Cloud RAN Architecture for 5G

A Telefónica White Paper
Prepared in collaboration with Ericsson

WHITE PAPER

Cloud RAN Architecture for 5G

Executive Summary

Increasing traffic demand, limited spectrum availability and mass adoption of mobile broadband are challenging the traditional ways to build cellular networks. In this new environment, mobile operators are seeking new ways to increase network capacity, coverage and user experience while reducing time to market for new services and reduce costs.

To accomplish this, operators need to cost-effectively use all network assets, including multiple standards, frequency bands, cell layers and transport network solutions. This means that, above all, cellular infrastructure must be flexible and support simplified deployment and management of increasingly heterogeneous radio access networks (RANs).

At the same time, in the mobile packet core network, Network Function Virtualization (NFV) has emerged as a viable approach to increase network flexibility in order to, for example, reduce time to market for new services. Making use of open application interfaces and sharing of data centers enables a large number of applications and services to be provisioned cost-effectively over fixed and mobile broadband networks and for many device platforms. The question now is how operators can best make use of virtualization technologies in RANs while at the same time building cost-, spectrum- and energy-efficient networks that offer a seamless user experience.

There are major distinctions, however, between cloud computing in the RAN compared to the core network and service layer. For example, the bulk of the cost of a mobile network lies in the large number of distributed base station and antenna sites, as well as in the last-mile transport network links – not in central nodes and sites. Consequently, the costs associated both with the central parts of the network and its distributed elements and last-mile links must all be taken into account for a cost-effective network evolution.

Future Cloud RAN architectures will therefore exploit a combination of virtualization, centralization and coordination techniques, all of which interact with each other in a variety of ways within the network. Cloud RAN will be composed of a mix of Distributed RAN, Centralized RAN and Virtualized RAN architectures, allowing for spectral efficient solutions over the transport infrastructure available. This paper outlines these architecture options and identifies their key benefits and challenges.

WHITE PAPER

Cloud RAN Architecture for 5G

Telefónica

Table of Contents

Introduction

Mobile network capabilities are evolving quickly, continuously pushed by new requirements relating to latency, traffic volumes, data rates and need for reliable connectivity.

To efficiently meet future demands the LTE RAN architecture will need to support improved resource pooling, capacity scalability, layer interworking and spectral efficiency over various transport network configurations.

Cloud RAN architectures will support these needs by exploiting Network Functions Virtualization techniques and data center processing capabilities, as well as improved radio coordination for distributed as well as centralized RAN deployments.

Telefónica

1. Evolving the radio access network architecture

Mobile broadband is approaching a point where cellular infrastructure –originally designed for mobile telephony– is a viable substitute for fixed broadband in many markets. At the same time, the mass adoption of smartphones and other connected devices is increasing the need for speed application coverage, lower latency and greater capacity in mobile broadband networks.

This evolution is being driven by both the evolution of 4G and the coming introduction of 5G, which will deliver a downlink user experience of 10-100 Mbps everywhere and 1-10 Gbps locally, with a latency of less than 1 ms.

In essence, the main challenges for mobile network infrastructure to meet this need over the next three to five years will be to:

- manage large amounts of new spectrum, primarily on higher bands (>4 GHz) and then combined with existing bands, so that by 2020 a typical mobile network operator could have access to more than 100 MHz in five to ten bands;
- deploy new sites to support new use cases which may also require new business models; and
- develop efficient hardware and software solutions that enable the speedy introduction of new services in a sustainable way, from both an energy and cost perspective.

5G use cases such as enhanced mobile broadband (MBB) (see Figure 1) may also drive the further evolution of the RAN architecture.
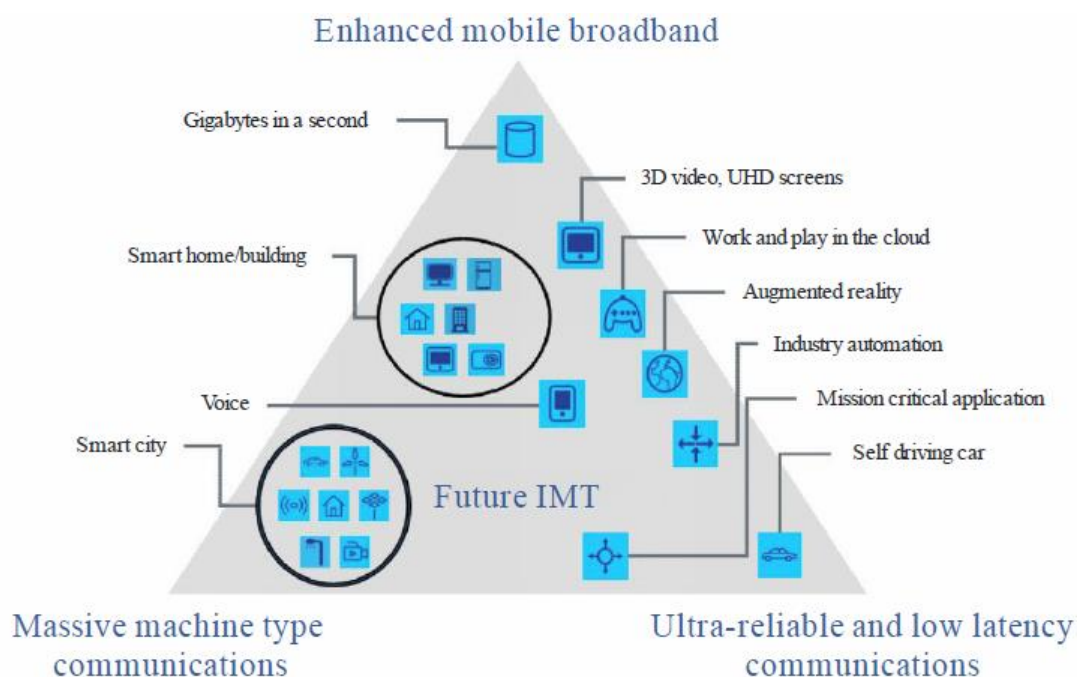

Figure 1. 5G Use Cases (taken from ITU-R [5]).

In this paper we focus on the evolution of radio access network architectures towards 5G. For this purpose we analyze the high level requirements on the RAN architecture associated to current

Radio Access Technologies (mainly LTE) and envisaged future 5G technology. Further, we review the different RAN architecture options, ranging from the traditional Distributed and Centralized RAN architectures– to new options enabled by Network Function Virtualization and improved transport network capabilities. Finally, we analyze how the new architectural options allow addressing the 5G needs.

2. RAN architecture evolution drivers

From a network architecture perspective, the main options for LTE up to the present have been either fully distributed or fully centralized baseband deployment, each with its own pros and cons.

In essence, the choice of architecture is a tradeoff between:

— need for spectral efficiency and user experience,

— efficient utilization of equipment and sites, and

— cost and availability of 'last mile'-transport network connectivity.

These needs in turn drives the need for coordination (between cells and bands), centralization (of hardware), and virtualization (of software) that are discussed further next.

2.1. Coordination

Going forward it is anticipated that radio access networks will become more heterogeneous, composed of multiple layers with different cell sizes and bandwidths, which calls for a tighter interworking between technologies and cell layers to ensure a seamless use experience and maximum spectral efficiency.

Such interworking, or radio coordination, can in LTE today be achieved in a distributed RAN architecture as well as in a centralized RAN architecture.

Radio coordination between cells and bands is becoming increasingly important to maximize spectrum efficiency and user experience. Generally speaking, radio coordination mechanisms can be categorized as:

1. Mobility management (handover)
2. Traffic management (load balancing)
3. Interference management (interference control)
4. Joint reception and transmission
5. Carrier Aggregation
6. Dual Connectivity

2.2. Centralization

Centralizing base station processing has several advantages from operational, hardware and spectrum efficiency point of view.

First, it simplifies network management, upgrades and troubleshooting due to less site visits.

Second, it enables hardware resource pooling: thanks to statistical multiplexing an execution platform can perform the same tasks with less hardware or capacity.

The highest degree of pooling is achieved with a fully centralized baseband approach, with a star connection long-haul Common Public Radio Interface (CPRI) between the pooled baseband and the distributed remote radio heads. This is because processing of the lower layers constitutes such a large part of the computational effort. However, the potential gains in hardware pooling may in practice be outweighed by the (high) cost for long-haul CPRI. Furthermore, the computational complexity with a completely pooled common baseband approach will prohibit scalability of each "baseband pool", see Figure 2.
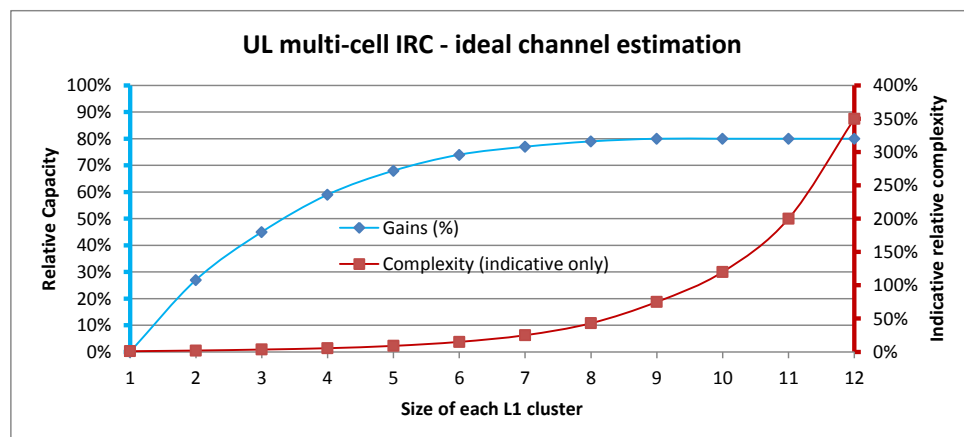
Figure 2. Example of how the relative network capacity gains with coordination (here Uplink CoMP) diminish with the size of the common baseband cluster, and how the computational complexity at the same time increases exponentially with the cluster size.

Also from a hardware utilization point of view, the statistical multiplexing gain achieved can be showed to diminish after pooling of 10-100 cells (depending on cell load distribution and quality of service requirements).

A selective centralization of higher radio resource control and user plane handling as well as improved coordination across baseband units may however still provide substantial benefits and will be discussed further in this paper.

2.3. Virtualization

Work in Network Functions Virtualization (NFV) [4] was initiated in order to solve some of the main challenges network operators face today when trying to deploy new services following the traditional approach. They include, among others, excessive time to market for the deployment of new services, increasing costs of energy, capital investment challenges and the rarity of skills necessary to design, integrate and operate increasingly complex hardware-based appliances that support them, and constraint of innovation in an increasingly network-centric connected world. NFV aims to address these problems by leveraging standard Information Technologies (IT) virtualization techniques to consolidate many network equipment types onto industry standard high volume servers, switches and storage. This equipment could be located either in data centers, network nodes, or at end user premises.

Network Function Virtualization has entered the telecom networks starting in the packet core network. As a natural evolution, it is now investigated if virtualization techniques can be applied also in the radio access network.

The basic idea is to execute RAN functionality on more generic and generally available execution hardware and software platform, together with cloud core applications and other latency-critical services, sometimes even in a virtualized Platform as a Service (PaaS) environment.

However, the lower layers of the RAN protocol stack are real-time critical. Many of the time synchronization requirements that ensure the performance of the radio access protocol are on the

microsecond level and, in some cases, the nanosecond level. Contrary to server-type functions, RAN functionality is thus not easily hosted by the PaaS model.

The question at hand is hence: what parts of the RAN functionality are viable to move to a virtualized environment? Will the needs and prerequisites change when 5G is introduced?

3. Future radio access network challenges

This section provides an overview of the expected challenges for 5G radio access and the impact it may have on the radio access network architecture.

5G standardization is at the time of writing (2016) in an initial phase in 3GPP and fundamental parts of the radio interface have not yet been specified. In this paper the working assumption is however that 5G will be developed in two standards:

- The evolution of LTE, aimed at enhanced functionality while securing backwards compatibility with LTE.

- A new radio access technology aimed primarily for spectrum bands where LTE is not deployed. In this paper we refer to the new radio access technology as "5G RAT" or "NR", the latter being the working name in 3GPP fora.

The NR concept has been designed to meet all the foreseen 5G system requirements including new use cases as well as a wide range of spectrum bands and deployment options.

3.1. Spectrum for 5G

One of the primary goals of NR is the ability to cope with RF carriers having significantly wider bandwidths than existing cellular technology, ranging from several hundreds of MHz up to a few GHz.

Such an enormous amount of spectrum can only be released for cellular usages at frequencies in higher bands, well beyond 6 GHz, which sometimes is referred to as "millimeter-wave spectrum". The industry is currently looking at all frequency bands from 6 GHz up to 100 GHz for 5G, and especially spectrum bandwidth that currently is unused or under-utilized (by non-cellular incumbents like satellite service providers or military players). The propagation conditions at such high carrier frequencies are however unfavorable, which needs to be mitigated for these bands to be usable in cellular networks.

The frequency bands targeted include:

- Full coverage layers at lower frequency regions below 6 GHz.

- Partial coverage layers at higher bands, up to 30 GHz or even beyond where large bandwidths are available.

While licensed spectrum remains a cornerstone for 5G wireless access, unlicensed spectrum (stand-alone as well as license-assisted) and various forms of shared spectrum will be natively supported.

3.2. Physical layer and radio resource management concepts

To enable use of high carrier frequencies Massive MIMO and beamforming techniques will be used to extend the reach as much as possible.

In addition to beamforming, transmission and reception using multiple access points simultaneously may be used to reduce the chances of suffering a radio link failure in standalone deployments. Such techniques however introduce additional complexity and overhead.

Alternatively, a reliable link can be maintained through legacy infrastructure at a lower frequency (e.g. a 4G access point), thus leading to non-standalone deployments. Legacy LTE control schemes

are proven to be robust enough for these tasks and, at least conceptually, an LTE layer could control the radio resources of the higher-frequency access layer. Both standalone and non-standalone deployments are sketched in Figure 3.
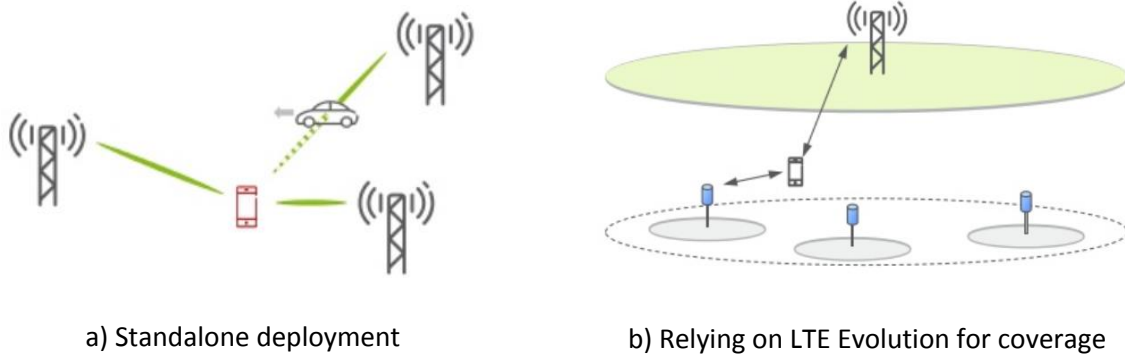

a) Standalone deployment   b) Relying on LTE Evolution for coverage

Figure 3. Multi-connectivity examples foreseen in 5G.

To achieve a tight interworking with LTE the eNBs (that are LTE, NR or LTE+NR capable) are connected to each other and to the core network via new RAN interfaces, to be standardized. Figure 4 shows the high level logical architecture for a system supporting both NR and LTE.

Together with the spectrum harmonization, NR will have to support lower latency, which requires shorter and more flexible Transmission Time Intervals (TTIs), new channel structure etc.

Both FDD and dynamic TDD, where the scheduler assigns the transmission direction dynamically, are part of NR, however most practical deployments of NR will likely be in unpaired spectrum.

"Ultra-lean design", where transmissions are "self-contained" with reference signals transmitted along with the data, minimizes broadcasting of signals resulting on significantly improved energy efficiency.

In summary, to support all use cases, frequency bands and deployments, NR needs to be very flexible in its design.
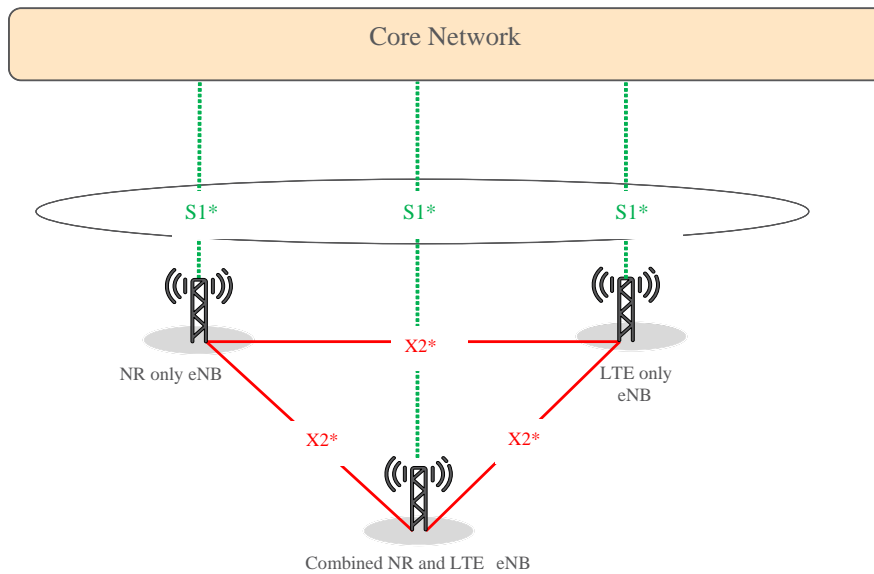
Figure 4. High-level logical architecture for NR and LTE.

3.3. Network slices

5G is expected to support a wide range of services and associated service requirements in a wide range of scenarios. One way to address these different use cases efficiently is through the use of network slicing (Figure 5).

Network slicing is an end-to-end concept where the user or operator of a network slice (e.g. an MTC sensor network) sees the network slice as a separate logical network having similar properties of a dedicated network (e.g. separate management/optimization), but in fact realized using a common infrastructure (processing, transport, radio) which is shared with other network slices. Physical network resources are separated from the logical network using the principles of Network Function Virtualization (NFV) and Software Defined Networking (SDN).
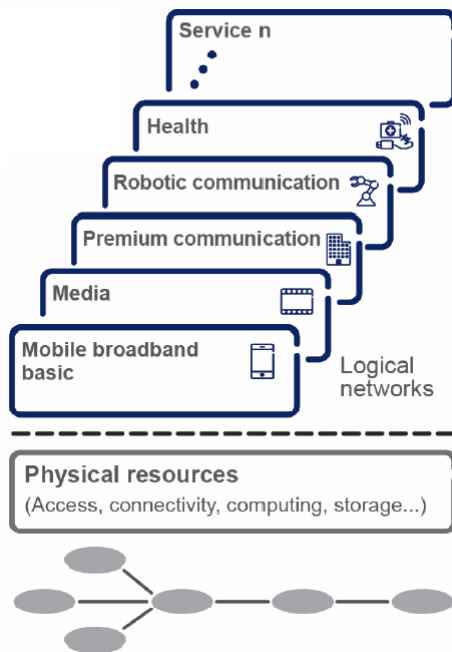
Figure 5. Support of multiple network slices acting as independent networks over the same physical resources.

3.4. LTE Evolution into 5G

As described above, the evolution of LTE is an integral part of 5G and tight interworking between LTE and NR is envisioned at least in the early stages of deployment of 5G. As a consequence, some of the 5G features described above are already being promoted as part of the evolution of the LTE standard.

LTE Release 13 specifications are about to finish and the following is a brief summary of some of its most relevant features that may influence 5G/NR architecture:

1. *Active Antenna Systems (AAS), and associated SON techniques*: AAS systems have the ability to dynamically adjust the radiation pattern so as to introduce cell split, beamforming, and dynamic sectorization in the vertical and horizontal planes. These techniques can be of importance in ultra-dense deployments for proper interference management. Exchanging the necessary control information between neighbor nodes may impact 5G architecture and should therefore be taken into account if AAS systems are to be embraced by 5G.

2. *Elevation beamforming and Full-Dimension MIMO (FD-MIMO)*: AAS systems can be regarded as the basis for so-called FD-MIMO systems, where 3D multi-user MIMO techniques are investigated. An analysis on how the complexity scales with the number of antennas would be of utter importance here, as it might impact design choices like how many antennas to consider or the use of distributed vs. centralized architectures.

3. *Enhanced signaling for inter-eNB Coordinated Multipoint (CoMP)*: Signaling procedures for inter-eNB CoMP are introduced to exchange control information among the nodes in the coordinated set, assuming a distributed approach with no central coordinating node. Such

Telefónica

enhancements can be further re-used in 5G/NR ultra-dense architectures making use of coordination techniques.

4. *Licensed Assisted-Access (LAA) using LTE[1]*: The interest of LAA in 5G architectures can be high because initial 5G rollouts will likely rely on LTE carriers for coverage and control. Such cross-carrier control mechanisms would have to be extended in this case so as to control non-LTE (e.g. NR) carriers.

5. *Dual connectivity enhancements*: Dual connectivity will play a key role in 5G architectures. Even in the case of standalone deployments, where control procedures operate autonomously as part of the 5G network, multi-connectivity is likely to play a key role at higher frequencies for improved resiliency or data rate.

Much of the ongoing technical work in Release 13 is intended to continue during Release 14. In parallel, additional work is planned (in the form of Study Items) that can be relevant for 5G architecture. The reader is referred to [1] for further details on them.

---

[1] LAA studies the use of LTE in unlicensed spectrum, as a complement to networks in licensed spectrum. Mobility and critical control signaling hence rely on licensed spectrum carriers, while less demanding traffic can be handled by an unlicensed spectrum carrier in an opportunistic way.

4. Overview of RAN architectures

The following sections provide an overview of each of the architectural options being considered for 5G networks.

4.1. Distributed RAN

In a fully distributed baseband deployment, the interface between the RAN and core network is located at the radio site. Today, most LTE networks use a distributed baseband deployment only. In fact, one of the key advantages of LTE has proven to be its flat architecture, which enables quick rollout, ease of deployment and standard IP-based connectivity (Figure 6).
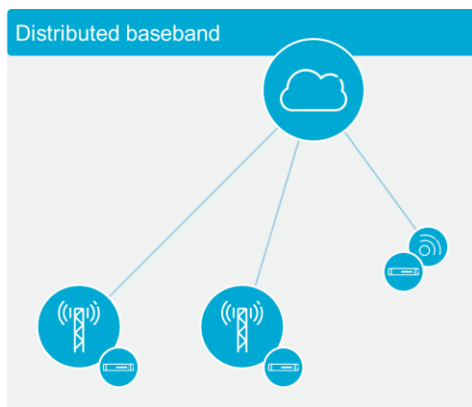

Figure 6. A distributed RAN with distributed baseband deployment.

Baseline X2 coordination

Thanks to collaboration between base stations over the IP-based X2 interface, LTE handovers remain seamless from a user perspective. In addition to basic mobility and traffic management functionality, X2 coordination is evolving to support carrier aggregation and coordinated multipoint reception (CoMP) across sites and layers – see Figure 7.

Baseline X2 coordination features include Automatic Neighbor Relations (ANR) and Reduced Handover Oscillations, Load Balancing, etc.
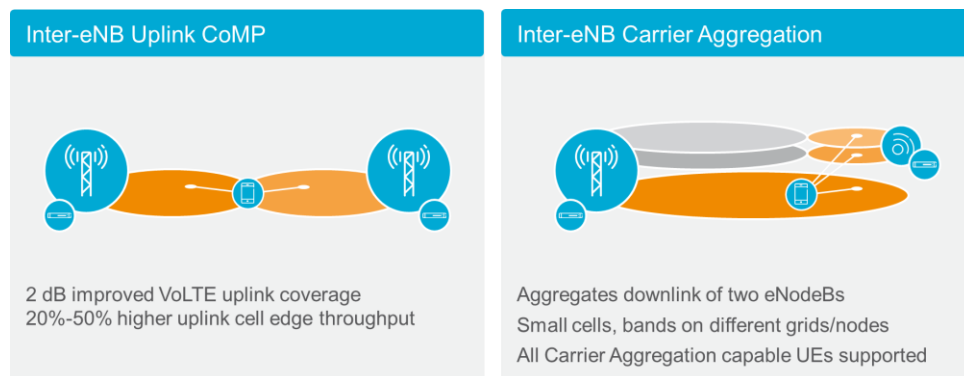

Figure 7. Examples of Tight X2 coordination features.

Tight X2 coordination – between eNodeBs with short latency

In addition to basic mobility and traffic management functionality, X2 coordination is evolving to support carrier aggregation and coordinated multipoint reception (CoMP) across sites and layers.

If the backhaul latency is in the order of a few milliseconds, features like carrier aggregation and CoMP can be supported over the X2 interface.

Intra-site common baseband coordination

On top of the fully distributed topology with collaborative functionality over X2 it is straightforward to exploit common baseband for co-sited sectors & cells. This enables advanced joint signal processing – including combining signals from several sectors, and interference mitigation mechanisms – which will increase performance. Examples of features include carrier aggregation and CoMP.

Aside from improving efficiency in the coordinated cells, inter-cell interference is reduced which as a side effect improved performance also in the surrounding cells. Combined with the inter-site X2 coordination the overall network performance is sufficient for many scenarios.

4.2. Centralized RAN

To boost performance in traffic hotspots such as offices, stadiums, city squares and commuter hubs, centralized baseband deployments have become increasingly interesting for operators.

In a fully centralized baseband deployment, all baseband processing (including RAN L1, L2 and L3 protocol layers) is located at a central location that serves multiple distributed radio sites – see Figure 8. The transmission links between the central baseband units and distributed radio units use CPRI fronthaul over dedicated fiber or microwave links. This CPRI fronthaul requires tight latency and large bandwidths.
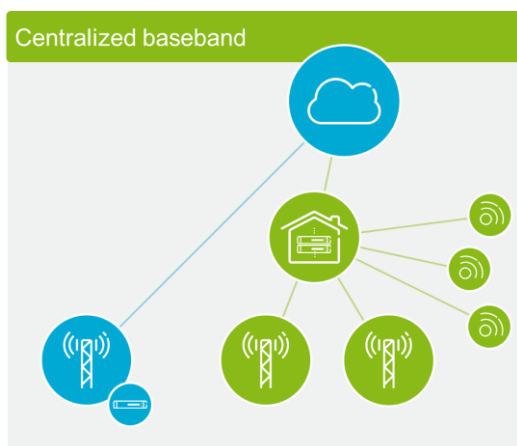

Figure 8. Centralized baseband deployment (green) complementing a distributed baseband deployment (blue).

State-of-the-art signal processing technology can enable large centralized baseband configurations that host a number of remote radio units. These remote units are fully coordinated with joint

transmission and reception across all antenna elements, cells and bands.

The potential for better performance with a fully centralized baseband deployment is unmatched: the downlink data rates at cell edge can in such highly loaded scenarios be improved up to 40%-70%, enabled by coordinated scheduling functionality, and uplink cell edge data rates can be improved by up to 2-3 times or more depending on interference levels and signal strength. The gain in coverage driven scenarios is uplink coverage; cell edge throughput can be improved by a factor 2 thanks to uplink coordinated multipoint reception.

However, in many situations, CPRI connectivity requirements will be too strict for Centralized RAN architectures to be affordable.

4.3. Mixing Distributed and Centralized RAN deployments

Going forward many networks will likely consist of a combination of distributed and centralized baseband deployments, mainly depending on availability of fiber and performance needs. (Note that the cost of transport network connectivity is similar regardless of the use of CPRI compression and/or CPRI over Ethernet, since it is related to the strict transport delay requirement.)

For a common baseband to be more widely adopted, the cost of fronthaul hence needs to drop significantly. Alternatives to fiber based CPRI, including microwave solutions and other options that enable somewhat relaxed fronthaul requirements are being investigated in the industry. Common baseband could also be used in the future for coverage limited deployments in suburban and urban areas. Primarily as a way to extend uplink range and enable carrier aggregation in a flexible fashion across layers having non-uniform coverage.

As a further evolution of Centralized RAN architectures, an interesting option is also to connect baseband units at the L1/L2 level of the protocol stack, as opposed to X2 – interworking on a Radio Resource Control (RRC) and Packet Data Convergence Protocol (PDCP) – and CPRI operating on I/Q antenna stream level. With such an interworking the baseband units can be interconnected through a high-speed, high-quality Ethernet switched network which is much more efficient than dedicated, point-to-point fiber connections required for CPRI links in typical C-RAN deployments. Full performance benefits can then be achieved although cells are not hosted in the same baseband unit.

With such a tight L1/L2 interworking, baseband units can be aggregated in a fully meshed fashion, enabling borderless coordination across centralized as well as distributed baseband deployments. The end-user will always benefit from coordination features like carrier aggregation and CoMP throughout the entire network, even when covered by different cell sites that have different baseband units.

Further, as baseband units can be geographically separated and the architecture is truly meshed, the network can be migrated stepwise as the need for capacity diffuse from the inner city traffic hot spots to a wider area.

4.4. Virtualized RAN

Distributed and Centralized architecture have served the industry well for the currently deployed 4G networks. However when introducing high bandwidth layers with partial coverage in 5G, as previously discussed, there is a need to revisit the RAN architecture.

Virtualized RAN is addressing the challenges brought on by the vastly different throughput

capabilities and limited coverage exhibited by this new spectrum.

Key aspects of Virtualized RAN

The Virtualized RAN architecture exploits NFV techniques and data center processing capabilities and enables coordination and centralization in mobile networks, as summarized in Figure 9.

The Virtualized RAN architecture supports:

- resource pooling (cost-efficient processor sharing),
- scalability (flexible hardware capacity expansion),
- layer interworking (between the application layer and the RAN), and
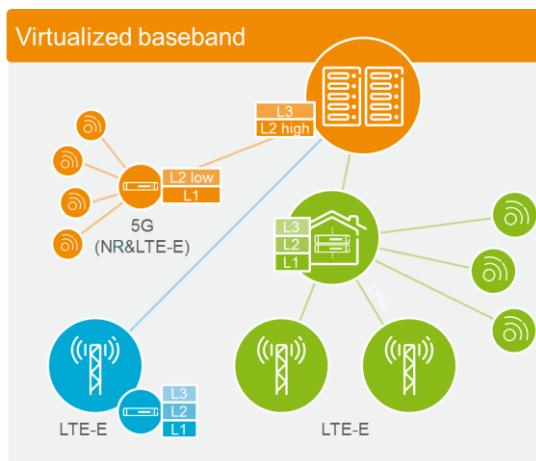- robust mobility.

Figure 9. In a Virtualized RAN (parts of) the baseband functionality will be hosted in a separate, data center, processor environment.

Virtualized RAN can be viewed in several different ways and there are many different and complementary aspects and benefits that can be considered. However, a key aspect of Virtualized RAN is the fact that certain benefits can be achieved by the split and separation of the higher asynchronous layers of the radio access protocol stack.

The main benefit of separating higher and lower layers of the RAN protocol stack into separate nodes ("functional split") is related to the need for tight interworking between small and large cells on different frequencies and on different deployment grids. By allowing for a tight interworking, transport network resources can be used more effectively and the high bands with partial coverage can be used as much as possible (while ensuring a reliable connectivity through the lower band).

Another benefit of Virtualized RAN is that the functionality that has been separated out from the baseband unit and virtualized, running on generic hardware, can benefit from more flexible scalability of capacity, co-hosting with Core Network functionality, and features provided through the NFV framework and future implementations of so called network slices.

Control plane

Virtualized RAN allows operators to centralize the control plane (seen together with PDCP split in Figure 5) –which does not have extreme bitrate requirements– to bring RAN functionality closer to applications.

Cloud core and NFV frameworks also bring applications closer to the RAN, and this proximity enables scalable and shared common and commercial-off-the-shelf (COTS) execution platforms to be used and leveraged for cost-effectiveness and flexibility. For instance, if cloud core function are pushed out into the network and RAN is centralized to some degree, there will eventually be some degree of colocation of core and RAN functionality – either with RAN and core together on a server in a distributed fashion, or with RAN and core executing in a centralized data center environment. This will enable substantially lower latencies for the interconnection between RAN and core.

This kind of selective centralization of the control plane –shown in Figure 6– can provide user experience benefits such as mobility robustness, while spectral efficiency can be ensured through a level of radio resource coordination across radio sites.

User plane

From a user plane perspective, Virtualized RAN can also provide optimization benefits for certain deployment scenarios driven by dual connectivity needs. With dual connectivity in a fully distributed deployment, data can be routed first to one site and then rerouted to the second site.

This results in what is referred to the "trombone effect" in the transport network, which means that data is sent inefficiently back and forth over the same transport network segment. This can be avoided by placing the routing protocol higher up in the transport network aggregation hierarchy, which improves user plane latency.

The L2 user plane layer (PDCP) is predominantly a routing protocol, but it also includes a fair amount of processor-heavy ciphering. Optimized ciphering accelerators can be used to provide a low-latency and high-bandwidth performance implementation in a more energy- and cost-efficient way, as a complement to a more generic packet data processing environment.

5. 5G architecture design aspects and open issues

In the previous sections we have identified the new challenges on the RAN from the new use cases supported by 5G, as well as the expected characteristics of the 5G RAT. We have also analyzed the different architectural options for the RAN. Now, in this section we analyze how the different architectural options available may help to meet the new 5G requirements on RAN.

5.1. RAN functional splits

The functional split between centralized and distributed RAN nodes is important as it leads to varying interfaces and requirements for the transport network, as well as different possibilities for cells coordination at control- and data-plane levels.

Centralization of RAN functions enables smart pooling of resources in multi-cell environments where not all the cells are likely to demand full computing capabilities at the same times. Moreover, centralization makes it easier to perform joint radio resource management (JRRM) techniques without costly data shuffling among the nodes.

An initial possibility is to locate the split point somewhere inside the physical (PHY) layer. In this case, CPU-intensive tasks (for which little pooling gain can be expected) may run in a distributed way while the remaining tasks can benefit from centralization and eventual pooling gains. One example could be locating the split point between the Precoding and Resource Mapping steps in an LTE-like PHY processing chain (Figure 10, left). The attractiveness of this option is that it enables techniques such as distributed massive MIMO or CoMP without heavy data exchange among the nodes. The fronthaul traffic can be based on a frequency-domain description of the signals, exploiting the inherent trunking gains resulting from the aggregation of multiple traffic-dependent flows, hence alleviating the transport requirements. However, fronthaul traffic rates would not be constant over time which complicates the resulting interfaces.

Another choice for the split point could be the boundary between PHY and MAC layers (Figure 10, middle). The advantage in this case is the lower resulting fronthaul rates, as only transport block bits need to be exchanged with much reduced capacity requirements. However only MAC-level (and above) functions would be centralized and coordination possibilities are limited compared to intra PHY split.

Alternatively, a third split point could be defined at the uppermost level of the data plane protocol stack, namely the PDCP layer (Figure 10, right). This functional split enables multi-connectivity by splitting the traffic into multiple flows directed towards different access nodes. PDCP centralization has the additional benefit of exploiting eventual pooling gains from CPU-intensive header compression protocols (like Robust Header Compression, ROHC), which can benefit from statistical multiplexing gains at the aggregation point.
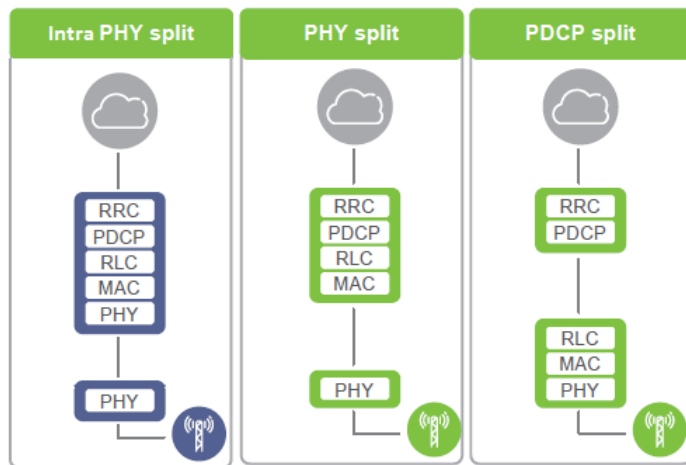
Figure 10. Key options for functional splits: Intra PHY split (left); PHY-MAC split (middle); PDCP split (right).

Whatever the functional split point is defined, it is essential that the architecture ideally supports all the possibilities by leveraging on generic interfaces with varying degrees of traffic multiplexing and/or routing capabilities. Significant progress remains to be done on RAN architecture and interfaces so that functions can be flexibly instantiated according to any suitable definition of the corresponding network slices.

5.2. Transport network aspects

Beyond the expected benefits of the above described functional split options, it is to note that introduction of any of them in a real network will come at a twofold price:

- New physical interfaces ought to be defined at the split points.
- Stringent requirements at the transport network would have to be met in order to enable seamless operation of the RAN protocol stack, by fulfilling the throughput and timing requirements set by the air interface protocols, frame structure, and numerology.

Definition of interfaces between network functions is expected to be complex and falls outside the scope of this White Paper [2]. Some guidelines can however be given on the fronthaul requirements that would arise when defining a given functional split. Referring to the three possibilities described in section 5.1, the following high level observations can be made:

1. Intra-PHY split point: This choice will likely demand high throughput values at the fronthaul network, although smart definition of the interface could yield significant throughput savings compared to CPRI (with rates of several Gbps per sector), eventually enabling statistical multiplexing. In terms of latency, current implementations and even evolutions of CPRI are well below 1 ms one-way delay [2], [3]. This stringent requirement applies to any split option located below the HARQ level, i.e. when HARQ is part of the set of centralized processing functions.

2. PHY-MAC split point: Throughput would be greatly reduced in this case compared to intra-PHY split, as a result of carrying transport bits (with rates of several hundreds of Mbps per sector) instead of conveniently processed PHY-layer samples. Latency would however stick

to the same tight boundaries because of the HARQ cycle.

3. PDCP split point: The potential attractive of this option would be the much relaxed latency requirements compared to the previous ones, in the order of several tens of ms, as in today's backhaul links. Throughput figures would not be much different from those in PHY-MAC split.

From the above considerations, no single optimal solution can be found that meets the trade-off between RAN performance and transport network requirements. Realistic deployments will therefore likely have to adapt to the available transport infrastructure, on a case-by-case basis. For this reason 5G networks should ideally support different functional splits.

6. Conclusions

Opportunities for new mobile services and deployment use cases, in conjunction with a growing concern for economic sustainability, will challenge the traditional LTE RAN architecture as networks evolve into 5G.

To meet this demand, the following developments will drive the need for more flexible radio access network architectures:

- Deployments of heterogeneous networks with a mix of macro and small cells, and new bands with substantially different coverage.
- The adoption of NFV into mobile core networks.
- The need for decoupled and independent scalability of processing capacity for different RAN protocol layers.
- The need to support a variety of transport network capabilities.

The industry is now looking into how to evolve the LTE RAN architecture into a Cloud RAN, by exploiting coordination, centralization and virtualization techniques.

A promising concept is to implement parts of the RAN protocol stack as Virtual Network Functions (VNFs) in a data center environment; introducing a so called Virtualized RAN architecture.

Aside from benefits of NFV, such as the capability to instantiate network functions on demand, a split architecture introduced with Virtualized RAN will improve interworking between cells and layers. Such radio coordination will especially be important when introducing 5G in higher frequency bands that may have partial coverage. In practice Cloud RAN will enable different levels of radio coordination depending on the degree of RAN centralization and the capabilities of the transport network infrastructure.

To conclude, Cloud RAN will therefore be an important tool to support the introduction of 5G infrastructure and services in a cost effective way.

Key takeaways

- The support of new mobile services under strong economic constraints will challenge the traditional LTE RAN architecture

- The industry is looking into how to evolve the LTE RAN architecture into a Cloud RAN

- A promising concept in Cloud RAN is to implement parts of the RAN as Virtual Network Functions in a data center, leading to so-called Virtualized RAN

- A split architecture with Virtualized RAN will improve interworking between cells and layers with different levels of radio coordination

- Cloud RAN can therefore improve the economic sustainability of the networks and become an important tool for the introduction of new 5G services

Testimonials

Telefónica

"5G is an exciting journey towards defining the network of the future. That is the reason why Telefonica continues advancing its role as a key player in the development of 5G technologies, contributing with White Papers like this in advancing future architectures and options to cope with future demands. The support of new mobile services under strong economic constraints will challenge the traditional LTE RAN architecture. This White Paper shows how Cloud RAN can improve the economic sustainability and become an important tool for future 5G networks."

Enrique Blanco, Global CTO Telefónica

Ericsson

"The new capabilities of 5G span several dimensions: lower energy requirements, greater capacity, bandwidth, security, reliability and data rates, as well as lower latency and device costs. The architecture that allows for this tremendous flexibility is of great importance for the industry, and we believe that Cloud RAN architectures will satisfy these demands. The collaboration with Telefonica in this area is extremely important for Ericsson as it allows us to understand the true requirements of a global operator, which helps us to focus on providing the right solutions at the right time."

José Antonio López, CEO Ericsson Iberia

Acronyms

3GPP	3rd Generation Partnership Project		**LTE**	Long-Term Evolution
4G	4th Generation		**MAC**	Medium Access Control
5G	5th Generation		**MBB**	Mobile Broadband
AAS	Active Antenna Systems		**MIMO**	Multiple Input Multiple Output
ANR	Automatic Neighbor Relations		**MTC**	Machine-Type Communications
C-RAN	Cloud-RAN		**NFV**	Network Function Virtualization
CoMP	Coordinated Multi-Point		**NR**	New Radio
COTS	Commercial off-the-shelf		**PaaS**	Platform as a Service
CPRI	Common Public Radio Interface		**PDCP**	Packet Data Convergence Protocol
CPU	Central Processing Unit		**PHY**	Physical Layer
eNB	Evolved NodeB		**RAT**	Radio Access Technology
FDD	Frequency Division Duplex		**RAN**	Radio Access Network
HARQ	Hybrid Automatic Repeat Request		**RF**	Radio Frequency
IP	Internet Protocol		**ROHC**	Robust Header Compression
IRC	Interference Rejection Combining		**RRC**	Radio Resource Control
IT	Information Technologies		**SDN**	Software Defined Networking
ITU	International Telecommunications Union		**SON**	Self-Organizing Networks
			TDD	Time Division Duplex
JRRM	Joint Radio Resource Management		**TTI**	Time Transmission Interval
LAA	Licensed-Assisted Access		**VNF**	Virtual Network Function

References

[1]   RP-151210 Draft Report of 3GPP TSG RAN meeting #70, Sitges (Spain), December 7-10, 2015.

[2]   China Mobile Research Institute, "White Paper of Next Generation Fronthaul Interface", V 1.0, June 2015.

[3]   A. Checko, "Cloud RAN fronthaul", iJOIN Winter School "5G Cloud Technologies: Benefits and Challenges", Bremen, Feb. 2015.

[4]   Joint-operator white paper introducing NFV published October 2012: http://portal.etsi.org/NFV/NFV_White_Paper.pdf

[5]   ITU-R M.2083-0, "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond", September 2015.