Challenges ahead of RAN virtualization in LTE

Desislava C. Dimitrova
University of Bern, Bern, Switzerland
Email: dimitrova@iam.unibe.ch

Lucio S. Ferreira
INOV-INESC/IST, Lisbon, Portugal
Email: lucio.ferreira@inov.pt

Andre Gomes
One Source, Coimbra, Portugal
Email: gomes@onesource.pt

Navid Nikaein
Eurecom, Sophia Antipolis, France
Email: navid.nikaein@eurecom.fr

Alexander Georgiev
CloudSigma, Bulgaria
Email: alexander.georgiev@cloudsigma.com

Anna Pizzinat
Orange Labs, France
Email: anna.pizzinat@orange.com

I. Motivation of LTE network virtualisation

Seamless wireless communication has become an elementary building block of any society. In fact, mobile communications are among the few technologies experiencing rapid adoption in developing countries. While developing countries still have customer growth potential (and thus financial growth), developed countries are showing saturation of the market (slowing financial growth). Not only that but the demand towards mobile services grows posing higher requirements towards the network, its density and operational cost. The combination of increasing service requirements and decreasing network profit has motivated network operators worldwide to search for the next big breakthrough that can ensure their market advantage. Cloud computing may offer them just that.

Network virtualization, i.e., moving network processing in software, has already proven a viable option for the mobile network core. Letting this software run in the cloud benefits from on-demand provisioning of and pay-per-use principle for computation, storage and networking resources. Naturally the question arises: Could the same strategy work for the Radio Access Network (RAN) in LTE and LTE-Advanced systems? The motivation for such change is straightforward - up to 60% of all network operational costs are spent on the RAN [1]. Furthermore, operators typically dimension the RAN for peak loads, while in reality the offered traffic may vary drastically both geographically and temporally. In fact, network measurements show that 50% of network sites generate 10% of the revenue, while 20% of the base stations carry 50% of the traffic [2]. Placing the RAN functionality in the cloud offers the possibility of common base station management, cutting down costs and improving the spectrum use.

RAN virtualisation, however, is not trivial. Although fibre cabling is already a reality in many urban areas [3], it only provides the infrastructure to connect the cloud-based processing with remote antennas. It remains to be shown that a cloud infrastructure - Data Centres (DCs) with General Purpose Platforms (GPPs) - can meet the processing requirements of a pool of base stations (eNBs in LTE terminology). Specifically, two questions require attention. First, what are the computational needs of the eNB processing, taking into account the underlying hardware. Second, can the time constrains of the radio interface be guaranteed.

The Mobile Cloud Networking (MCN) project [4] addresses the challenge of extending the cloud computing paradigm to radio communication networks and is thus concerned with the above set of challenges. The MCN vision of the future radio network is depicted in 1 and shows that virtualised RAN (vRAN) is a central part of that vision. In this abstract we would like to briefly introduce the MCN vision of a virtualised RAN architecture and discuss a first set of observation on its practical realisation on GPPs.

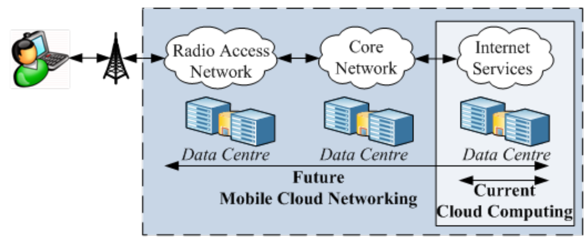

Fig. 1. Mobile cloud networking vision (extracted from [5]).

II. Architecture for virtual RAN

RAN virtualisation empowers the instantiation of software-based RAN components on GPPs located in DCs, where resources such as computation, storage and networking, are freely available and provided on-demand. The successful realisation of the virtualised RAN potential depends on the adequate identification of the individual RAN components and the ability to manage them in real time.

RAN decomposition in interconnected functional components is necessary in order to move the RAN processing in software. Generally, processing in hardware is faster, for which reason heavy computational operation such as these of the RAN signal processing are performed in hardware. Current hardware platforms implementing the LTE eNB are therefore specifically designed and tuned for the operations they should support [6]. Consequently, eNB virtualisation brings two changes. First, processing is moved into software, implying slower processing capacity. Second, GPPs are used as the underlying hardware, losing the advantages of specialised hardware. Naturally, GPPs will fall behind and should be larger in numbers to keep to performance targets. It is easy to see that eNB virtualisation requires augmentation of GPP resources. In order to maximise the performance, the heavy eNB processing should be decomposed into functional models with lower computational needs, which can be mapped to fitting computing infrastructure.

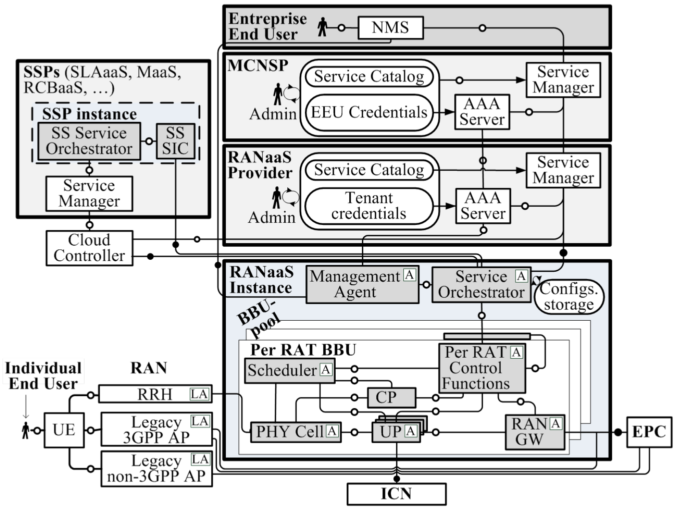

Fig. 2. MCN RANaaS architecture (extracted from [5]).

The benefits of virtualised RAN can be only explored through an appropriate management platform that supports real-time monitoring of the virtual RAN performance and on-demand provision of RAN functionality and underlying computational resources. The combination of virtualised RAN and management platform enables us to introduce the concept of RAN as a Service (RANaaS), i.e., the offering of RAN when needed and where needed [5].

MCN proposes a RANaaS architecture which is able to capture the functional RAN decomposition in software-based components as well as to support the appropriate management of the components, based on the network operator needs. The architecture is illustrated in Figure 2.

The individual RAN components are commonly referred to as Service Instance Components (SIC). Each SIC has a specific functionality mapping to typical eNB functions. In a virtualised RAN the eNB is split between remote radio heads (RRH) for RF transmission and base-band units (BBU) for baseband processing and higher layer functions. BBUs include layers 1, 2 and 3 processing of the LTE radio stack, corresponding to user and control planes [7]. BBUs can be instantiated on virtual machines of data-centres, following the cloud paradigm. Naturally, RRH can not be fully virtualised since physical antennas are needed for the radio transmission. Figure 2 shows the SIC associated with virtualised eNB.

The virtualised eNBs are under the control of a Service Manager (SM) and a Service Orchestrator (SO). The SM allows operators to request for RANaaS instance (accompanied by detailed description) from a provider of the service and acts as the interaction module between provider and customer-operator. The SM provides to the SO relevant information on the requested RANaaS instance such as coverage area, expected traffic, etc. that enable the SO to create the necessary SICs. The orchestration of the RANaaS instance is initiated and managed by the SO. Important responsibilities of the SO are the monitoring of the RANaaS instance and taking scaling decisions to react to changes in performance, e.g., degradation of the service-level agreements (SLAs).

The management components use the facilities of a Cloud Controller (CC) to interact with the underlying hardware infrastructure. Furthermore, various Support Services (SSs) provide specific functionality to vRANaaS, such as monitoring, load balancing, mobility and bandwidth availability prediction. The support services are accessible via the SO.

III. Profiling for virtual RAN

We mentioned earlier that there are two major aspects to be considered when talking about RAN virtualisation, namely, (1) the hardware resources required to support the eNB processing and (2) the LTE frame processing deadlines.

The first aspect relates to the fact that the eNB processing is a set of different processing modules at the physical (PHY) and higher layers (MAC). Furthermore, the processing at the receiver and transmitter parts differs leading us to expect different processing requirements for downlink and uplink at the base station. In addition, the size of the system bandwidth (5, 10, 25MHz) is also expected to have an influence, e.g., larger bandwidth can accommodate more users and thus correspond to heavier processing. Note that while some processing modules are common for all users (e.g., OFDM signal generation), other processing operations are done on per-user basis (e.g., modulation and coding). In order to gain understanding on which modules are processing-heavy and what are general expectations towards the processing we ran several 'profiling' experiments with an LTE emulator. In particular, we chose the OpenAirInterface emulator by Eurecom [8] due to its comprehensive representation of the LTE radio stack and its open source.

We conducted tests with 5, 10 and 20 MHZ bandwidth gathering information on the processing time of the PHY and MAC layers of the eNB for both the transmitting and the receiving direction and increasing the modulation and coding scheme (MCS) used. Also both full bandwidth use and a minimal (a single resource block group allocation) were emulated. The experiments are run on a VM in a public cloud with 2GHz and 2GByte of memory. Results for the PHY layer are shows in Figures 3 and 4 for eNB transmitting and receiving, respectively. Interpreting the results quickly indicate that: (i) as the bandwidth increases so does the processing time, implying the needed hardware setup is system specific; (ii) as it can be expected the processing time changes for full bandwidth allocation but remains relatively constant for minimum bandwidth use; (iii) increasing the MCS index requires more computational resources, implying profiling with realistic traffic combinations should be our next step; (iv) uplink processing (i.e., receiver chain) clearly requires more processing time, suggesting it is worth investigating a possible separation of the downlink and uplink processing on different VMs.

Looking in addition to the MAC layer indicates that in general the processing requirements are lower than those of the PHY layer. The increase in the eNB processing is explained with the scheduling, which is performed over all users and thus increases in the number of users. More importantly, looking at both the PHY and MAC processing, we quickly can deuce that the total processing time can easily surpass 3ms. This is important since there is a 3ms deadline in LTE to process a

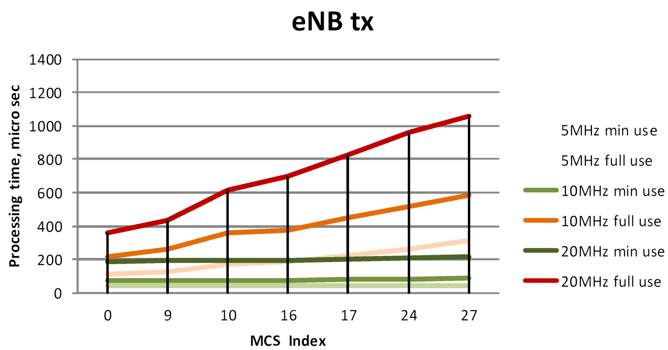

Fig. 3. PHY layer profiling - downlink

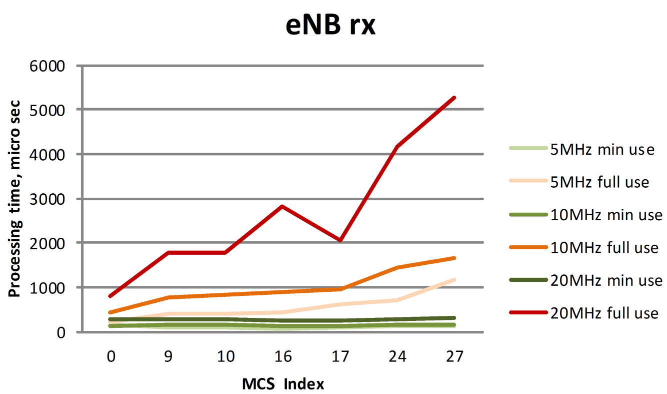

Fig. 4. PHY layer profiling - uplink

frame. This deadline should be kept at all time for the network to operate properly and is dictated by the functioning of the HARQ (Hybrid Automatic Request Response) mechanism. Therefore, in our current work we are looking at which PHY and MAC processing modules in the transmission and receiver chain contribute to the processing delays in order to evaluate the feasibility to run a single eNB with various load on a GPPs platform.

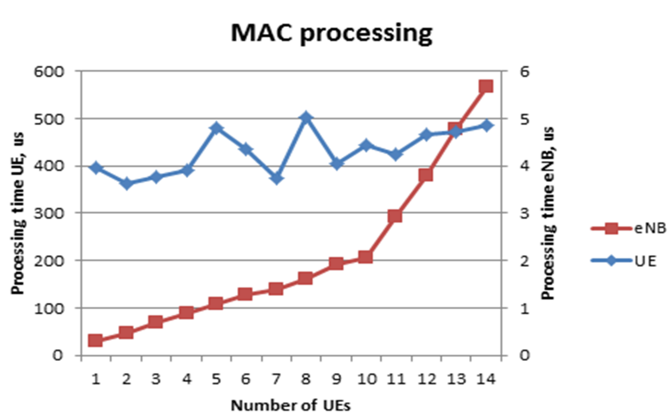

Fig. 5. MAC layer profiling - uplink & downlink

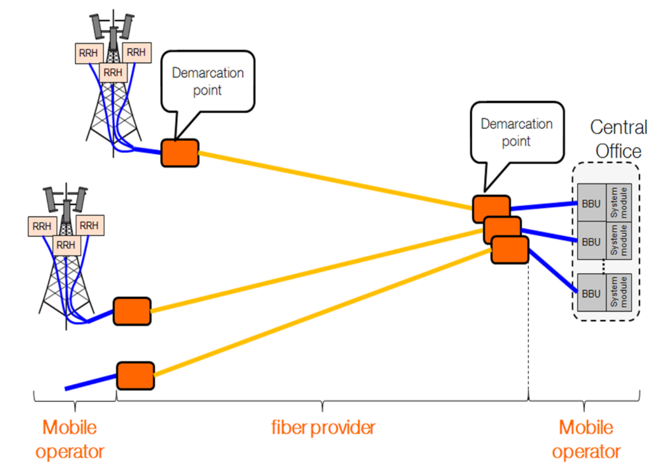

Fig. 6. Fronthaul with network demarcation points.

IV. Fronthaul Requirements and BBU Pool Dimensioning

For building a fronthaul solution it is mandatory to keep into account for three interdependent requirement types: technical aspects, business aspects and regulation constraints [9].

Business and regulation constraints introduce the need to define Service Level Agreements (SLAs) for the fronthaul. Different levels of SLA can be envisaged depending on the chosen fronthaul solution but the basic and necessary one is the capability to monitor the optical link and detect if there are failures.

Technical aspects include CPRI bandwidth, latency on the fronthaul segment and synchronization and jitter. All of them will have an impact on the choice of fronthaul solution and on the dimensioning of the BBU pool. In particular, the fronthaul solution will have to be scalable to support today's CPRI rates of 2.457Gbit/s for each LTE 20MHz 2X2 MIMO sector as well as the 10.137Gbit/s for each sector expected in the future.

Another challenge in the RAN virtualisation is the requirement set to the optical link. The number of RRH on a cell site can be up to 15 considering different radio access technologies with 3 sectors. As fiber is a scarce and expensive resource, wavelength division multiplexing (WDM) is the natural choice to reduce the number of needed fibers between each cell site and the BBU pool location. A preference long-term solution is single fiber optical distribution network (ODN), which allows mutualization with the existing fixed access (Fiber To The Home) infrastructure.

The requirement with expected biggest impact for the dimensioning of the BBU pool is latency in the RAN. As previously mentioned there are limits for the processing of the LTE radio frame, which can be calculated into limits on the time spent in processing and the time spent in transport. The latter limits the maximum distance between cell site and BBU pool location. Figure 7 illustrates the network and equipment segments which have to be considered in an uplink case. For each of these segments, minimum and maximum timing values should be defined for several Mobile network generations. The

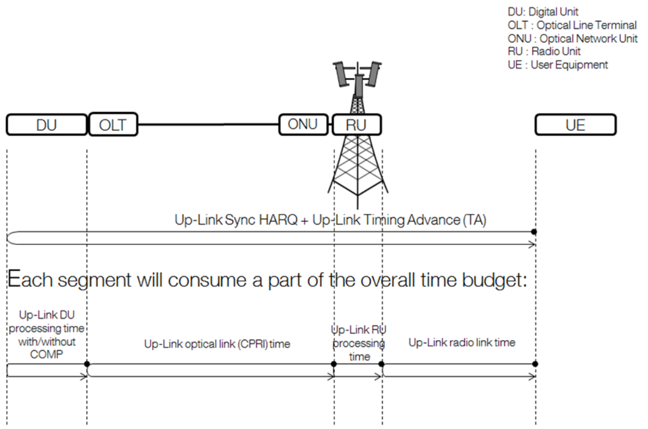

Fig. 7. Network and equipment segments to be considered for fronthaul latency.

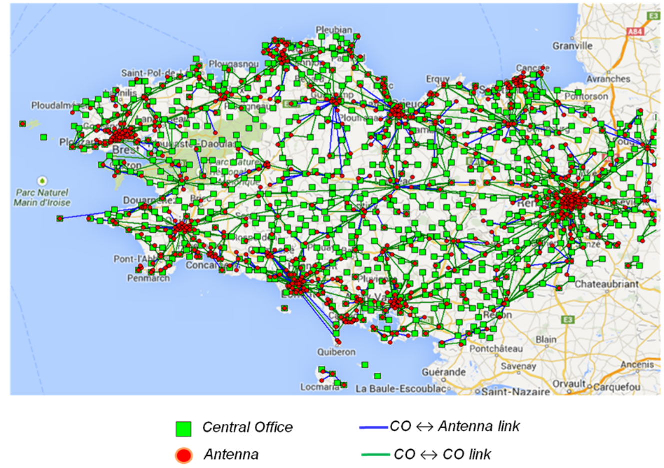

Fig. 8. Antenna sites in Brittany region.

asymmetry timing value (for downlink and uplink) of each segment should be also defined. Moreover, in order to be future proof, these timing values should take into account the most stringent LTE Advanced requirements. As indicated in the figure the HARQ deadline is what dominates the delay calculation since retransmission due to missed deadlines have direct negative impact on peak data rates. After subtracting the mobile equipment processing time (DU=BBU and RU=RRH), considering maximum timing advance (667 μs, for LTE) and the assumption of a 10km cell radius, the remaining time for round trip time propagation between RRH and BBU is in the order of a few hundreds of microseconds.

On the basis of these requirements, an exercise of BBU pool dimensioning has been performed on a real network configuration based on French Brittany region as a good representative of dense urban, semi rural and rural areas. This region covers about 30 000 km^2 with 3.2 M inhabitants. Under a mobile network point of view, the considered area comprises for one operator 860 antenna sites (we make the assumption that antenna sites are multi-RAT: 2G, 3G and LTE). Central offices are divided in 3 types: Central Office (CO), Main CO and Core CO. Over the Brittany region there are 2 mobile Core CO and about 20 main COs. The choice of main COs for BBU pool location seems the best compromise to cover with C-RAN a big part of the population without distributing too much the BBU pools.

Acknowledgements

The research leading to these results was partially funded by EUs Seventh Framework Programme Mobile Cloud Networking MCN project (FP7-ICT-318109) and the Network of Excellence in Wireless Communications NEWCOM# project (FP7-ICT-318306).

References

[1] C. Chen, C-RAN: the Road Towards Green Radio Access Network, in *Proc. of ICST Workshop on C-RAN*, Kunming, China, Aug. 2012.

[2] H. Guan, T. Kolding and P. Merz, Discovery of Cloud-RAN, in *Proc. of NSN Cloud-RAN Workshop*, Beijing, China, Apr. 2010.

[3] A. Pizzinat, P. Chanclou, C-RAN architecture and fronthaul challenges, in *Proc. of LTE World Summit*, Amsterdam, The Netherlands, June 2013.

[4] Mobile Cloud Networking project (FP7-ICT-318109), May 2014 (http://www.mobile-cloud-networking.eu/).

[5] L.S. Ferreira, D. Pichon, A. Hatefi, A. Gomes, D. Dimitrova, T. Braun, G. Karagiannis, M: Karimzadeh, M. Branco, L.M. Correia, Än Architecture to offer Cloud-Based Radio Access Network as a Service, in *Proc. of EUCNC 2014 - European Conference on Networks and Communications*, Bologna, Italy, June 2014.

[6] S. Narnakaje, T̈Is KeyStone architecture enables Layer 2 and transport processing, in *Texas Instruments white paper*, 2010, (http://www.ti.com/lit/wp/spry148/spry148.pdf).

[7] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, John Wiley, West Sussex, UK, 2011.

[8] EURECOM, OpenAirInterface, May 2014 (http://www.openairinterface.org/).

[9] P. Chanclou, A. Pizzinat, et al., Optical fiber solution for mobile fronthaul to achieve CRAN, in *Proc. Future Networks Mobile Summit*, Lisbon, Portugal, June2013.