final Considerations on Synchronous Control Functions and Agile Resource Management Framework for 5G", March 2017

[6] 5G PPP project 5G-Crosshaul, http://5g-crosshaul.eu/

[7] 5G PPP project 5G-XHaul, http://www.5g-xhaul-project.eu/

[8] "Common Public Radio Interface (CPRI); Interface Specification", V6.0, August 2013

[9] 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.801 "Study on New Radio Access Technology: Radio Access Architecture and Interfaces", Version 14.0.0, March 2017

[10] 3rd Generation Partnership Project (3GPP), Technical Report (TR) 23.799 "Study on Architecture for Next Generation System", Version 14.0.0, December 2016

[11] F. Yang, M. Huang, S. Zhang, S. Sirotkin.: "Radio Access Network Re-architecture to meet 5G Requirements – A SDN-based Paradigm Shift for Cellular Networks", ICC 2016, 5G RAN Design WS, May 2016

[12] European Telecommunications Standards Institute (ETSI): Industry Specification Group "Network Function Virtualization" (ISG NFV), http://www.etsi.org/technologies-clusters/technologies/nfv

[13] 5G PPP project METIS-II: Deliverable 4.2 "Final air interface harmonization and user plane design", April 2017