



Tech Unveiled

Ericsson Cloud RAN

Content

01 Introduction

02 The four components of
Cloud RAN

03 The case for integrated high-
performance RAN processing

04 Evaluating processing
options in 5G RAN

05 How to get the most out of 5G
mid-band in Cloud RAN

06 The innovation potential of
Non Real-Time RAN Intelligent
Controller

07 Next generation Cloud RAN
management and orchestration

08 Leverage Ericsson Spectrum
Sharing benefits in Cloud RAN



01

Introduction

Introduction

Not long ago, it would have been unlikely to think about service providers bringing virtualization to their network deployments. But a lot has changed since then – 5G came with force and with it came the demand for new levels of flexibility in architecting, scaling and deploying telecom networks. Cloud technology offers new innovative alternatives for such Radio Access Network (RAN) deployments complementing existing proven purpose-built solutions. Cloud RAN refers to realizing RAN functions over a generic compute platform instead of a purpose-built hardware platform and managing the RAN application virtualization using cloud-native principles.

Cloud RAN entails the migration from custom-built network nodes to network functionality implemented in software running on a generic hardware compute platform. Vendor-agnostic hardware facilitates innovation across a wide range of software ecosystems, and Cloud RAN also creates the potential to generate additional network architecture flexibility, bringing platform harmonization and simplification.



System verified

Ericsson can offer a full stack Cloud RAN solution



Cloud agnostic

Our Cloud RAN software is designed to be cloud agnostic



Leading radios

We offer seamless utilization of our radio portfolio



Cloud native

Solution to maximize efficiency and TCO gains



Seamless interworking

With Ericsson Radio System and Ericsson Spectrum Sharing



know more about Ericsson Cloud RAN

02

The four components of Cloud RAN

The four components of Cloud RAN

5G calls for new levels of flexibility in architecting, scaling and deploying telecom networks. Cloud technology offers new innovative alternatives for such Radio Access Network (RAN) deployments complementing existing proven purpose-built solutions. Open RAN is an industry term for open radio access network architecture with open interoperable interfaces and hardware-software disaggregation, which enable big data and AI-driven innovations in the RAN domain. To realize this in scale requires leveraging cloud technologies and web-scale paradigms which leads us to Cloud RAN.

Cloud RAN refers to realizing RAN functions over a generic compute platform instead of a purpose-built hardware platform and managing the RAN application virtualization using cloud-native principles. Cloudification of RAN begins with running selected 5G RAN network functions in containers through Commercial off-the-shelf (COTS) hardware platforms. It starts with control plane and user plane in the Centralized Unit and continues with latency sensitive radio processing functions in the Distributed Unit. In this journey, Cloud RAN will adopt leading practices and become a foundation for openness and enable innovation in 5G.

In this chapter we will explore:

- The four components of Cloud RAN
- Key considerations in Cloud RAN
- The value Ericsson brings into Cloud RAN

The four components of Cloud RAN

Ericsson has broken down Cloud RAN into four components that create the foundation for a successful Cloud RAN implementation.

COTS hardware including accelerators

Selecting the right hardware platform and environment.

Cloud native architecture

Realizing RAN functions as microservices in containers over bare metal servers using cloud native technologies such as Kubernetes and applying DevOps principles.

Management, orchestration and automation

Bringing end-to-end life cycle management of services across Cloud RAN, transport, 5G core and underlying cloud infrastructure.

RAN programmability

Deploying non-RAN functions in the virtualized system to add new functionality and additional value.

COTS hardware including accelerators

Selecting the right hardware platform and the virtualization environment is an important first step. Commercial-off-the-shelf server hardware with processors, Network Interface Cards (NICs) and hardware accelerators is the foundation. Specific server configurations are based on workload demands, especially for Distributed Units (DU) with its higher processing demands and stricter requirements on latency. This requires the use of accelerators to manage Layer 1 pipeline functions and is especially critical to meet the demands of Massive MIMO radios.

In coming blogposts, we will discuss in more detail different types of processing architectures including how to best realize an efficient hardware architecture for Cloud RAN.

Cloud native architecture

Disaggregating RAN software and hardware in Cloud RAN has the potential of independent innovation on software and hardware. However, simply forklifting existing 5G RAN software to a COTS platform is not enough. To really realize the value of Cloud RAN one needs to embrace cloud native architecture.

Cloud native architecture facilitates RAN functions to be realized as microservices in containers over bare metal servers, supported by technologies such as Kubernetes. Such software modularity and decomposition allow for:

- Independent life cycle management following DevOps principles and Continuous Integration and Continuous Delivery (CI/CD)
- Scaling of component RAN microservices
- Application-level reliability with platform abstraction
- Simplified operations and maintenance with network automation

Management, orchestration and automation

As the industry is transforming and introducing 5G networks, the management, orchestration and automation has become a cornerstone of 5G networks. With the introduction of Cloud RAN, edge compute, container-based architectures, and network slicing, the management and orchestration systems have become critical for success. These systems will now have to manage physical network functions (PNF), virtual network functions (VNF), cloud native network functions (CNF) and end-to-end life cycle management of services across Cloud RAN, transport, 5G Core and underlying cloud infrastructure.

Under this context, Cloud RAN management is not just about managing hybrid network functions in RAN, but also about bringing cloud native principles, programmable networks and model driven management and orchestration systems. This brings key values needed in 5G networks such as automation, flexibility, faster time to market, managing complex networks, and better customer experience.

In an upcoming blogpost, we will elaborate more about how we enable automation, management and orchestration to deliver 5G services across all industries.

RAN programmability

RAN programmability is essentially the capacity to tailor the behavior and performance of the RAN through applications with access to multi-domain information. This will also open up the RAN environment to a broader ecosystem and allows developers to build applications completely independently from innovations in the RAN domain. This unlocks new innovation value, enhancing the CSPs role and creates opportunities for monetization.

Programmability of RAN can be done both within the framework of 3GPP and within the complementary specifications of O-RAN. The O-RAN architecture includes two flavors of RAN Intelligent Controller to aid programmability of RAN:

- **Non-Real Time RAN Intelligent Controller (Non-RT RIC)** – operating in the control loop of 1 second and above.
- **Near Real Time RAN Intelligent Controller (Near-RT RIC)** – operating in the control loop of 100s of milliseconds

Ericsson sees that the Non-RT RIC is offering the greatest benefit. With its placement higher up in the architecture, it can access a larger pool of RAN data as well as external data sources such as weather, traffic, first responder events etc. to provide policy guidance to RAN on radio resource allocation and the 1 second control loop is short enough to capture almost all possible use-cases.

Given the benefits of being able to bring multidomain knowledge at higher layers, Ericsson is currently focusing its R&D efforts on the Non-RT RIC, as opposed to the Near-RT RIC which overlaps considerably with RAN functionality already provided in feature rich RAN offerings as offered by Ericsson.



Key considerations in Cloud RAN

Cloud RAN journey

Evolution to Cloud RAN is a multi-year journey. A key learning from the virtualization journey of the core is that pre-verified solutions work best and are the most efficient, and that customer unique cloud environments come with additional work, cost and TTM overloads. The Ericsson approach to Cloud RAN is built to be cloud agnostic but to extract maximum value, CSPs should consider a balanced approach to disaggregation. From a technology maturity perspective, Cloud RAN will benefit greatly from clear industry practices and O-RAN innovation with a high degree of alignment towards 3GPP architecture. Ericsson actively contributes to several of these alliances in order to enable innovation at global scale for the benefit of the whole industry. From our experience and domain expertise, we see a five-step journey towards Cloud RAN.

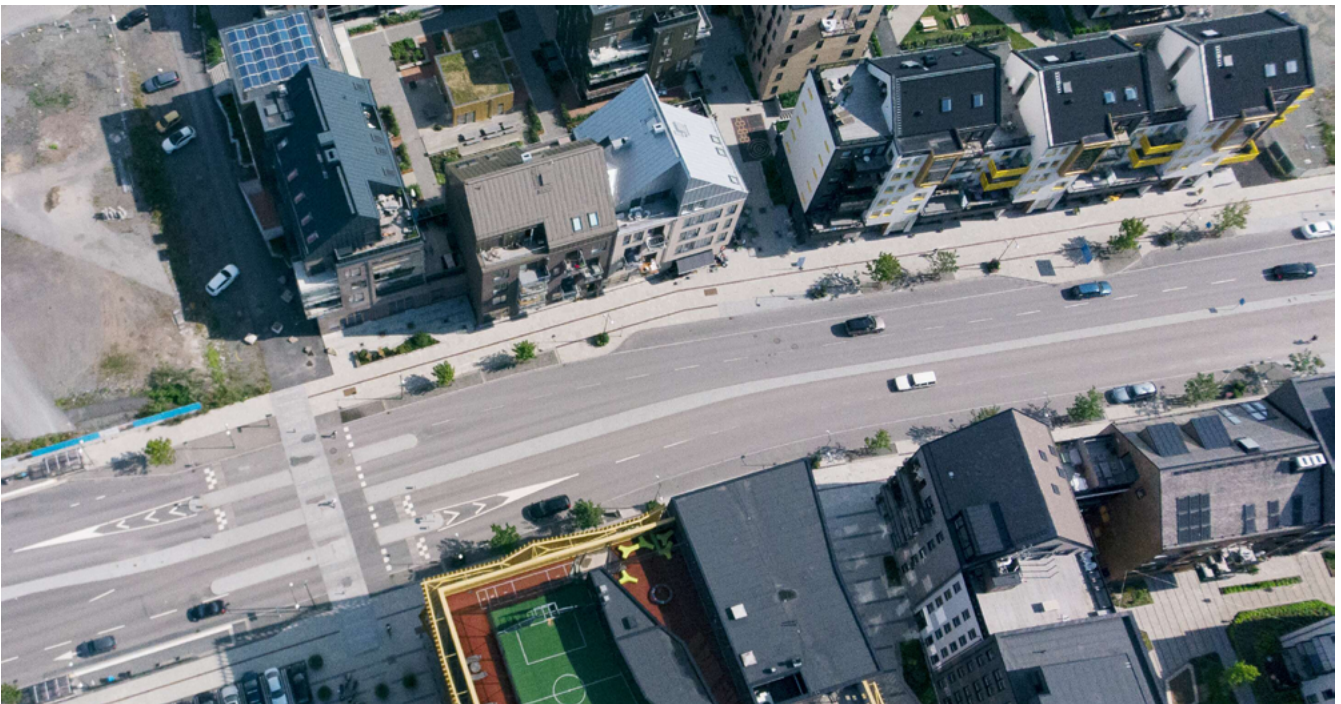
1. **Evaluate the key use cases and deployment scenarios to decide the right cloud infrastructure.** This typically includes the infrastructure hardware, cloud platform and RAN applications that will be hosted on it and the interworking and compatibility with current installed base.
2. **Virtualization of Central Unit (CU) and especially CU-UP.** This is the lower risk starting point and opens up for flexible distributed edge placement of vCU-UP and virtual User Plane Function (vUPF).
3. **Fully cloud-native realization of both CU-CP and CU-UP.** This provides independent scaling for centralized control and user planes and creates locational flexibility in deployments
4. **Cloud-native realization of the DU function.** This brings challenges around server and accelerator selection, cloud infrastructure and OAM, capacity dimensioning, power efficiency planning and security planning.
5. **Introduce Cloud Native SMO (Service Management and Orchestration) in parallel –** Bring RAN programmability to the network and start to capitalize on this foundation for innovation.

Building a disaggregated system

Building a disaggregated system, that actually works as a cohesive entity is complex. The disaggregation of hardware and software means that the RAN solution no longer has a unified node concept, instead consisting of software functions, NFVI and hardware versions, all with their own life cycles and release cadence. Additionally, Radio Access Networks have rather specific processing needs and requirements compared to typical IT applications or even core network workloads.

To build and integrate such a system requires telecom expertise, the ability and experience to deliver pre-verified solutions and solid network evolution capabilities and skills. Cloud RAN will be introduced into live networks with multiple radio access technologies, running predominantly as physical functions, with no impact on the quality of experience for its users. This demands strong knowledge not only of Cloud RAN but of the entire existing telecom stack.

To do this with an IT-focused approach at scale is unproven and current business models do not cater for back-to-back support agreement with RAN component suppliers, indicating an area where Ericsson can contribute to the maturity of the industry over time.



Why partner with Ericsson in your Cloud RAN journey?

Ericsson is a partner for your entire Cloud RAN journey, from design, solution creation, implementation and operations.

Global RAN expertise: Ericsson brings significant RAN expertise and experience from delivering leading and robust networks throughout the globe.

Cloud RAN industry practices: Ericsson is a leading contributor to the most important open forums, actively supporting efforts to establish industry practices in Cloud RAN.

Cloud RAN processing: We bring deep understanding of the demanding processing requirements in 5G RAN and active adoption of cloud technologies for application in the RAN domain.

Cloud RAN compatibility: Ericsson brings a leading radio portfolio meeting the unique needs of widely different networks and bring features that ensure Cloud RAN interworks seamlessly with installed base. One great example is the unique Ericsson Spectrum Sharing solution.

Ericsson pre-verified solutions: Ericsson can deliver a pre-verified Cloud RAN solution ranging from cloud infrastructure, RAN software, COTS hardware and accelerators and Radio gateways for optimized transport performance.

Cloud-native transformation experience: We have successfully helped our customers through a similar multiyear cloud-native core transformation journey.

We have a leading 5G position and rich knowledge of building and running several generations of robust and reliable radio access networks globally. We are delighted to share our expertise and help our customers unveil a whole new world of possibilities with Cloud RAN.

Authors:



Eric Parsons
Head of Product Development Unit, Cloud RAN



Gabriel Foglander
Strategic Product Manager, Cloud RAN

03

The case for
integrated high-
performance RAN
processing

The case for integrated high-performance RAN processing

The pandemic has shown us how critical role connectivity plays in our lives. In a short time, video communication and network quality has become a key part of our daily lives. At Ericsson we innovate to help people stay connected. At the heart is the unique Ericsson Many-Core Architecture which powers our networks, enabling the industry's most scalable and energy efficient RAN Compute portfolio, typically 2x more power efficient versus competition. Let's explore the Ericsson Many-Core Architecture solution!

Great network performance is often taken for granted, but the on-going pandemic has shown how important it really is for our daily lives. In the June 2020 edition of Ericsson Mobility Report, 83 percent of surveyed network users admitted that ICT has helped them to cope with the lockdown. For many, the saving grace has been video for work conferencing, socializing and entertainment. Even before the current crisis, Ericsson estimated that over 60 percent of all network traffic was video-related. Hopefully, we will see the end of the pandemic soon, but it won't be the end of growing demand for video.

We expect video content to make up over three-quarters of all network traffic by the end of 2025. Our research shows that, as video-centric apps become part of our everyday lives, people have come to expect top performance and, importantly, that they are willing to pay more for an optimal user experience. Here's a look at how the innovative Ericsson Many-Core Architecture (EMCA) can help CSPs deliver the kind of network performance users expect, both today and tomorrow.

A growing need for RAN processing power

RAN processing power will become a critical resource as we move further into 5G. There are three key drivers that will place the greatest processing demands on our RAN:

Radios have become more advanced to cater to the increase in speeds and overall required network performance, progressing from two to four to eight transceivers including 5G Massive MIMO (64 transceivers and massive beamforming)

There has been a significant increase in bandwidth to address the capacity required for new use cases – from 20MHz on the low band for 4G, to 100MHz on the mid-band for 5G and 800MHz on the high band for 5G

The time transmission (TTI) requirement (that is, the critical loop in which the scheduler works and processing has to be carried out) has decreased from 1ms on the low band for 4G, to 0.5ms on the mid-band for 5G and 0.125ms on the high band for 5G. This is an enabler of several use cases in 5G. For instance, ultra-reliable low latency use cases for mission-critical applications and enhanced mobile broadband use cases, such as gaming.

Quality is a processing challenge

Delivering the top network performance of the future is a processing challenge. In order to ensure that everyone gets a great user experience, the RAN Compute hardware will need to be able to execute advanced algorithms for:

many users and devices with...

many quality of service levels in...

many network slices in...

many cells with...

many resource blocks with...

many MIMO layers and...

many antenna branches

And all this processing needs to be completed without fail within down to 0.0001 seconds!

The EMCA engine

This is a massive processing challenge, requiring top computing performance, but also without sacrificing power efficiency. Ericsson has met this challenge head on by developing EMCA – a unique processing architecture that allows us to design solutions both for present and future computing needs.

Ericsson uses tailor-made Application Specific Integrated Circuits (ASIC), containing hundreds of specialized signal processors (DSPs) and accelerators, to implement the EMCA System on a Chip (SoC). SoC is an advanced ASIC with digital logic, mixed signal content, and processors.

Our software's architecture is designed to fully leverage the capabilities our silicon provides, making EMCA a computing solution, rather than just a complex integrated circuit.

EMCA is the essential parallel processing power behind all of our Ericsson Radio System products and solutions. As a unified architecture, it is used to implement the same hardware and software in all of our networks (regardless of their capacity, size and form factor), ensuring optimal network performance and efficiency.

Ericsson Uplink Booster: the unsung hero of 5G

The sharp rise in video conferencing has caused a 40 percent increase in uplink traffic. That's why we have developed a 5G mid-band (sub 6GHz TDD) base station solution that boosts the uplink behind the scenes, becoming the unsung hero of 5G to cater for the needed capacity.

Known as the 5G Uplink Booster, it is implemented in the radio – on top of EMCA – to enable extensive coverage by securing a strong radio signal at the cell edge. It offers extreme receiver resolution and full interference rejection.

This ensures 10dB coverage gain in the mid-band, maximizing uplink MIMO performance through our zero-compromise architecture.

The improved uplink performance can boost user app coverage by an impressive 90 percent, and cell-edge uplink speeds can be increased by a factor of 10.

RAN Compute with EMCA

EMCA gives our products unparalleled performance, and allows for scalability across multiple generations of hardware and technologies. Ericsson's RAN Compute portfolio runs on EMCA, which enables our products unparalleled energy efficiency, typically 2x more power efficient than competition. Furthermore, its building practice allows for linear scaling of both user- and control-plane capacity with fewer boards on site and as a result, fewer site visits. This could potentially result in 50% less site visits corresponding to costs savings of millions of dollars over the years.

Thanks to the EMCA architecture a flexible usage is enabled which allows the RAN Compute products to run a multitude of configurations across 2G/3G/4G/5G/IoT on a single board. And by the way, they are designed for new technologies as in the case of being 5G-ready.

EMCA also facilitates continuous innovation in RAN. EMCA can impact the transport network, for example, by making it more flexible and efficient. It can be applied anywhere within the RAN in the radio or baseband, minimizing transport costs as a result. Without this highly effective silicon processing hardware, none of our innovative software features could be made available – the invaluable solutions that serve to boost network performance.

To realize the full potential, it also houses extremely sophisticated software with several million lines of code – nearly four times as much as Mars Curiosity rover – to provide high performance wireless communication that constantly pushes the boundaries of efficient and secure use of spectrum.

Here are two such key innovative software features that EMCA enables:

Ericsson Spectrum Sharing: the fastest way to 5G

Ericsson Spectrum Sharing (ESS) is a software that enables our partners to scale up 5G fast by introducing it throughout their LTE networks. What's more, service providers can use the same radio, CPRI link and RAN Compute baseband to activate spectrum sharing, making it a highly cost-effective solution.

Reusing the same hardware and using low-band spectrum for 5G means networks can be built with fewer sites, which is positive from a sustainability perspective.

It also transforms the way 5G is introduced nationwide. And it can deliver cutting-edge connectivity to everyone, everywhere in the fastest possible way.



Staying ahead of the game

Innovation is in our DNA. At Ericsson, we believe in creating future-proof network technologies that meet not only today's challenges but are also ready to take on tomorrows'. That's why our networks have been ready for 5G since 2015 and is also the reason behind EMCA, the unified architecture used across the Ericsson Radio System portfolio.

In the world of connectivity, technology evolution happens fast. Our dedication to future-proof technologies ensures that radio products will continue to perform effectively for many years to come – extending the life cycle of network equipment of service providers and driving down their total cost of ownerships (TCO).

The long-term network quality we enable also makes sense from a business perspective. Market share rankings clearly indicate that network performance has a major impact on service providers' market performance, and that network quality is key to lowering churn and increasing average revenue per user (ARPU). By making sure we stay a step ahead in our innovation, we also enable our customers to stay ahead of the market demands and differentiate against the competition.

Authors:



Anders Söderlund
Network Evolution Expert



Fadi Hannah
Product Marketing Manager

04

Evaluating processing options in 5G RAN

Evaluating processing options in RAN: Purpose built and Cloud RAN

With the advent of technologies such as Cloud RAN and the specifications developed in the O-RAN alliance, service providers are faced with questions relating to new and different options to process RAN functions:

- Is it best to execute in a de-centralized way, close to the antenna, or in a more centralized location?
- How is purpose-built hardware and cloud infrastructure best combined for RAN processing?
- How can total cost of ownership (TCO) be optimized across the whole system?

A new processing baseline for 5G

Before addressing the above questions, let's look at service providers two key expectations on 5G:

1. Manage the anticipated traffic growth of 4X from now until 2025 as predicted in Ericsson Mobility Report
2. Leverage mobile networks to support new use-cases such as critical IoT, fixed wireless access, Ultra reliable low latency connectivity etc.

To meet these expectations service providers are building 5G networks by:

- Adding new and wide frequency bands, i.e. mid band and high band.
- Deploying advanced antenna systems with massive MIMO for additional coverage and capacity.
- Using shorter time transmission interval (TTI) to lower latency.

These expectations and actions impact and increase the requirements on the way service providers design their 5G networks and decide on the best RAN processing option.

The 5G RAN protocol stack

Now let's turn our focus to the 5G RAN protocol stack to outline the different parts where RAN processing is important. 3GPP has defined a Higher Layer Split interface (F1), separating the Central Unit (CU) from the Distributed Unit (DU). In summary, the 5G protocols contain:

Function	Subfunction	Main processing function	Scales with:
CU	Control plane (CU-CP)	Radio Resource Control	Subscribers/connections
	User plane (CU-UP)	Packet Data Convergence Protocol	Throughput
Higher Layer Split (HLS)			
DU	Digital Unit	Radio Link Control	TTI, bandwidth & branches
		Medium Access Protocol	TTI, bandwidth & branches
		Physical Layers	Branches & bandwidth
	Lower Layer Split (LLS)		
	Radio Unit	Beamforming (if applicable)	Branches & bandwidth

Figure 1. 5G RAN protocol stack

The lower you go in the protocol stack, the higher the demand on processing – Layer 1 & 2 combined comprise 90% of processing demands. The wide 5G mid and high bands together with massive MIMO technology exponentially increase the processing

demand on L1 and beamforming (Figure 2). In addition, the transport overhead increases the lower you go in the protocol stack and for this reason the de-centralized processing has the advantage of lowering the requirements on RAN transport.

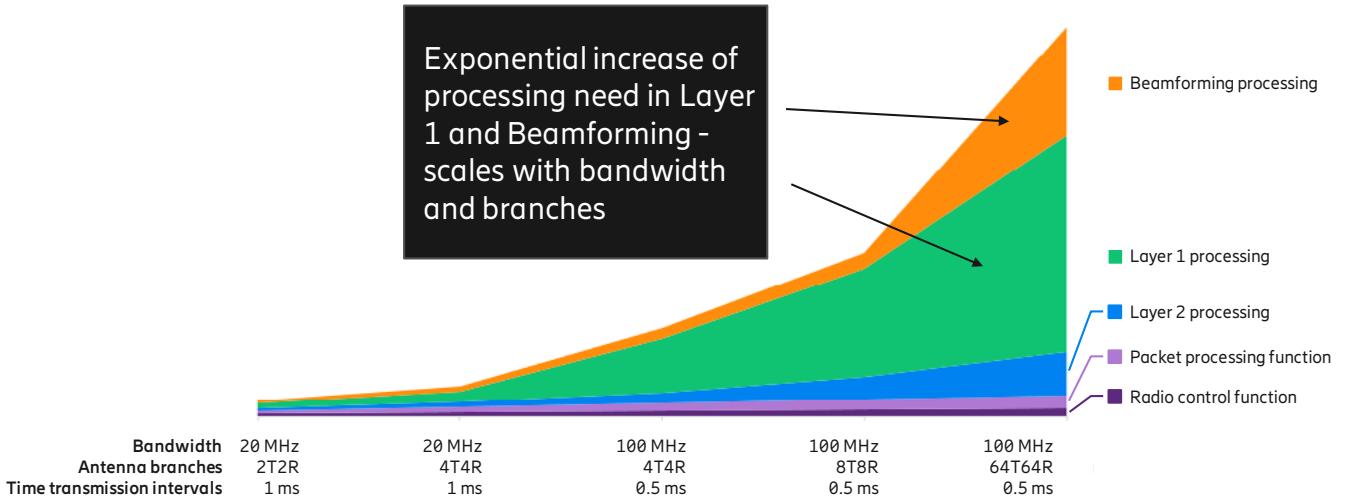


Figure 2. Processing demand per protocol layer

Processing in RAN is diversifying

When 4G was built networks were homogeneous with a de-centralized baseband on every antenna site. Since 5G RAN is more heterogeneous with a combination of low, mid and high bands, the processing is better suited to be diversified to meet different needs. Below is a list of possible set ups (non-exhaustive) :

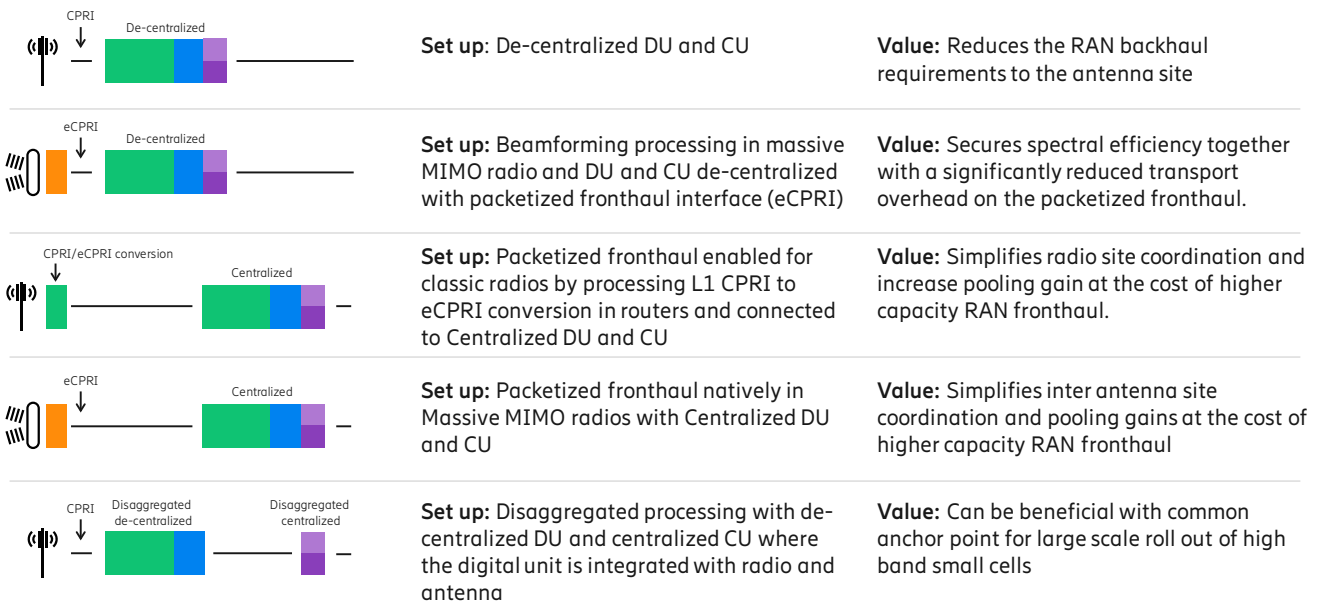


Figure 3. Illustrative deployment set ups and service provider value. (legend as in Figure 2)

Processor options in RAN

Purpose built hardware, such as Ericsson RAN Compute Baseband, uses system on a chip (SoC) which is a tailor-made chip based on Ericsson Many-Core Architecture (EMCA) for lower protocol layers and uses X86 processor for the higher layers.

X86 processors are used in Cloud infrastructure today. Ericsson has shown in a live demo that the complete stack for low band 5G is feasible to process on such infrastructure. For mid band and massive MIMO, hardware acceleration will be needed as a complement for the demanding L1 processing.

Now let's look at the merits of de-centralized and centralized processing for both purpose-built and cloud infrastructure.

De-centralized processing

- **Purpose-built hardware** - For de-centralized processing on the antenna site the purpose-built hardware can integrate into massive MIMO radio for beamforming and full Interference Rejection Cancellation (IRC)-receiver processing, into routers and gateways for CPRI to eCPRI conversion and into basebands and outdoor radio processors.

- **Cloud infrastructure** - If rack space is available on the antenna site both purpose-built and cloud infrastructure processing is feasible.

Centralized processing

- **Purpose-built hardware** – Purpose-built hardware has the advantage of maximizing the compute power in a sub-rack resulting in a more compact installation.
- **Cloud infrastructure** - For centralized processing on cloud infrastructure the advantage is that synergies can be found when multiple applications such as RAN, core and content/compute can share infrastructure resources in the cloud.

Ericsson Radio System and 5G cloud core is the foundation

Ericsson has more than 100 5G commercial agreements, all built on a foundation of Ericsson Radio System (ERS). It is powered by a tailor-made system on a chip (SoC) with special purpose processors, based on Ericsson Many-Core Architecture (EMCA). EMCA is future proof and optimized for maximum performance with a small form factor and low power consumption. It enables integration into many Ericsson Radio System products including massive MIMO radios, routers and RAN Compute allowing for a coherent and integrated system performance proven in the field.

It is less attractive for service providers to invest in cloud infrastructure for 2G/3G/4G since this is already deployed and a mature technology handled by multi standard, purpose-built hardware. In addition, Ericsson Radio System radios and RAN

compute portfolio supports 5G with a remote software installation.

Service providers now deploy 5G core based on cloud infrastructure, capitalizing on cloud native technologies for a more efficient and TCO optimized deployment. As service providers evolve this cloud infrastructure it is important to consider the specific demands of 5G RAN use cases so that the infrastructure can be shared across domains (e.g. running both 5G Core and RAN).

Ericsson provides the possibility to complement purpose-built with cloud infrastructure

Ericsson Radio System interfaces evolve to also enable RAN processing on cloud infrastructure:

1. Massive MIMO radios powered by EMCA SoC for the heavy processing in L1, maximizing spectral efficiency and providing a packetized fronthaul interface (eCPRI). The packetized fronthaul enable processing also on cloud infrastructure.
2. EMCA SoC is used in routers to implement L1 CPRI to eCPRI conversion for remote radio units. The packetized fronthaul enable processing also on cloud infrastructure.
3. High band street macro with EMCA SoC for the lower layers can benefit from a common anchor point with interconnection to the core. This anchor point can be realized as a central unit (CU) on cloud infrastructure, with potential to manage load balancing more holistically with new scaling capabilities.

Processing options in 5G core networks are also diversifying. Core network processing has traditionally been done on few centralized sites. But there are benefits with more de-centralized core processing such as low E2E application latency and Security. Such de-centralization of 5G core brings cloud infrastructure into edge or far-edge data centers creating processing synergies with the RAN since these new cloud sites can also process RAN functions.

Ericsson foresees a stepwise introduction of cloud infrastructure for RAN processing to complement Ericsson Radio System. Here are three examples:

1. Introducing unique interworking features such as Ericsson Spectrum Sharing and Carrier Aggregation across purpose-built RAN and cloud infrastructure.
2. Centralized RAN deployments in strategic locations can serve many cells provided that sufficient transport capability is in place. This type of deployment enables better resource pooling, efficient scaling capabilities on cloud infrastructure and potentially also removing the need for hardened server deployments on de-centralized sites.
3. Enterprise and industry edge are use cases that could benefit from RAN deployed on cloud infrastructure by simplifying operations & maintenance, enabling local data sovereignty and use of Machine Learning/Artificial Intelligence applications.



Check points for evaluating processing options in RAN

As a conclusion there are some technology agnostic check points that are generally applicable for service providers when evaluating processing options in RAN.

Evaluate 5G RAN processing:

to cater for high load to support the expected traffic growth

to ensure support for future use-cases

in combination with transport

considering pooling gains of centralized vs. de-centralized

to optimize full RAN and transport TCO

quantifying the value in operations and maintenance by harmonizing onto uniform COTS hardware

taking into consideration the potential co-location synergies of core break-outs, far edge- and centralized RAN-deployments

We are in the beginning of an exciting journey with a multitude of RAN processing options that will meet the expectations of managing the anticipated traffic growth and support new use cases. Ericsson is convinced that a complementary approach of purpose-built and Cloud RAN processing is the most ROI friendly approach for service providers. This as it gives a broad toolbox to realize benefits in a variety of deployment scenarios.

Authors:



Anders Söderlund
Network Evolution Expert



Gabriel Foglander
Strategic Product Manager,
Cloud RAN

05

How to get the most
out of 5G mid-band
in Cloud RAN

How to get the most out of 5G mid-band in Cloud RAN

The new shining spectrum star - mid-band - brings the full potential of high capacity to explore new use cases with high bandwidth demand. Openness and Cloud RAN solutions are in focus, leading to thoughts on how mid-band can succeed in Cloud RAN. Let's guide you through some key considerations to get the most out of 5G mid-band in Cloud RAN.

Starting a new exciting journey

To provide a great customer experience, most of today's 5G deployments rely on mid-band and it is critical for communications service providers to optimize these deployments to get maximum benefits.

In this blog post we will show how Ericsson addresses the 5G mid-band challenges leveraging cloud infrastructure, giving service providers some food for thought and discuss necessary considerations required for a successful solution:

What are the challenges on mid-band deployment and how could they be solved in a cloud infrastructure?

How can we leverage the design philosophy and in-depth experience from the Ericsson Radio System into Cloud RAN?

What are the implications of differences in processing need between low-band and mid-band?

How can we reflect cloud infrastructure aspects into total cost of ownership (TCO) evaluations?

Contemplating the questions above, let us begin the journey together: mid-band using cloud infrastructure – the beginning of a fascinating journey.

Solving the mid-band coverage challenge

5G mid-band runs on wide band Time Division Duplex (TDD) spectrum below 6GHz delivering the high capacity promised by 5G. However, due to shorter wavelengths, it provides less coverage compared to low-band. The limiting factor is the uplink and Ericsson has two main solutions to address this challenge – Uplink Booster and 5G carrier aggregation.

Uplink Booster

Uplink Booster is an innovative solution available in Ericsson Radio System today, that increases coverage on 5G mid-band by moving the uplink beamforming processing from the baseband to the massive MIMO radios, where it runs on Ericsson Many-Core Architecture (EMCA). This creates up to 10dB gain, equivalent to a 3.7x extension of coverage area, improved uplink throughput and higher spectral efficiency.

Technically, Uplink Booster is based on the Ericsson lower layer split (E-LLS) which makes it possible to split the processing of the protocol stack and offload a big part of the layer 1 processing to the radio unit, in particular the instantaneous channel estimation and the full interference rejection combining (IRC). In addition, we use the evolved common packet radio interface (eCPRI), which significantly reduces the fronthaul data requirements by 90 percent. This is an example of a key advantage in our solution, which applies for both Ericsson Radio System and Cloud RAN.

Coverage extension with 5G carrier aggregation

Carrier aggregation in 5G is another powerful solution in the Ericsson portfolio. Our resolution is to decouple the downlink from the uplink, keeping the downlink data on the mid-band while the uplink data, and the control plane, are moved to the low-band. In simple words, when combining mid-band with low-band, the uplink, which is normally weaker, will reach the base station on low-band 5G, which travels longer. And the 5G mid-band signal will travel as far as the base station allows, ensuring also an extended coverage, without any limitation from a weaker signal from the user equipment.

This has two primary benefits:

- Coverage increases with low-band's longer reach, adding 7dB gain, translating to extend the coverage area by 2.5x, compared to dual connectivity deployments.
- The overall network capacity increases about 27 percent, due to the extended mid-band coverage, so we can also offload traffic on to the mid-band.

As we look to bring 5G carrier aggregation to Cloud RAN, we build on the in-depth knowledge, experience and innovations gained in Ericsson Radio System. Always building for the future, our comprehensive portfolio of Ericsson Radio System and future Cloud RAN solutions works seamlessly together.

Another value of our portfolio is the high output power radios for mid-band, designed to cope with the challenge of the high-demanding processing and coverage in mid-band. Typically, each megahertz (MHz) of bandwidth requires 2 watts (W) in output power. With the 100MHz bandwidth in the mid-band, this translates to radios which can deliver at least 200W. With our portfolio service providers have a one stop-shop with high output power radios and high performance for both Ericsson Radio System and Cloud RAN deployments.

Processing for mid-band compared to low-band

In previous blogposts we have discussed the processing power needed for 5G mid-band compared to low-band. Below, we examine how a combination of cloud infrastructure and accelerators can tackle this processing challenge.

The processing needed for a fully loaded, low-band cell on a commercial-off-the-shelf (COTS) server, excluding processing for operating system and common functions, is roughly 1 core. Comparing to a fully loaded, mid-band cell without any accelerator this would be about 16 cores. The difference in processing volume comes from the amount of data produced and consumed in the mid-band system, differences in bandwidth, layers and traffic models. It is roughly 20 times larger in downlink compared to a low-band cell (see figure 1 for rough calculation).

	Bandwidth	Layers in DL	Layers in UL	Total DL bandwidth (Bandwidth x Layers)	Total UL bandwidth (Bandwidth x Layers)
Low-band	20MHz	4	1	80MHz	20MHz
Mid-band	100MHz	16	8	1600MHz	800MHz
				~ 20x processing need	~ 40x processing need

Figure 1. Comparing mid-band versus low-band processing needs

A pure soft x86 based implementation for mid-band is not commercially viable in a deployment scenario in line with expected traffic growth due to this massive processing need. However, offloading processing of certain layer 1 functions to more specialized hardware, so-called accelerators, will make RAN on cloud infrastructure viable also for mid-band.

It is vital to consider the complete picture when dimensioning your cloud infrastructure. Communications service providers need to consider processing requirements for radio functions but also the processing required for management, orchestration, fronthaul and backhaul termination and handling the life cycle of the components in the system.

In these kinds of disaggregated systems, there is a thin line between over and under dimensioning. A strategic choice must be made to either create processing head room for new features and users in the system or to have a tighter dimensioning of hardware which would limit physical footprint and power consumption. Either way, it is important to take careful consideration of target deployment scenarios and expected traffic growth.



Choice of accelerators

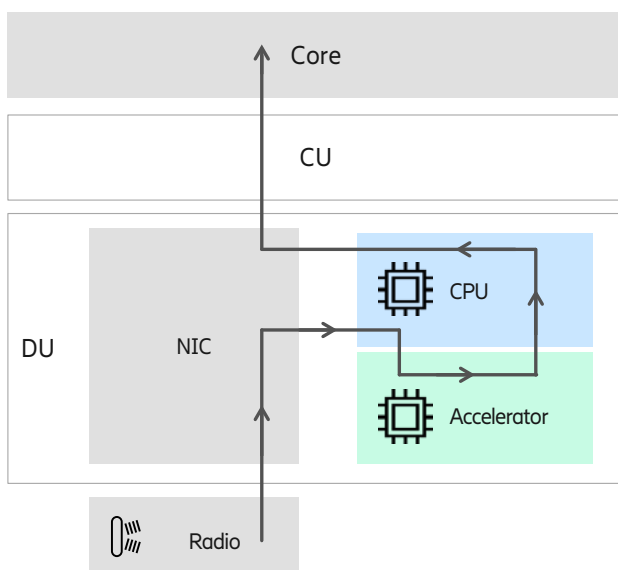
One important aspect of accelerators is programmability. In general, a fully programmable solution is the best and most versatile solution for scalability, which could be automated based on traffic load and current communications service providers' needs. That said, in the end there will be a trade-off between programmability and compute efficiency – a more hardened solution will always be

more efficient at the cost of less flexibility. There are different propositions in the market when it comes to acceleration architecture and hardware. The accelerator hardware typically used are Field Programmable Gate Arrays (FPGA), Graphics Processing Units (GPUs), or Application-Specific Integrated Circuits (ASICs).

Looking at the architecture for layer 1 acceleration, there are two fundamental approaches - look-aside acceleration and inline acceleration.

The principal difference between them is that in look-aside acceleration, only selected functions are sent to the accelerator, and then back to the CPU while in inline acceleration parts of or the whole data flow and functions are sent through the accelerator (see figure 2).

Look-aside acceleration



Inline acceleration

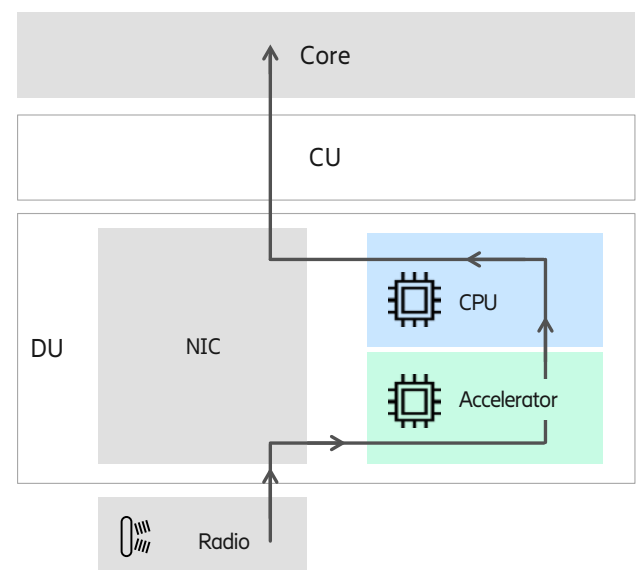


Figure 2. Illustrating look-aside versus inline acceleration

Look-aside acceleration

In the case of look-aside acceleration, the CPU is free to use its cycles to process other useful tasks, while the accelerator is working on the data to be accelerated. Once the CPU receives processed data back from the accelerator, it can switch back to the original processing context and continue the pipeline execution until the next function to be accelerated comes up.

The look-aside acceleration requires massive data transfer between CPU and accelerator and therefore a tight integration is preferred.

Inline acceleration

In the inline acceleration case, a part of or the entire layer 1 pipeline can be offloaded to the accelerator, potentially allowing for a less data heavy interface between CPU and accelerator. The acceleration solution can in this case be a mix of programmable and "hard" blocks, again there is a trade-off between flexibility and efficiency.

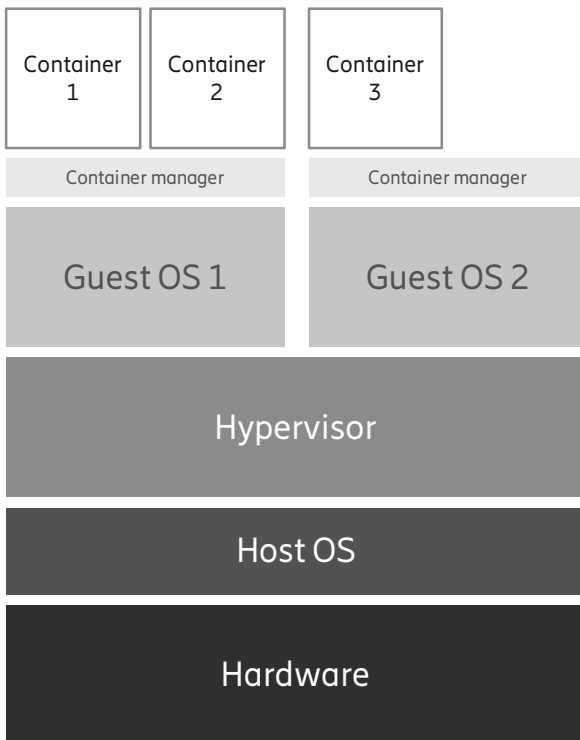
The importance of openness

The market for Cloud RAN is still in its early days and acceleration solutions will evolve over time. Ericsson believes that, it is important to ensure a range of available acceleration options in the ecosystem. The ideal state in a truly disaggregated system is complete independence of software and hardware selection. In practice, the differences in architecture between look-aside and inline acceleration require software adaptation to maximize performance. Therefore, Ericsson is actively contributing to O-RAN Alliance's working group 6 (WG6) with the ambition to allow for specification of many accelerator profiles and open capability negotiation between hardware and software.

Optimizing TCO with container deployments on bare metal

The usage of accelerators reduces significantly the number of CPU cores needed in the cloud infrastructure. To further reduce footprint, increase processing efficiency and achieve lower TCO, we believe, the current best approach is using Kubernetes containers on bare metal. Kubernetes has emerged as the platform of choice for deploying cloud-native applications. To run this on bare metal, which essentially means removing the virtualization layer (see figure 3), is a natural development of the industry.

Kubernetes on virtual machines



Kubernetes on bare metal

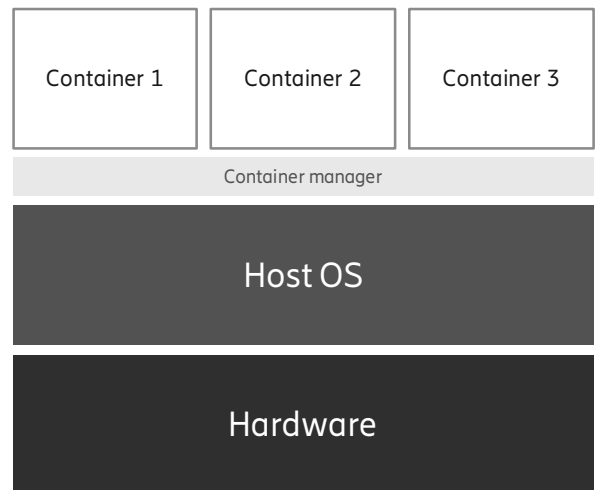


Figure 3. Comparing Kubernetes on virtual machines versus on bare metal

Ericsson has compared deploying our cloud-native applications on virtual machines or on bare metal. The analysis shows that it's possible to achieve TCO savings of at least 30% in the bare metal scenario. The savings impact both the CAPEX and OPEX side.

From a CAPEX perspective, there are three main savings. Firstly, a more efficient hardware utilization, leading to a reduced footprint. Secondly, a higher application efficiency, translating into a 10-20% performance advantage over virtual machines. Finally, the elimination of virtual machine specific fee with simplified cost structure.

From an OPEX perspective, there are also three main savings. Firstly, the software upgrades are faster and simplified, as virtual machine layer and associated upgrade complexities are eliminated. Secondly, there is a unified competence across applications, since the system is less diversified. Finally, the life-cycle management of the system is simplified, as the application onboarding is unified.

Checklist for communications service providers considering a Cloud RAN deployment in mid-band

We have now exposed the main perspectives of 5G mid-band deployments using cloud infrastructure. In summary, we see six key points, which service providers should keep in mind.

Extending coverage on 5G mid-band by leveraging the low-band is vital to meet the needs and expectations of users.

A high-performing mid-band deployment requires high output power radios with high demands on processing.

Dimensioning of the cloud infrastructure should take the whole picture into account and consider processing needs for future scenarios.

Cloud infrastructure is a feasible option for mid-band with the addition of hardware accelerators.

Choice of accelerators is a strategic decision and depends on communications service provider's deployments and traffic scenarios.

Cloud-native deployments on bare metal versus virtual machine deployments will increase efficiency and lower TCO.

Ericsson's powerful portfolio brings Cloud RAN to 5G mid-band

5G mid-band will be a critical building block for any service provider. We stand by communications service providers as a guide in their future mid-band journey towards Cloud RAN, supporting in strategic decision making and delivering a comprehensive and powerful portfolio in line with time frames required by leading customers.

In order to solve the 5G mid-band coverage challenge, we build on our in-depth radio knowledge to develop high-performing future Cloud RAN solutions, leveraging

our design philosophy and feature set for Ericsson Radio System, such as Uplink Booster and 5G carrier aggregation.

Adding an extra boost, our high output powered radios enable communications service providers to fully utilize the larger bandwidth for maximum performance. By using Kubernetes on bare metal, we can also reduce the TCO for Cloud RAN significantly.

Ericsson stands ready to guide its customers to get most out of mid-band 5G on Cloud RAN.

Authors:



Mathias Riback
Head of Product Line,
Cloud RAN



Oscar Toorell
Head of Technology,
Cloud RAN

06

The innovation
potential of Non
Real-Time RAN
Intelligent Controller

The innovation potential of Non Real-Time RAN Intelligent Controller

Non Real-Time RAN Intelligent Controller (Non-RT RIC) will realize a new intent-based network management framework in the O-RAN architecture. We believe Non-RT RIC will bring the most value to the industry by enabling new services and improving user performance. It can influence the RAN behavior, enabling a wide variety of new use-cases, and new capabilities that does not exist in current mobile networks.

The value of intent-based management of RAN

As the mobile networks evolve from 4G to 5G, service providers face a new reality of requirements and expectations at the same time as new technologies can help them mitigate the effects of increased complexity. In a typical radio-access network there are literally millions of decisions taken every second about which user to serve over the radio interface and how. Each of these decisions contribute to the service quality and the prioritization among users and services in case of conflicts.

Traditionally these micro decisions are governed by a combination of supplier design choices and network configuration parameter settings done by the service provider. In the relatively simple 2G systems, the effect of a configuration change was mostly possible to understand. In today's more sophisticated multi-service 5G networks it is virtually impossible, in a cost-efficient manner, to predict the effect a given set of configuration changes will have on the end-user services.

However, the intent of the RAN remains the same, to offer connectivity to the service providers' customers in a profitable way. The idea of intent-based management for RAN is to evolve the configuration of the RAN from setting technical parameters (such as hand-over thresholds) and instead allowing service providers to specify the connectivity service itself, prioritizing across users and services, based on business intent. Non-Real-Time RAN Intelligent Controller (Non-RT RIC) is a concept developed by O-RAN Alliance to realize Intent-based management, built on principles of automation and AI and Machine learning. The Non-RT RIC brings genuinely novel capabilities to the system and addresses use cases that were previously out of reach, with the ability to set policies per user and data enrichment information for RAN optimization.

Intent-based management based on Non-RT RIC can be applied to Cloud RAN to enable a high degree of network programmability and can equally well be applied to purpose-built RAN to enable a wide variety of automation and optimization use-cases that are not possible today.

Deep dive on the non-RT RIC technology

The Non-RT RIC is part of the Service Management and Orchestration system (SMO) as defined by the O-RAN Alliance and consists of a platform plus a set of microservices (named rAPPs by O-RAN Alliance) that represent the intelligence (see figure 1). The design of the system is based on these principles:

- 1. Access to information:** There is a wealth of contextual information – not available in the RAN – with potential to improve the radio-resource management and RAN performance in general. This includes application-level information, cross-domain information, UE positions and mobility trajectories as well as external information.
- 2. Dynamic optimization:** Traditionally management and orchestration have been done on the timescale of hours. With automation and improved interfaces, the Non-RT RIC can optimize the RAN on a time scale down to half a second.
- 3. User level service assurance:** Optimizing the RAN on a user level (in addition to per-node level) enables the Non-RT RIC to address a wide set of use-cases that were previously out of reach.
- 4. AI/ML over engineered programs:** The intelligence in RAN control is gradually moving to AI/ML-based software, and the Non-RT RIC is designed for AI/ML from day one.
- 5. Innovation for openness:** It is possible to build an open eco-system of intelligent controller software where applications (rAPPs) feed each other with data and insights

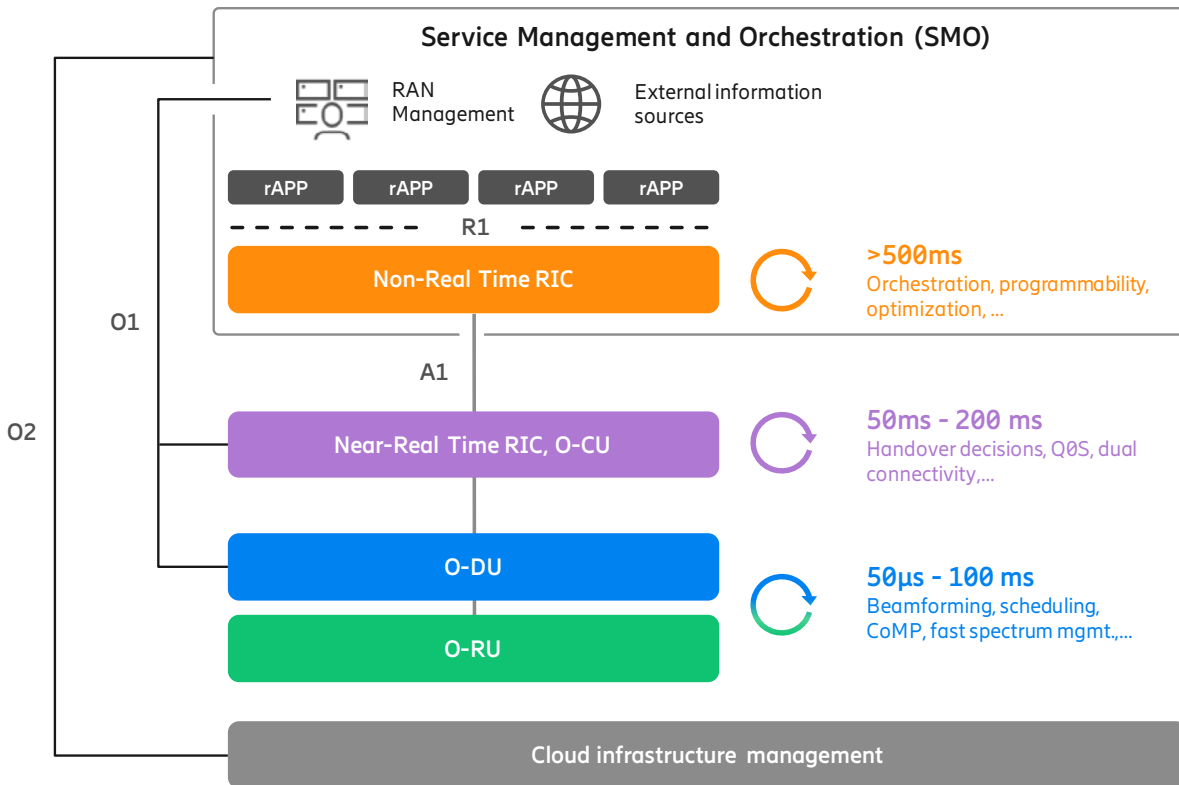


Figure 1 - The Non-RT RIC uses the interfaces A1, O1 and O2 to interact with the RAN on timescales above 500ms

The placement of the Non-RT RIC in the SMO and not in the RAN is to secure access to contextual data and use it to optimize the RAN, something that the RAN nodes CU, DU and Near RT-RIC can't do. One example is the coordinated optimization of radio and transport to adjust user-level throughput based on application performance and contextual prioritization, which can't be achieved in existing networks.

Near-RT RIC is a new element, defined in O-RAN, inside the RAN aiming for radio resource management. Currently, it is unclear what performance benefits Near-RT RIC can provide beyond that of a 3GPP technology.

Shorter time scales improve network control and user experience

The Non-RT RIC introduces a new control loop in the system on a time scale between traditional management (minutes/ hours) and the RAN-internal control loops (between 50 microseconds and 100 milliseconds). Setting interacting control loops on different time scales is a proven design pattern to maintain system stability and separation of concern between the loops so by setting the Non-RT-RIC time scale to around half a second we avoid potential conflicts and instabilities that

can occur. As a contrast, the CU-CP and the Near-RT RIC both act on the same 100 ms time scale which is a potential source of conflicts. Non-RT RIC adds the ability to send optimization policies enrichment information targeting an individual user by adding the new A1 interface that complements the per-node configuration interface O1.

Combining user-level granularity, shorter control loop time scale and the access to outside-RAN information, the Non-RT RIC brings novel system capabilities that were not available in any earlier generations of mobile systems. Adding rapid development of AI/ML technologies to the mix leads us to predict that the Non-RT RIC will drive significant innovation and address use-cases that were previously not possible to solve and that we will look into in the next section.

To capitalize on this, the Non-RT RIC is designed with openness and innovation in mind and follows a micro-service architecture where intelligent applications (the "rAPPs") are deployed on top of the Non-RT RIC platform. A standardized information model and the open interfaces A1, O1 and O2 makes it possible to have rAPPs from both incumbents, third parties and service providers alike, that together, can steer any RAN.

rAPPs are integrated to the rest of the system through what is called the R1 interface, presently technically studied in O-RAN. The R1 interface will support easy onboarding of rAPPs and will give service providers controlled authorization so that rAPPs can be onboarded at low cost, trialed in limited parts of the network, evaluated and then either scaled out to a larger part of the network (if successful) or fail early, pulled out and re-designed (if not so successful).



Categories of use-cases enabled

The non-RT RIC enables a wide variety of new use-cases. Some of these use-cases can be seen as enhancement of existing functions, while other are entirely new. On a high-level, the non-RT RIC use-cases can be categorized as follows:

1. **User-level policies:** The Non-RT RIC can influence the RAN behavior by means of declarative policies, per user and per user group. This enables optimized treatment of individual users based on their special requirements. In the next section we will explore this in detail.
2. **RAN enrichment data:** The Non-RT RIC can provide additional information to the RAN, for better resource optimization. For example, the Non-RT RIC can provide information about weather conditions to the RAN so that it can change behavior, such as prioritizing frequencies that are less affected by rainy conditions.
3. **Enhanced self-organizing-network (SON) functions:** Using fast closed-loop automation the Non-RT RIC can enable more advanced SON use-cases. For example, the dynamic orchestration of radio and transport together can provide more efficient end-to-end resource usage.

Service assurance for first responders – a real-life application of Non-RT RIC

Imagine that a large fire is ongoing in an industrial area. First responders have arrived on-site, bravely fighting the fire. Each first responder streams a live video, creating situational awareness for the task force leader. The live feed of some first responders are more important than others – the closer they are to the fire, the more important information.

Non-RT RIC provides the ability to prioritize radio resources all the way down to the individual first responder based on contextual, external information such as the location of the fire (e.g. captured through location sensors in the building) and the location of the first-responder (e.g. captured through GPS). In existing networks this is not possible. Today, it is only possible to perform service assurance per group of users (i.e. all the first responders) or per node (i.e. everyone close to the fire).

With Non-RT RIC, an rAPP can be designed to provide this kind of granular service assurance, improving the quality-of-experience where it is really needed and, in this case, potentially taking out the fire faster. It also opens up completely new doors for innovation and business opportunities for service providers.

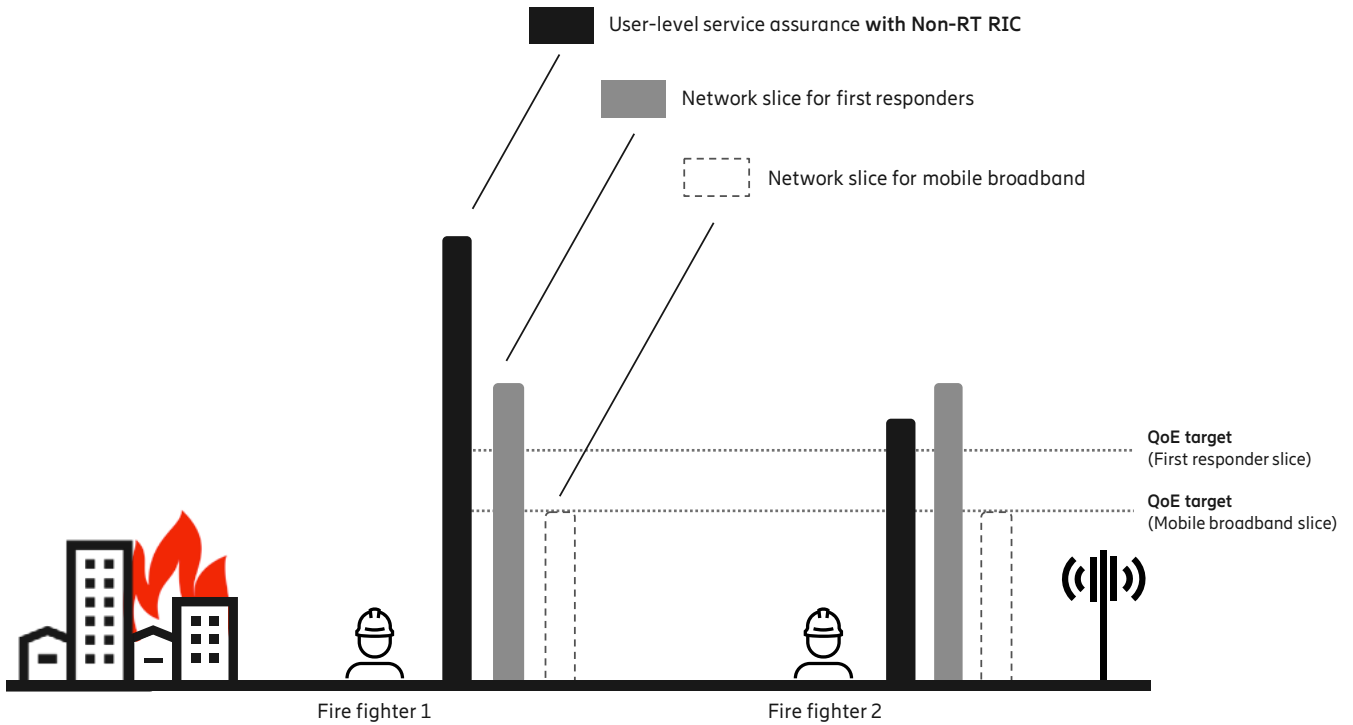


Figure 2

Ericsson is very excited to see the industry move towards an architecture that enables an intent based management of the Radio Access Network, AI-enabled closed-loop automation and end-to-end optimization. We believe this will bring industry value by enabling new use-cases and improving

end-user performance. Service providers can benefit from this innovation no matter the type of RAN infrastructure - purpose-built, cloud native or open. Ericsson stand ready to guide and collaborate with our customers as they embark on this journey.

Authors:



Erik Westerberg
Senior Architect



Matteo Fiorani
Strategic Product Manager,
Cloud RAN

07

Next-generation
Cloud RAN
Management and
Orchestration

Next-generation Cloud RAN Management and Orchestration

As the global telecoms industry begins to deploy fifth-generation wireless networks, technologies such as cloud-native RAN and automation platforms will play major roles in their evolution. In this blog post, we'll cover impact on management and orchestration, the challenges that emerge, our view on the future of management and orchestration and what Ericsson brings to the table.

5G evolution realizing new values

5G is more than radio and core. It encompasses virtualization, management and orchestration, transport, software-defined networking (SDN), and adaptive policies powered by artificial intelligence (AI) and machine learning (ML). As more and more technological innovations enter the market, 5G will continue to leverage the capabilities of these technologies to evolve further.

The 5G networks are designed to serve consumers and enterprises alike and as we add capabilities like network slicing, edge computing, service exposure and new RAN applications, we are in effect creating a new innovation platform. An innovation platform that will enable exciting new business models and use cases that will deliver innovative services to countless industry verticals.

The impact on orchestration

The architecture of the 5G network is substantially different from previous network generations. One key change is the introduction of micro services-based architecture and network functions distributed over different cloud infrastructures. In addition, many services will go across different network functions and domains in the network and they need to work in unison to ensure end-user experiences. This impacts the way we manage and orchestrate the networks.

The openness and disaggregation started in transport and core with the introduction of Software Defined Networking (programmable networks) and virtualization. A similar shift is expected to happen in the access domain and certain aspects need to be considered in the orchestration systems as well.

Open source and standards become even more important than previously in this context. Ericsson is a leading supporter of several open-source initiatives. These include the Open Network Automation Platform (ONAP) and Open Radio Access Network (O-RAN). We also work hard to encourage industry alignment to ensure speed of adaptation and realization of values from the 5G platform.

The challenges from an orchestration perspective

As service providers' networks evolve, and new technologies are added, the network is becoming ever more sophisticated and complex.

As network functions are virtualized, networks will end up consisting of Physical Network Functions (PNF), Virtual Network Functions (VNF) and Cloud native Network Functions (CNF).

Network-related services and end-user services will flow across different network functions and multiple domains.

Services will become more advanced with the introduction of network slicing, exposure and RAN applications.

Multiple domains (radio, transport, core and cloud infrastructure) end up having their own OSS systems for different reasons.

The challenge that arises is that service providers need to deliver on stringent -service level agreements towards enterprises, with many different business models and multiple partners. Orchestration and management need to deliver on this challenging task.

What's needed is an intent-based AI/ML-enabled automation platform that is open and highly flexible to scale. This Automation platform is key to manage these complex networks and services and to achieve the necessary network and service flexibility and speed up time to market.

As cloud-native RAN is introduced in the network, Ericsson brings the knowledge and capability from our experience of managing and orchestrating end-to-end 5G networks that includes cloud-native 5G core.

Key considerations for services providers

As Cloud RAN is introduced, certain aspects need to be considered in the management and orchestration systems.

Artificial intelligence and machine learning will play important roles in traffic steering and optimization decisions that will improve network efficiency and customer experiences and is an absolute prerequisite in next generation management and orchestration.

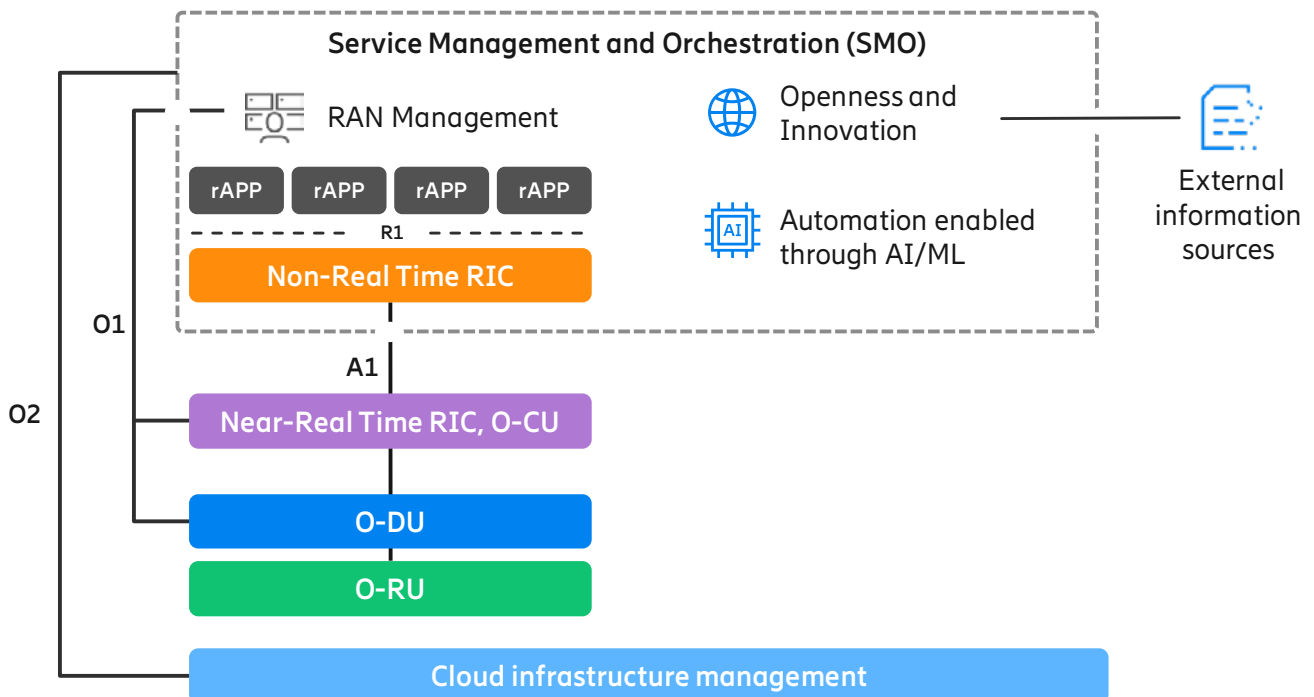
Since 5G will provide services to many vertical industries, the use cases demand Quality of Experience (QoE) such as high bandwidth, low latency, high security and ultra-reliability. Managing QoE proactively will be central to maintaining SLAs and continuous optimization is needed. The introduction of cloud-native RAN and Service Management and Orchestration will add new capabilities to assure best possible Quality of Experience.

The virtualized radio and core will be deployed on many different cloud infrastructures and distributed multiple data centers and far edges of the networks very close to where the services are being consumed. The orchestration system will need to work across all these different cloud environments and manage the end-to-end lifecycle management of network functions and services.

These are some key considerations that service providers need to think of as they evolve their management and orchestration. In the next section we will explore our view on how to realize this and the values we bring.

Ericsson’s view on next-generation management and orchestration

The next-generation management and orchestration (as defined by O-RAN Alliance) consists of a number of interworking components – the Service Management and Orchestration (SMO), Non Real-Time Radio Intelligent Controller, rAPPs and the interfaces O1, O2 and A1 allowing for communication between these components.



Service Management Orchestration

Ericsson brings their unique experience in building, deploying and managing cloud native networks and orchestration systems to enable seamless and modular introduction of Service Management Orchestration (SMO) for both greenfield and brownfield networks. We work closely with our customers to create unique use cases, while at the same time driving rapid innovation by enabling an open app ecosystem for both customers and partners. SMO the management and orchestration for the future and have a few very important characteristics:

AI/ML capabilities will be crucial in SMO for innovation, automation and operational efficiency. A key capability to achieve a high degree of automation is to use artificial intelligence and machine learning together with advanced policy management and we call that intent- based AI/ML.

To achieve a high degree of automation, Ericsson brings unique know-how to package AI/ML powered use cases into hardened and scalable solutions using our telco expertise.

SMO should have infrastructure visibility and be able to scale compute resources up and down. This also helps optimize or reduce the overall capacity costs.

SMO can handle end-to-end network slicing where RAN slicing is an important component. For example, a SMO can continuously monitor and ensure that a specific enterprise SLA is maintained in an end-to-end network slice that is providing a mission critical, low latency, ultra-reliable service.

SMO needs to be aware of applications and their performance, network behavior, traffic and multiple domain information, and have the ability to dynamically place application workloads across multiple cloud infrastructures.

The SMO need to simultaneously handle life cycle management of PNFs, VNFs and CNFs. In the cloud native world software management changes drastically with the introduction of high degree of automation, continuous integration/continuous development of micro-services. To support very fast introduction of new SW and features into the network, canary deployment and A/B testing will grow in relevance.

Non Real-Time RAN Intelligent Controller

The Non Real-Time RAN Intelligent Controller is key component of the future management and orchestration as defined by O-RAN. Placed inside the SMO layer it can access external information to dynamically use policies to optimize the RAN powered by AI and ML.

The Non Real-Time RIC plays a central role in SMO hosting the ecosystem of applications acting across RAN, Core and Transport, to generate and apply new policies and network optimizations. Additionally, from a pure innovation platform point of view, this component will expose the SMO capabilities and the network data, to a vast variety of applications, that will coexist and collaborate, to achieve new business outcomes.

Finally, it needs to be flexible, aware and secure. It needs to be flexible enough to foster innovation, aware enough to understand the network behavior, and secure enough, since it provides access to operators' networks and data pools.

The power of RAN applications (rAPPs)

As seen above, rAPPs are powerful and important tools in the future management and orchestration. rAPPs can be written for a various management tasks such as assurance, optimization, automation, fault management, performance management, configuration management, software management, domain-specific applications.

The introduction of rApps, made possible by an end-to-end automation platform, brings innovative aspects in the RAN domain. It means that tasks as those above can be realized outside the traditional node context. Working on their own or together with other rAPPs, they can create improvements across all network domains.

In summary, SMO and the next generation of management and orchestration is essential to cope with the rising complexity of data and network systems, achieving not only a more manageable end-to-end network but also achieving the business propositions sought by operators: increased QoE and time to market for services creation and rapid innovation to achieve the full 5G potential.



Ericsson is exploring the possibilities

Ericsson views the rapidly progressing automation capabilities as essential to maintaining operational efficiency while also maximizing the business potential of the increasingly sophisticated 5G networks. Building on our leadership in RAN, core, management and orchestration systems, we bring a unique experience in delivering end-to-end 5G networks and OSS solutions (management and orchestration) to manage the smooth introduction of open and interoperable cloud-native RAN. In addition, Ericsson brings the following key capabilities in its automation platform that is driven by management and orchestration systems:

A programmable network with intent based AI/ML capabilities in management and orchestration solution to drive efficiency and automation.

A management and orchestration system that manages the PNF, VNF and CNF that comes with the introduction of cloud-native RAN.

An O-RAN aligned management and orchestration system that can manage radio, transport, core, cloud infrastructure and multi-layer networks.

Ability to develop unique use cases, onboard new applications in the management systems and rapid innovation by enabling an open app ecosystem for customers and partners.

An intelligent management and orchestration system that can provide network and service flexibility to support various new business models and use cases.

Ericsson is exploring the possibilities offered by SMO architecture to create an attractive solution for next-generation management and orchestration and accelerate the automation journey together with our customers.

Authors:



Balaji Ethirajulu
Strategic Marketing
Manager, OSS



Pedro Sousa
Strategic Product Manager,
Cloud RAN

08

Leverage Ericsson
Spectrum Sharing
benefits in Cloud
RAN

Leverage Ericsson Spectrum Sharing benefits in Cloud RAN

The award-winning innovation – Ericsson Spectrum Sharing – will take a leap into Cloud RAN. Follow-us and get to know more on how Ericsson Spectrum Sharing will seamlessly interwork between purpose built (Ericsson Radio System) and Cloud RAN.

Ericsson Spectrum Sharing – running on Cloud RAN

Ericsson Spectrum Sharing (ESS) is an innovative, award-winning, and market-disruptive solution that has set a new standard in the wireless industry. ESS enables service providers to utilize its scarce spectrum assets in the most efficient way possible by instantly – on a millisecond level - sharing spectrum between 4G and 5G.

ESS introduces a new way of rolling out 5G Frequency Division Duplex (FDD) spectrum that re-uses hardware, spectrum, sites and increases coverage of the mid/high band so that operators can shift Capex investments from new sites to new 5G use-cases. Ericsson Spectrum Sharing brings 5G to everyone, everywhere, much faster.

So, the question is, will this powerful solution work on Cloud RAN, leveraging the benefits it brings while keeping the value of ESS? The very short answer to this question is – yes!

In this blogpost we will explore the details on how ESS will work seamlessly across our Cloud RAN and our purpose-built Ericsson Radio System, and how we can take advantage of decisions made already when designing the original solution, as well as our deep radio knowledge and experience to move this innovation into the cloud paradigm.

How Ericsson Spectrum Sharing works

To understand how we extend ESS from purpose-built infrastructure to Cloud RAN, we must go back to the design of the original solution. In short, the power of ESS is to realize a combined 4G/5G carrier, that both 4G and 5G devices can use. This entailed a number of challenges:

Secure the right 3GPP toolbox required - There was and is no feature called 4G/5G sharing in 3GPP. Therefore, we worked with the 3GPP standardization to secure the needed functionality. This resulted in a list of tools as a baseline for the initial ESS support, mainly on the device side.

Get day 1 support from the device ecosystem for a swift rollout - Early discussions with the major device chipset suppliers was critical so that time plans, interoperability tests, and early field testing were aligned according to needs.

Additional Ericsson innovations beyond 3GPP were needed - Additional innovation beyond 3GPP was needed to be designed. Some examples include avoidance of colliding signals on the air-interface, initial access additions for the period until the device is configured and aware of dynamic spectrum sharing and scheduler coordination between 4G and 5G with 1ms periodicity.

Ensure total performance of both 4G and 5G - On the journey to ESS some very tricky performance challenges had to be solved which resulted in 1ms scheduling periodicity between scheduler decisions and Resource Block Group level allocation in the air-interface. Last, but not least, 4G feature parity had to be secured in order to not jeopardize 4G performance as delivered from legacy deployments

Efficient update of Ericsson Radio System - ESS was designed to be deployed as a remote software installation, without site visits and can be deployed using existing basebands and radios in the field when adding 5G to an existing site. To achieve this, we used our unique advantage of 5G ready Ericsson Radio System radios and RAN Compute basebands. This means high usability of the solution and also reduced operator Capex and Opex.

Optimizing and future proofing the architecture - As described above, ESS means combining 4G and 5G data streams to one carrier on the air-interface. The physical placement of this combination function was an important decision. There are three principal options – on the Ericsson Radio System mixed mode basebands, on an external unit between Ericsson Radio System basebands and radios, or inside the Ericsson Radio System radios.

We decided to implement it in the mixed mode baseband to avoid duplication of CPRI lines out of the basebands. This would have resulted in higher costs and site visits, but the architecture and realization are still prepared for all three options to leave room for future innovation. For the same reason, the architecture is also prepared to be implemented on several deployments, not only on a single mixed mode baseband, but also on multiple, single mode basebands. Without explicit requirement, the architecture is prepared even for the execution of 4G or 5G on Cloud RAN.

See figure 1 for the final implementation for ESS on a purpose-built platform.

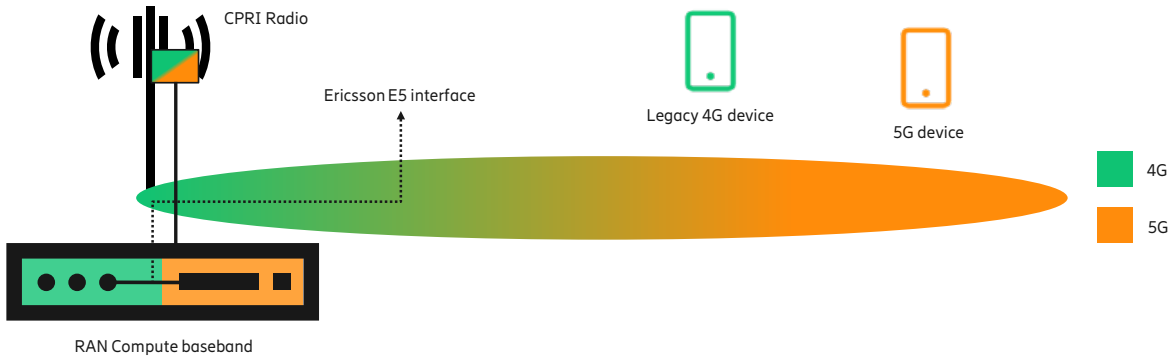


Figure 1

Ericsson Spectrum Sharing implementation extended to Cloud RAN – How do we do that?

Ericsson now brings ESS also to our Cloud RAN. This means that service providers leveraging both Ericsson Radio System as well as Cloud RAN by Ericsson can enjoy the full benefits of ESS.

With Cloud RAN we realize RAN functions (such as ESS) on a generic compute hardware platform, Commercial off-the-shelf (COTS), instead of a purpose-built hardware platform. Cloud RAN also means managing virtualized RAN applications based on cloud-native principles, e.g. running several, loosely coupled, microservices in containers, handling resilience and scaling by replication and independent life cycle management following DevOps principles.

Cloud RAN offers a very attractive alternative for deploying 5G RAN networks and at the same time get the benefits of a generic compute platform (The four components of Cloud RAN). Our Cloud RAN solution complements and extends our existing and well proven purpose-built offering.

We are now realizing an extension of ESS also into Cloud RAN and the solution is a direct continuation of the concepts and implementation on the ERS side. As mentioned above, when we decided on the ESS architecture for the purpose-built platform, we considered a future where it could be deployed on multiple architectures.

Also, there is no need for any 3GPP additions. Cloud RAN will re-use the necessary specifications and the Cloud RAN with ESS implementation will be completely transparent to the device ecosystem. This means that there is no need for repeated interoperability with associated field testing, resulting in short time to market.

ESS on Cloud RAN will use the same innovation toolbox for initial access, scheduler coordination and Quality of Service adaptation and there is no performance penalty due to Cloud RAN – the crucial 1ms scheduling period is preserved and as well as the granularity of spectrum allocation (Physical resource block allocation). However, there were some clever engineering tricks involved to make it happen.

Key innovations for running Ericsson Spectrum Sharing on Cloud RAN

A key innovative solution to bring ESS to Cloud RAN is Ericsson Elastic RAN and specifically the ability to enable instant 1ms sharing between different platforms. With our Elastic RAN as an enabler, ESS on Cloud RAN is realized with 5G executing on Cloud RAN whilst 4G (or both 4G and 5G) is executed on the Ericsson Radio System RAN Compute baseband with an inter-system interface in between. This configuration can be seen in Figure 2.

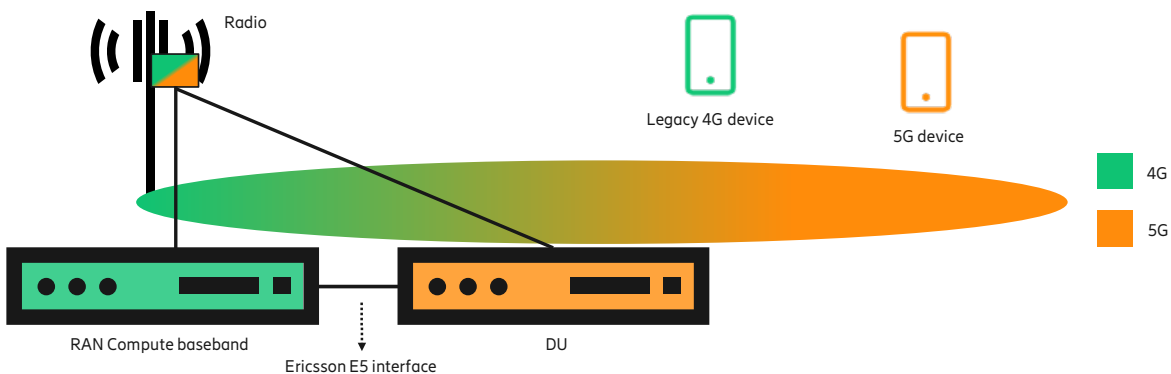


Figure 2

The 1ms sharing between different platforms enabled by Ericsson Elastic RAN is critical for a successful deployment of ESS since the 5G Cloud RAN scheduler needs to interwork with the 4G scheduler on that timeframe.

In the RAN Compute platform, the communication between the 4G and 5G parts is done through the Ericsson unique E5 interface that is very stable and continuously proven in the field, since more than 80 service providers already has the ESS software.

As we introduce ESS with 5G on Cloud RAN, the same reliability of communication between 4G and 5G is necessary. Since the original architecture was designed with this in mind, it makes perfect sense to use the Elastic RAN technology to re-use and extend the field proven internal E5 interface to go beyond the borders of the purpose-built baseband and to also allow connections with the Cloud RAN platform. We see this as a powerful proof-point of our technology leadership, expertise and our commitment to work with our customers to seamlessly move from one technology paradigm to the next.

Ericsson Elastic RAN and NR Advanced RAN Coordination as enabler for further network evolution

But it does not stop here. By using our Elastic RAN technology, the Cloud RAN can not only interwork with a single RAN Compute baseband, but with several, which can be single mode basebands or mixed mode basebands (ref Fig 1.) or a mixture of both. Our award-winning Elastic RAN technology has also been further enhanced for NR Carrier Aggregation features and renamed to NR Advanced RAN Coordination. Thus, NR Advanced RAN Coordination will enable the next step of our network evolution such as Carrier Aggregation between Ericsson Radio System and Cloud RAN.

The value of choosing Ericsson Cloud RAN

Ericsson Spectrum Sharing is a breakthrough in mobile communication and changes the way spectrum can be utilized. We believe that this will be the new normal when new technologies are introduced. Now, we take a next step by bringing ESS to our Cloud RAN.

With the extension of an installed base of Ericsson Radio System to a Cloud RAN based architecture, service providers can expect to leverage the full capability of Ericsson Spectrum Sharing - providing 4G and 5G spectrum sharing on an instant - 1ms - basis.

In addition, our solution for extending Ericsson Spectrum Sharing onto Cloud RAN also enables service providers to fully utilize already made investments and keep flexibility when rolling out 5G.

ESS will work seamlessly, no matter if it runs on purpose-built hardware or on a combination of purpose-built and cloud infrastructure.

Our underlying design philosophy and technology will secure a seamless network evolution for service providers' beyond ESS. As the next step, we will bring our innovative Carrier Aggregation across both purpose-built and cloud infrastructure.

This is only the start of our journey, with more to come that will unlock the full potential of networks beyond today.

Authors:



Sibel Tombaz
Head of High-Band and
Active Antenna Systems



Gunnar le Grand
Head of Distributed Unit
Base and Low Band

Contact your local
Sales team:



Watch Tech
Unveiled on video



For all ordering details refer to your Ericsson sales representative.

Ericsson is a world leader in communications technology and services with headquarters in Stockholm, Sweden. Our organization consists of more than 100,000 experts who provide customers in 180 countries with innovative Solutions and services. Together we are building a more connected future where anyone and any industry is empowered to reach their full potential. Net sales in 2017 were SEK 201.3 billion (USD 23.5 billion). The Ericsson stock is listed on Nasdaq Stockholm and on NASDAQ in New York.

ericsson.com