ERICSSON

CLOUD RAN

THE BENEFITS OF VIRTUALIZATION, CENTRALIZATION AND COORDINATION

Mobile networks are evolving quickly in terms of coverage, capacity and new features, continuously pushed by new requirements relating to latency, traffic volumes and data rates. By introducing Cloud RAN architectures, operators will be able to meet these accelerating demands through the use of Network Functions Virtualization techniques and data center processing capabilities in their networks, which allows for resource pooling, scalability, layer interworking and spectral efficiency.

INTRODUCTION

Increasing traffic demand, limited spectrum availability and mass adoption of mobile broadband are challenging traditional ways of building cellular networks. In this new environment, mobile operators are seeking new ways to increase network capacity and coverage while reducing time to market for new services and achieving lower total cost of ownership (TCO).

To accomplish this, operators need to cost-effectively combine multiple standards (GSM, CDMA, WCDMA, LTE and Wi-Fi, among others), frequency bands, cell layers and transport network solutions, while at the same time reducing latencies and handling substantially increased data rates. This means that, above all, cellular infrastructure must be flexible and support simplified deployment and management of increasingly heterogeneous radio access networks (RANs).

While current RANs support a large variety of transport network and baseband configurations, ranging from high-capacity macro base stations to low-cost pico base stations, potential synergies with the emerging platforms for Network Functions Virtualization (NFV) and availability of high performance data center solutions have yet to be exploited.

In the mobile packet core network, 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, to calculate the return on investment benefit of implementing Cloud RAN, 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.

A Cloud RAN architecture 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.

This paper outlines future Cloud RAN architecture options and identifies their key challenges and opportunities.

MEETING TRAFFIC DEMAND

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 100Mbps everywhere and 1-10Gbps locally, with a latency of less than 1ms.

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

> manage large amounts of new spectrum, primarily on higher bands (> 4GHz) and then combined with existing bands, so that by 2020 a typical mobile network operator could have access to more than 100MHz in five to 10 bands
> deploy new sites to support new use cases – such as wireless offices and connected cars – which may also require new business models
> 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 and technologies, such as critical machine-type communication and full dimension multiple-input, multiple-output (FD-MIMO) – also known as massive beamforming – may also drive the further evolution of the RAN architecture, exploiting advances in backhaul and antenna systems, among other things – see Figure 1.
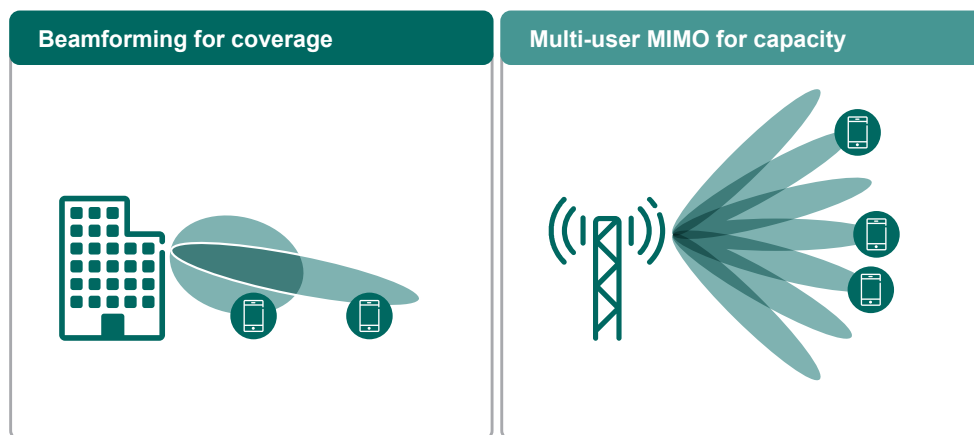

Figure 1: Examples of antenna techniques that increase requirements on Common Public Radio Interface (CPRI) and L1, and which couple L1 more closely to the antenna site.

CURRENT RAN ARCHITECTURES

RANs have evolved tremendously over the past decade. State-of-the-art radio access technologies (RATs) currently support:

> large spectrum bandwidths, multi-band carrier aggregation and multiple-input, multiple-output (MIMO) antennas
> tight radio coordination between sites and cells, including mobility and interference management
> a wide range of base station configurations, including base station hotels, non-intrusive metropolitan macro cells, energy-efficient coverage site solutions, and micro and pico cells
> flexible last-mile transport, ranging from best-effort IP transport connectivity to microwave links and fiber fronthaul for main-remote and centralized baseband deployments
> modularized software architectures that enable multiple RATs to share the same processing resources and that are scalable to large and small base station deployments.

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

DISTRIBUTED BASEBAND

Today, most LTE networks use a distributed baseband deployment only. In a fully distributed baseband deployment, the interface between the RAN and core network is located at the radio site. 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.

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 multi-point reception (CoMP) across sites and layers – see Figure 2.
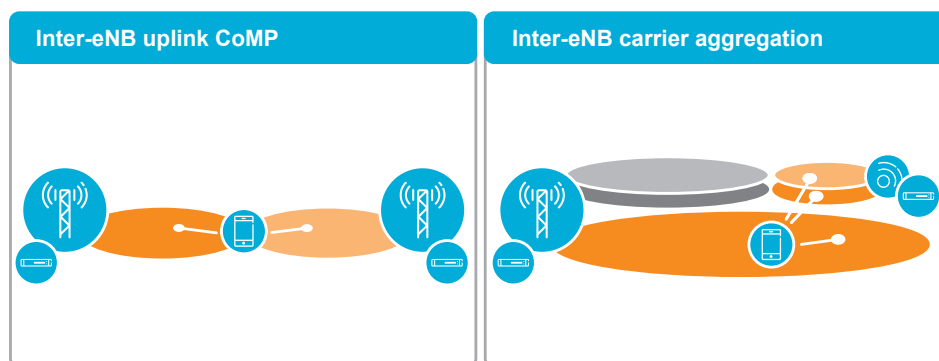

Figure 2: Examples of valuable coordination features possible with a fully distributed deployment.

CENTRALIZED BASEBAND

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 3. The transmission links between the central baseband units and distributed radio units use Common Public Radio Interface (CPRI) fronthaul over dedicated fiber or microwave links. This CPRI fronthaul requires tight latency and large bandwidths.

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. However, in many situations, CPRI connectivity requirements will be insurmountable, especially considering likely future 5G requirements such as extreme data rates and massive beamforming, which will feature many more individually-controlled antenna streams.

With this limitation in mind, most future networks will likely consist of a combination of distributed and centralized baseband deployments, mainly depending on availability of fiber and performance needs.

In order to properly benefit from centralization, operators will need a RAN architecture that both takes advantage of the strengths of a centralized architecture and allows for more affordable transport solutions. This is where the Cloud RAN concept can help.
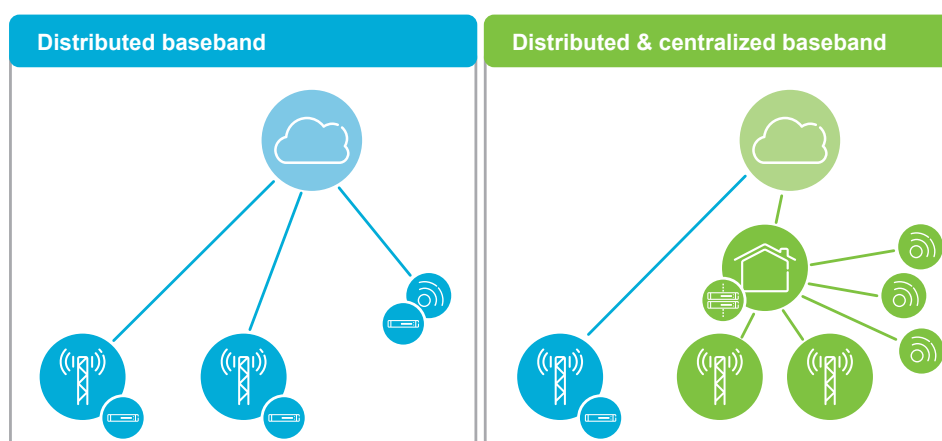

Figure 3: Current fully distributed baseband deployments versus a mixture of distributed and centralized baseband, and corresponding transport links (IP and CPRI) for small and large cells.

CLOUD RAN ARCHITECTURE FRAMEWORK

Driven by greater needs for coordination as well as increasing resource efficiency and advances in network virtualization, Cloud RAN architecture allows for the use of NFV techniques and data center processing capabilities such as coordination, centralization and virtualization in mobile networks, as summarized in Figure 4. This supports resource pooling (more cost-efficient processor sharing), scalability (more flexible hardware capacity expansion), layer interworking (tighter coupling between the application layer and the RAN) and spectral efficiency.

A Cloud RAN should support the following:

> separation of control and user plane to support flexible scaling of capacity for different functions of the RAN
> a variety of deployment options for anticipated network scenarios, including a wide range of transport network solutions, base station configurations and user applications
> alignment with legacy deployments, which reduces the overall network complexity thanks to a unified network architecture.
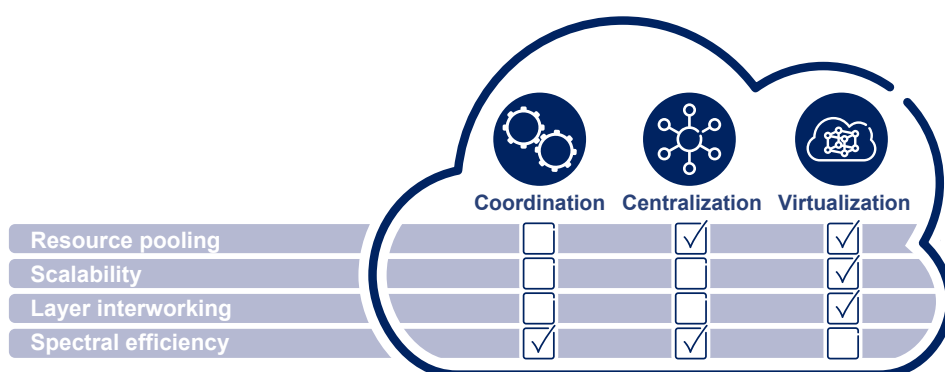
	Coordination	Centralization	Virtualization
Resource pooling		✓	✓
Scalability			✓
Layer interworking			✓
Spectral efficiency	✓	✓	

Figure 4: Some operator needs or opportunities mapped onto Cloud RAN components.

SELECTIVE CENTRALIZATION IN CLOUD RAN

The Cloud RAN will support different network architecture functional splits, as outlined in Figure 5, including different levels of functionality implemented as NFVs. By utilizing Cloud RAN, operators can 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, or further distribute the physical layer closer to the antenna (PHY split in Figure 5) to enable massive beamforming.

Cloud core and NFV frameworks also bring applications closer to the RAN, and this proximity
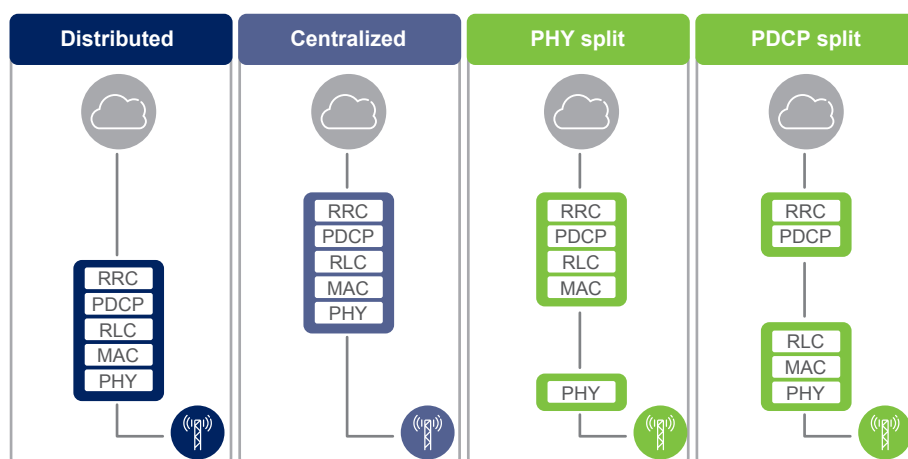
Figure 5: Examples of functional splits of the radio access protocol layer in a Cloud RAN.

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 functions 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.
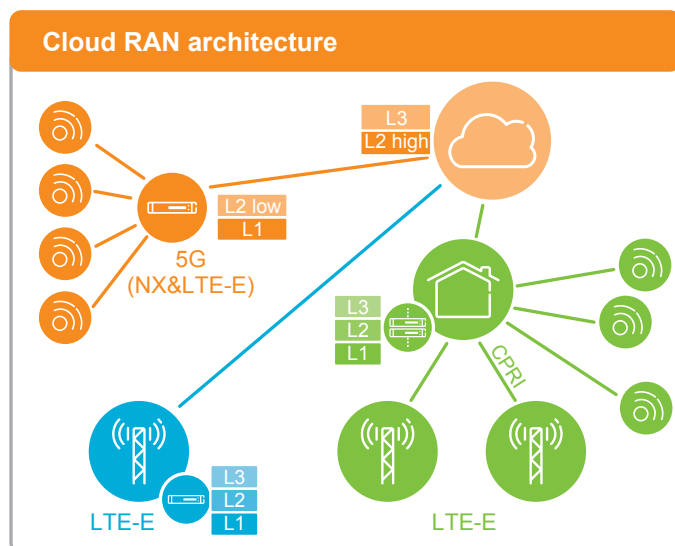

Figure 6: An example of a Cloud RAN realization showing various levels of centralization as well as a new flexible functional allocation of the different protocol layers together with virtualization in the cloud.

From a user plane perspective, Cloud 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 packet data convergence protocol (PDCP) layer 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.

IMPLEMENTATION ASPECTS

VIRTUALIZATION IN CLOUD RAN

The terms virtualization and cloud are often used interchangeably. They do work well together in many cases, including in a RAN context. However, each concept brings different things to the table.

In general, virtualization is a technique that can mean different things in different scenarios, and it is unlikely to mean the same thing in a RAN context as in, for example, a data server context. The reason for this is the substantial difference in real-time requirements imposed by the radio access protocol. Many of the synchronization requirements that ensure the performance of the radio access protocol are on the microsecond level and, in some cases, the nanosecond level. Thus, RAN functionality is not easily hosted by the so-called virtualized platform as a service (PaaS) model, as is possible with straightforward applications and server-type functions.

On the other hand, there is no need to virtualize all RAN functionality to provide the benefits of Cloud RAN. Virtualization as an execution environment technique can be used to provide isolation, scalability and elasticity, among other things, for the Radio Resource Control (RRC) protocol layer. When applied in this manner, virtualization can be used to simplify the management and deployment of the RAN nodes, for example, by allowing the definition of arbitrarily-sized base stations (in terms of the number of cells) and for more flexible scaling of higher layer functionality separate from the scaling of other layers.

Virtualization can also be used to leverage a common execution environment for RAN, core and application functionality, providing the ultimate in execution proximity and ensuring maximum responsiveness of, for example, a certain service, or, as it is sometimes called, a certain type of network slice. The possibility to virtualize network functions in this way makes it feasible to place the functionality on a more generic and generally available execution platform together with cloud core applications and other latency-critical services, sometimes even in a PaaS environment.

CENTRALIZATION IN CLOUD RAN

Centralizing base station processing with Cloud RAN simplifies network management and enables resource pooling and coordination of radio resources.

Pooling, or statistical multiplexing, allows an execution platform to perform the same tasks with less hardware or capacity. This is of greatest interest for tasks that require a large number of computational resources. It also means that the most desirable pooling configuration is a fully centralized baseband approach with a star connection long-haul 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. As mentioned earlier, however, there are not many cost-efficient solutions for long-haul CPRI.

By using separate (data center) processing capacity for higher layers, new features can be introduced without affecting the performance and capacity of distributed baseband units. The introduction of massive MIMO configurations – which will be of increasing interest with the move into ever-higher frequency bands – will also further highlight the need for optimized transport and baseband processing for centralized baseband configurations. This requirement is mitigated if centralization is isolated to the higher layers of the protocol stack.

There are some other potential centralization advantages, including:

> fewer X2 instances: centralizing the X2 control plane leads to fewer X2 instances and shorter X2 distances, i.e. latencies
> fewer handover failures: centralizing mobility and traffic management decisions leads to fewer handover failures and less network control signaling in complicated (heterogeneous) radio network environments
> memory trunking gains: pooling of the user equipment contexts from users served by more than one cell.

COORDINATION IN CLOUD RAN

Radio coordination between cells and bands is used to maximize spectrum efficiency and user experience. It is anticipated that there will be a need for different bandwidths of cells in different bands, as well as increasingly heterogeneous network deployments, and that this will call for increasingly tight interworking between technologies and cell layers to ensure a seamless user experience.

Centralized coordination functionality is advantageous in order to simplify and maximize performance of the network, including handovers, carrier aggregation and interference management.

Again, the centralized baseband configuration provides great potential to leverage interference management, joint processing and reception-combining techniques. But it is also still possible to have many levels of coordination in cases where centralization is limited to other higher layers of the radio protocol stack.

A CLOUD RAN TIMELINE

When can we expect to see the introduction of Cloud RAN on a large scale? In order to answer this question, it is necessary to consider once again the drivers behind Cloud RAN and the building blocks of the concept, such as centralization, coordination and virtualization. There are three primary developments that will shape the Cloud RAN environment going forward:

> deployments of heterogeneous networks with a mix of macro and small cells, and new bands with substantially different coverage, will benefit significantly from radio resource coordination, which becomes more effective with the centralization of the execution of the resources seen in Cloud RAN.

> the imminent adoption of NFV into mobile core networks. With the knowledge gained from this, focus will naturally shift towards the introduction of Cloud RAN.

> separation of the different protocol parts, driven by a need for decoupled and independent scalability. Cloud RAN could be used to simplify the choice of execution platforms as well as to provide a way of optimally placing functionality at the right places in the transport network aggregation hierarchy. These needs for scalability and optimal placement both stem from the capacity and data rate asymmetries anticipated when future very high frequency bands with massive bandwidth but limited coverage are anchored in existing low frequency bands with good coverage but limited bandwidth.

CONCLUSION

The Cloud RAN concept provides opportunities that can lower TCO and improve performance in the network. The concept supports excellent performance in low-latency network scenarios while still being able to function robustly in high-latency deployments. In comparison, traditional baseband centralization or base station hoteling approaches place more rigid requirements on the transport network in particular. Cloud RAN can therefore provide benefits of centralization, including tight layer interworking, without the need for a total rebuild of the transport network.

Cloud RAN can provide further benefits in terms of more separated and decoupled scalability of the different parts of the RAN functionality (enabling use of more COTS and general purpose processors, which can reduce the time to market for new services) and options to mitigate functional and capacity asymmetries introduced by future much more capable RATs.

Ultimately, Cloud RAN gives operators the chance to co-execute RAN functionalities with other network elements and functions on a generic execution platform or in a data center environment, which will enable them to offer new and exciting services more rapidly over the future mobile broadband networks that will form the backbone of the Networked Society.

Figure 7: Cloud RAN tools and benefits.

GLOSSARY

CoMP	coordinated multi-point
COTS	commercial off-the-shelf
CPRI	Common Public Radio Interface
FD-MIMO	full dimension multiple-input, multiple-output
L1, L2, L3	Layer 1-3 from the OSI protocol model
MIMO	multiple-input, multiple-output
MTC	machine-type communication
NFV	Network Functions Virtualization
PaaS	platform as a service
PDCP	packet data convergence protocol
PHY	physical layer corresponding to L1
RAN	radio access network
RAT	radio access technology
RRC	Radio Resource Control
TCO	total cost of ownership
X2 interface	base station to base station communication interface