5G Americas
1750 Avenue, N.E., B220
Bellevue, WA 98004
O: 425 372 8922
www.5GAmericas.org

February 10, 2021

Ms. Rebecca Dorch
National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Avenue, N.W.
Washington, D.C. 20230

Re: *5G Open Stack Challenge Notice of Inquiry*

Dear Ms. Dorch:

5G Americas, the voice for 5G and LTE in the Americas, writes to submit its white paper on the *Transition Toward Open and Interoperable Networks*[1] into the record of the National Telecommunications and Information Administration's ("NTIA") *Notice of Inquiry* on a 5G Open Stack Challenge. 5G Americas facilitates and advocates for the advancement and transformation of LTE, 5G and beyond throughout the Americas.[2] We do so as a Market Representative Partner of the Third Generation Project Partnership (3GPP), a consortium of standards-setting bodies from around the world.[3]

As NTIA recognizes in its call for comments and recommendations, 5G is vitally important to U.S. economic and security interests. NTIA notes that its 5G Open Stack Challenge is under the sponsorship of and in collaboration with the Department of Defense's ("DoD") 5G Initiative, and is intended to accelerate the development of an open 5G stack ecosystem in support of DoD missions. 5G Americas appreciates that in addition to the myriad 5G applications that will be consumed by the enterprise and individuals, government users like DoD and other agencies will benefit from customized 5G applications.

The pace of deployment of 5G in the U.S. and around the world is quickening as we move into 2021, with over three-quarters of Americans having access to a 5G signal today. In

---

[1] *Transition Toward Open and Interoperable Networks*, 5G Americas (November 2020) ("Open Networks") https://www.5gamericas.org/wp-content/uploads/2020/11/InDesign-Transition-Toward-Open-Interoperable-Networks-2020.pdf.

[2] 5G Americas Board of Governors includes AT&T, Cable & Wireless, Ciena, Cisco, Crown Castle, Ericsson, Intel, Mavenir, Nokia, Qualcomm, Samsung, Shaw, T-Mobile USA, Telefónica, VMware and WOM.

[3] 3GPP unites seven telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC), known as "Organizational Partners" and provides their members with a stable environment to produce the reports and specifications that define 3GPP technologies.

the U.S. and worldwide, there are over one hundred and fifty 5G network deployments. As a technology then, 5G has a degree of maturity, relative to the work already undertaken at 3GPP through Releases 15 – 17. Industry standardization of Open Radio Access Networks (ORAN) began more recently. NTIA and DoD's *Inquiry* would benefit from an understanding of the work undertaken by industry to date, and a greater exploration of the various components of ORAN. With that in mind, 5G Americas submits the attached white paper, to provide that backdrop of existing work on ORAN standardization, and also some of the technical considerations that mobile operators will face. The *Open Networks* white paper examines aspects of software and hardware disaggregation, open interfaces, multi-vendor interoperability, the ORAN ecosystem, and the role of Artificial Intelligence ("AI") and Machine Learning ("ML") in network management and automation. *Open Networks* also provides a list of operators' ORAN trials and deployments.

Open Radio Access Network and Open Stack

NTIA asks for comments on creating a 5G Open Stack Challenge. Most non-technical telecom policymakers are aware of discussions on ORAN. But as NTIA and DoD no doubt appreciate, an *ORAN* is not synonymous with *Open Stack,* but both could be components of an open system. ORAN architecture combines a modular base station software stack with off-the-shelf hardware, which allocates baseband and radio unit components from different suppliers to operate together, whether or not elements of the RAN are virtualized or disaggregated. NTIA asks in its call for comments "What 5G enabling features should be highlighted in the Challenge, such as software defined networking, network slicing, network function virtualization, radio access network intelligent controller, radio access network virtualization?" Because "open stack" is a broad term, and ORAN is at an early phase of commercial adoption by the global wireless community, 5G Americas believes it is worth noting how these terms inter-relate. Software Defined Networks ("SDN"), network slicing, and network function virtualization ("NFV") are features of 3GPP technology, deployable in either LTE (e.g. 4G) or 5G networks. An Open RAN may actually facilitate the efficient allocation of network resources to specific customer use cases via network slicing,[4] but such slicing, as well as SDN and NFV, are capabilities that can be deployed without an "open stack" per se. Likewise, a RAN Intelligent Controller—or RIC—or vRAN are system components that can be deployed in either a traditional or an ORAN system.

Interoperability

NTIA notes the challenge of interoperability in an open system, when it cautions in its request for comment that "interoperability among the community's implementation is not guaranteed." NTIA asks "What are the incentives in the Open 5G Stack ecosystem" that would "promote interoperability?" *Open Networks* reviews the incentives to deploy ORAN, such as

---

[4] See *Open Networks* at 7.

providing an operator additional flexibility to meet the requirements of a customer's particular 5G application. An operator may have customers with differing network requirements relative to performance, capacity, latency, etc. These various requirements could benefit from an open and flexible, software-programmable RAN architecture. With open programmable RAN interfaces, a centralized and virtualized baseband can provide pooled virtualized network functions that can dynamically allocate different resources to create an on-demand architecture. But these business incentives alone are not the key to achieving interoperability between different vendors' hardware and software. Additional development is needed for operators to manage the possibly increased complexity of heterogenous networks, with radio and baseband units and software from different vendors. These challenges are being worked on by various technical bodies and will be ameliorated with the development of more advanced AI software and corresponding ML

Conclusion

As 5G Americas *Open Networks* white paper catalogues, there is a great deal of work being done in industry development and other standards bodies on ORAN technology. To best channel the innovation from industry, NTIA and DoD should monitor and engage when appropriate in these industry discussions. 5G Americas cautions NTIA and DoD against developing USG-specific standards for open networks, but instead encourages the U.S. government to work within the rapidly emerging industry-developed ecosystem. 5G Americas supports a trusted, secure and innovative U.S. and allied 5G ecosystem and we commend NTIA and DoD for their efforts to unify the open 5G stack community by creating the 5G Challenge. We hope *Open Networks* will inform you on the work streams already underway by industry, so you can leverage this existing progress as DoD defines any customized mission-critical 5G requirements.

Best regards,

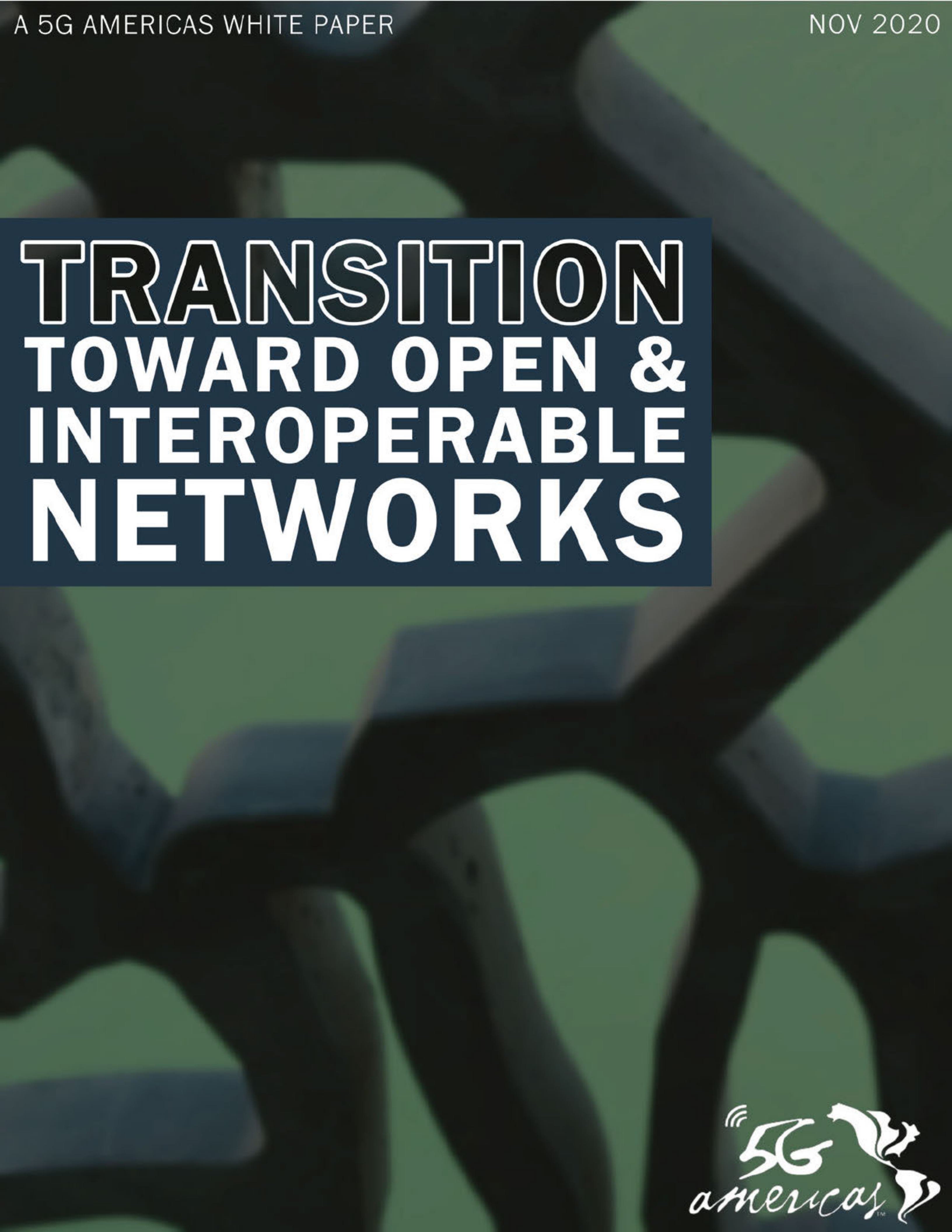
Chris Pearson

Chris Pearson
President, 5G Americas

cc: Evelyn Remaley, Acting Administrator
Charles Cooper, Office of Spectrum Management
Sheryl Genco, Institute for Telecommunication Sciences

A 5G AMERICAS WHITE PAPER

NOV 2020



TRANSITION TOWARD OPEN & INTEROPERABLE NETWORKS

Contents

1. Overview

1. Overview

This paper discusses a popular topic in wireless cellular communications today - Open Radio Access Networks (Open RAN). Starting with the introduction, ***Transition Towards Open and Interoperable Networks*** highlights key architecture aspects focused on disaggregating software from hardware and open interfaces to allow multi-vendor interoperability. After a brief look at various ecosystem bodies involved with Open RAN, there is an overview of various global operator trials and deployments taking place, as well as a discussion on key motivations and challenges that operators face as they consider the move from a traditional proprietary network to a multi-vendor Open RAN platform. Finally, the paper looks at the role of Artificial Intelligence (AI) and Machine Learning (ML) initiatives in enabling Open RAN architectures and associated functions for Self-Organizing Networks (SON), management, orchestration and automation to meet the variety of use cases for 5G and beyond.

Existing operator networks have many deployment models, which can be distributed or centralized, non-virtualized, virtualized or cloud enabled. While these have traditionally been implemented using proprietary platforms from one or more vendors, an Open RAN approach is applicable to every one of these models, but is neither dependent on, nor implied by, any of these models. For example, it is feasible to have an Open RAN that is not virtualized, and likewise a virtualized RAN that is not open.

Additionally, operators have pursued interoperability using the RU/DU interface to enable mix and match of RU and DU+CU vendors. This enables vendor interoperability but does not imply hardware disaggregation nor open hardware and open software. The general direction of Open RAN considers all these things. For example, some incumbent vendors today are implementing interfaces on their existing proprietary hardware platforms that are completely compliant with Open RAN specifications, while at the same time developing virtualized solutions. Both can happily co-exist.

In addition to having a range of deployment models, operators also have varying timeframes and are currently at various stages of Open RAN deployment, from simply exploring, or conducting trials to actually deploying it. This interest in Open RAN is driven by expectations of widening the supply chain, potential cost-effectiveness, shorter development cycles, faster time to market, greater innovation, improved security, and possibly better performance with best of breed vendor selection. There is no single path or timeframe for Open RAN deployment.

While the operator community overall is interested in Open RAN, each operator has a different mix of strategies, business objectives, spectrum, technical architectures, network configurations, and deployed components. This naturally leads to a range of deployment paths and timeframes, with some operators starting deployments in new networks or selected network segments on their way to scale. As the operator members in the 5G Americas workgroup can attest, deploying and operating a single vendor network has its fair share of pain points, with significant operator resources and investment needed for activities that are often taken for granted such as custom solutions training, troubleshooting, performance monitoring, feature testing and validation etc.

There are also dependencies on proprietary roadmaps and features for a given technology, and in migrations to future technologies. Operators expect these challenges to get amplified with multiple vendors in the mix when Open RAN is considered. While it is by no means an easy proposition, even for greenfield operators, the initiative is relatively much harder for brownfield operators who have to contend with the significant hassles prevalent in operating current “closed” networks while integrating a new open one. Nonetheless, Open RAN presents itself as an alternative worthy of consideration any time operators embark on network upgrades or enhancements.

2. What is Open RAN?

2. What is Open RAN?

Open RAN is a concept that encompasses interoperability of open hardware, open software and open interfaces. One way of implementing Open RAN is through the disaggregation of software from hardware which allows Radio Access Network (RAN) software to run on any common hardware platform such as those based on Intel x86 and ARM architectures. Open RAN can also be implemented using specialized hardware such as Application Specific Integrated Circuits (ASICs) and Digital Signal Processors (DSPs) provided the nodal specifications are open and interoperable. This disaggregation also applies to other hardware components such as Field Programmable Gate Arrays (FPGAs) and Graphics Processing Units (GPUs) so that the abstraction layer is common and open.

Open RAN platform software may be based on open source code or code from best-in-class coders, whether a person or a company, that can contribute to its development. This open software platform can be used to realize all RAN functions. Open RAN interfaces allow for disaggregating RAN into functional components such as a Radio Unit (RU), Distributed Unit (DU), Centralized Unit (CU) etc. to facilitate an open user plane, control plane, synchronization plane as well as management plane.

RAN openness enables a multivendor ecosystem. In turn, this ecosystem will allow innovation to thrive. Standardization targets global economies of scale across worldwide vendor supply chains. Together, standardization and openness promise to drive down costs and speed up innovation for 5G.

A typical RAN is composed of radio and baseband units. The interface between the radio and baseband units in existing RAN systems is based on published partial standards. However, its implementation in current products have proprietary variations that make multivendor interoperability between the baseband and radio impossible. In other words, if an operator uses vendor A for the radio, typically, it must use the baseband from vendor A as well. In addition, the software that runs on the baseband hardware is not designed to be run on another vendor's hardware. This creates vendor lock-in due to the proprietary vendor-specific product realization of the interface specification.

Hardware and software disaggregation in centralized, RAN deployments allows flexibility in scaling when compared with integrated platforms. An Open RAN approach targets enabling innovation, competition and driving down costs with a larger global vendor supply chain. It should be noted that for hardware and software disaggregation cases (where cloud infrastructure is procured separately from the RAN application), the system integration complexity, life cycle management and associated costs/cost improvements are factors that cannot be overlooked.

Open RAN seeks to create an avenue for introducing advanced RAN features and capabilities beyond the standard centralized feature sets and RAN scheduler-based features. This is achieved by leveraging a programmable open-software development approach. The benefits targeted by RAN programmability are achieved by the introduction of additional interfaces and operational complexity, requiring inter-operability testing and operational hardening of new products and features.

Finally, in the context of 5G RAN, an Open RAN approach can provide additional flexibility to meet 5G application requirements. Specifically, 5G supports vertical applications with different network requirements for performance, capacity, latency etc. Examples include Ultra-Reliable Low-Latency Communication (URLLC) and Massive Machine-Type Communication (mMTC) applications which are not fully supported in 4G networks. This requires a flexible software-programmable RAN architecture to accommodate the various vertical applications. For instance, with open programmable RAN interfaces, a centralized and virtualized baseband can provide pooled virtualized network functions that can dynamically allocate different resources through network slicing to efficiently and effectively create the architecture on-demand to fit the needs of the applications.

2.1 Goals of Open RAN

The goal of an Open RAN architecture is to evolve RAN networks towards an open and intelligent RAN, while complying with the 3GPP standards. Open RAN targets include:

Open, standardized interfaces between:

- the radio and the baseband
- the Element Management System (EMS) of the radio/baseband and the network management system
- the CU control plane and the CU data plane
- the CU and the DU

Disaggregated software from the hardware:

- to use general purpose processors for the RAN that allows software from different sources to run on one hardware platform
- to enable an open software development ecosystem that can interwork with standardized hardware

Open hardware:

- using standard processors (e.g., x86, ARM CPUs and GPUs) that allow software from different sources to run on them
- that uses standardized racks, chassis, power distribution, and cabling such as those from open19.org, Open Compute Project (OCP), etc.
- that has an open standard coherent accelerator processor interface

Open software:

- that is commercially viable to meet high performing KPI requirements that support real-time system needs
- that leverages adjacent software communities such as Open Networking Automation Platform (ONAP) and other open approaches to utilize existing solutions to speed time to market

One example of open software is the O-RAN Software Community (OSC) which is a partnership between the O-RAN Alliance and the Linux Foundation to support the software development for an Open RAN solution.

2.2 Ecosystem Survey and Implications

The development of robust, interoperable and Open RANs will require a broad ecosystem of partners, organizations, manufacturers, and operators. This section describes some of the most prominent organizations leading the Open RAN efforts today and their contributions to the development of Open RAN architecture.

2.2.1 O-RAN Alliance

The goal of O-RAN Alliance is to clearly define requirements, specify APIs and interfaces and drive standards to adopt them in the pursuit of an open and intelligent RAN. O-RAN Alliance has nine working groups and three focus groups. To achieve this goal, O-RAN Alliance has working groups to drive the specifications; an O-RAN Software Community (OSC) to drive software contributions; and O-RAN Testing and Integration Centers to drive industry adoption through testing and certification as well as global plugfests.

2.2.1.1 O-RAN Alliance Work Group Structure

Work within the O-RAN Alliance is split and streamlined into several different work groups:

WG1 - Use Cases and Overall Architecture Workgroup

Work Group 1 has overall responsibility for the O-RAN Architecture and Use Cases. WG1 identifies tasks to be completed within the scope of the Architecture and Use Cases and assigns task-group leads to drive these tasks to completion while working across other O-RAN work groups.

The WG1 OAM (Operations, Administration and Maintenance) architecture is based on a non-persistent Network Configuration Protocol (NETCONF) session, where asynchronous notifications are sent using ONAP/3GPP defined VNF Event Stream (VES) events that are signaled using JavaScript Object Notation representational state transfer (JSON/REST).

WG2 - Non-real-time RIC and A1 Interface Workgroup

The focus of Work Group 2 includes both the non-Real Time RAN Intelligent Controller (non-RT RIC) and the A1 interface. The primary goal of non-RT RIC is to support non-real-time intelligent radio resource management, higher layer procedure optimization, policy optimization in RAN, and providing Artificial Intelligence / Machine Learning (AI/ML) models to near-RT RIC and other RAN functions. The A1 interface supports communication and information exchange between the Orchestration/NMS (network management system) layer containing non-RT RIC and the eNB/gNB containing near-RT RIC. A key objective of the A1 interface is to support policy-based guidance of near-RT RIC functions/use-cases, transmission of enrichment information in support of AI/ML models into near-RT RIC, and basic feedback mechanisms from near-RT RIC.

As it relates to WG2, three key use cases are currently being discussed for the non-real time RIC:

- Traffic Steering
- QoE optimization Use Cases
- 3D-MIMO system configuration

WG3 - Near-Real-time RIC and E2 Interface Workgroup

The focus of Work Group 3 is to specify near-RT RIC open architecture and its functionalities, the Radio-Network Information Base and Network Topology, and modular on-boarding of new Control Applications. WG3 also specifies the E2 interface between near-RT RIC and CU/DU stack.

WG4 - Open Fronthaul Interfaces Workgroup

Work Group 4 specifies the Open Fronthaul Interface between Open RAN Distributed Unit (O-DU) and O-RAN Radio Unit (O-RU). It has published the open fronthaul interface specifications for the lower layer split, including Control, User and Synchronization (C/U/S) plane protocols, Management (M) plane protocols, and Multi-vendor IOT specifications, supporting both LTE and 5G NR systems.

WG5 - Open F1/W1/E1/X2/Xn Interface Workgroup

Work Group 5 is refining the definition of 3GPP's F1 interface for supporting the Higher layer Split (HLS) for 5G NR. While ensuring that the 3GPP split interfaces remains truly inter-operable between vendors, the focus is on F1, W1, E1, X2, and Xn interfaces. Significantly, WG5 has defined interoperable X2 profiles to enable multi-vendor deployments of Non-Stand Alone (NSA) 5GNR.

As it relates to WG5, one of the key challenges in 5G introduction is the lack of multi-vendor Non-Stand Alone (NSA) systems that will be typically used in the vast majority of deployments for 5G introduction. This lack of multi-vendor NSA means that operators are forced to source their 5G RAN equipment from the same vendor supplying their 4G RAN.

Hence WG5 publications provide profiles for the eNB to NR Dual Connectivity (EN-DC) related C-plane (Control Plane) procedures and functions together with U-Plane (User Plane) specifications to achieve interoperability among different vendors. The profile specifies the expected behavior of each node (e.g., call flow of each use case, definitions of Information Elements, etc.) which is not specified in 3GPP specifications. The profile specification provided in this document does not violate 3GPP specifications. The publication includes the detailed syntax of 34 message exchanges, defining interoperable profiles of the various optional information elements.

WG6 - Cloudification and Orchestration Workgroup

A "cloudified" or "virtualized" RAN is one that provides the flexibility of deploying multiple software implementations from different vendors on a common CPU-based (e.g., x86/ARM) platform with hardware accelerators (e.g., FPGA/DSP/ASIC/GPU) for specific functions, and conversely, allows multiple physical deployment scenarios in terms of centralizing or distributing each element with the same software implementation. The Work Group 6 "cloudification" charter is to identify use cases that will demonstrate the benefits of hardware and software decoupling of all O-RAN elements (including RIC, Open RAN Centralized Unit (O-CU), Open RAN Distributed Unit (O-DU), Open RAN Remote Unit (O-RU) and all deployment scenarios and to develop requirements and reference designs for the cloud platform including the NFVI (infrastructure), VIM (container/VM orchestration), and Accelerator Abstraction layers.

The first deliverable from WG6 introduces and examines different scenarios and use cases for O-RAN deployments of Network Functionality into Cloud Platforms and proprietary equipment. Deployment scenarios are associated with meeting customer and service requirements, while considering technological constraints and the need to create cost-effective solutions.

WG7 - White-box Hardware Workgroup

The goal of Work Group 7 is to specify and release the complete hardware reference design of a high performance, spectral and energy efficient white box base station. Within this scope, any kind of design material is included, such as documentation of reference hardware and software architectures, detailed schematic of reference designs and POC hardware, as well as test cases for verification and certification of all base station types and usage scenarios. Component selection for the implementation of example white box hardware is allowed for WG7 but is not mandatory in any specification. WG7 Draft Specifications proposed for adoption as O-RAN Alliance Final Specifications do not include mandatory requirements to use specific chipsets or components.

WG8 - Stack Reference Design Workgroup

The goal of Work Group 8 is to develop the software architecture, design, and release plan for the O-RAN Central Unit (O-CU) and O-RAN Distributed Unit (O-DU) based on O-RAN and 3GPP specifications for the NR protocol stack.

The first deliverable from WG8 introduces RAN deployment scenarios and requirements, describing RAN features and various functional blocks for O-RU, O-DU and O-CU.

WG9 – xhaul transport

WG9 is focused on the transport domain – consisting of transport equipment, physical media, and control / management protocols associated with the transport network underlying the assumed Ethernet interfaces (utilized for fronthaul, mid-haul and backhaul). The WG9 specifies deployment architectures, requirements, and solutions, identifies gaps and proposals towards existing transport SDOs. WG9 also coordinates requirements from other WGs, negotiating as necessary to align requirements among the other WGs.

The scope of the WG9 includes:

- deliver transport specifications including technical requirements, architecture, key components, management and control protocol for the various scenarios with potential solutions (e.g. PON, xWDM, DOCSIS, etc.) in support of open interfaces. Non-traditional network definitions including microwave and air-to-ground links are also considered.
- open-design specifications and multi-vendor interoperability in transport domain,
- definition of security handling on the networks related to the transport network,
- network topologies including tree and ring structures, and performance (e.g. timing) budgets for the various topologies,
- identify gaps in existing standards and drive requirements/use cases into relevant transport standards development organizations (SDOs), such as ITU Telecommunication Standardization Sector (ITU-T) or Institute of Electrical and Electronics Engineers (IEEE) etc.,
- requirements for transport optics and nodes, ensuring requirements conform to specifications already addressed by other WGs. Any new requirements for transceivers or nodes impacting other working groups, shall be proposed and treated in the affected groups to ensure alignment. A number of working groups have published output from their work.

All of the specifications are available for download at o-ran.org/specifications. The O-RAN specifications include a fair, reasonable, and non-discriminatory (FRAND) license for commercial use, as well as the ability to subsequently modify the O-RAN specifications; for example, to enable the baseline interoperable O-RAN specifications to be augmented with vendor differentiated functionality.

The three focus groups are:

- OSFG - Open Source Focus Group
- SDFG – Standard Development Focus Group
- TIFG – Test and Integration Focus Group

2.2.1.2 O-RAN Software Community

O-RAN Software Community (OSC) is a collaboration between the O-RAN Alliance and Linux Foundation with the mission to support the creation of software for the Radio Access Network (RAN). OSC uses O-RAN specifications while leveraging other LF network projects, to address the challenges in performance, scale, and 3GPP alignment.
Initial set of software projects being worked or considered include:

- near-real-time RAN intelligent controller (nRT RIC),
- non-real-time RAN intelligent controller (NRT RIC),
- cloudification and virtualization platforms, open central unit (O-CU),
- open distributed unit (O-DU), and
- test and integration effort to provide a working reference implementation.

2.2.1.3 O-RAN Testing and Integration Centers

One of the biggest goals of the O-RAN Alliance is to deploy O-RAN compliant elements and interfaces. To this end, O-RAN Testing and Integration Centers were developed by the Testing Integration Focus Group (TIFG). The purpose to setup the OTICs is to facilitate O-RAN community conformance and interoperability testing and to drive the ecosystem towards O-RAN compliant solutions.

The TIFG has been defining the OTIC Guideline and Criteria needed to guide interested hosting entities for setting up the OTICs and to ensure consistent approach to manage the relationship between O-RAN, OTIC hosts, and the participants. OTICs are being deployed globally initially across Asia, Europe and North America.

In addition to certifications, OTIC is also leading global plugfests. OTIC plugfests may vary from OTIC to OTIC but typically demonstrate key O-RAN capabilities and use cases in the areas of openness and intelligence. For example, an OTIC may demonstrate multi-vendor interoperability testing for indoor scenario; Open Front Haul decoupling; SW/HW decoupling; performance and functional evaluation of O-RU as well as FH interface conformance test by the involvement of O-DU simulator; and Proof of Concept of testing of RIC, e.g. E2 interface.

2.2.1.4 O-RAN Alliance Architecture

The O-RAN Alliance is committed to accelerating the evolution of the RAN, driving the adoption of virtualized network elements, white-box hardware and standardized, multi-vendor interoperable interfaces. Whereas internal RAN interfaces have always been defined by 3GPP, they have not seen widespread deployment in multi-vendor environments. Hence, O-RAN efforts are focused on “closing the interoperability gap” between published specifications which may have been incompletely defined, with O-RAN refining the specifications, removing ambiguity, and defining interoperable “profiles”.

While openness targets the enabling of a more competitive and responsive supplier ecosystem, O-RAN is also defining new concepts in the RAN, including how to expose access to local RAN data, and describing the functionality associated with a “Near Real-Time RAN Intelligent Controller”, as shown in Figure 21. This Near RT-RIC is aimed at enhancing network performance by enabling third party applications to interact with the closed loop control loops that are used to manage load, energy consumption and other parameters across multiple radios.

From a control and management plane perspective, O-RAN further decomposes the RAN into the Near-RT RIC, O-CU-CP, O-CU-UP, O-DU, and O-RU functions and illustrated in Figure 21. Significantly, the designation "O-" is used to differentiate the O-RAN functionality compared to 3GPP standardized functions; the combination of Near-RT RIC, O-CU-CP and O-CU-UP can be viewed as being logically equivalent to 3GPP's CU, and the combination of O-DU and O-RU can be viewed as being logically equivalent to 3GPP's DU.

The CP and UP functions are complemented by a high-level management architecture also illustrated in Figure 21. This shows the four key management interfaces being specified by O-RAN Alliance, namely:

- **A1:** Defined between the Non-RT RIC in the Service Management and Orchestration framework and the Near-Real Time RIC in the RAN
- **E2:** A logical interface defined between the near-RT RIC and an E2 Node, which for NR can be any combination of O-CU-CP, O-CU-UP, or O-DU
- **O1:** Defined between the Service Management and Orchestration framework and the O-RAN Network Functions
- **Open Fronthaul Management Plane (M-Plane):** Defined between the Service Management and Orchestration framework and the O-RAN defined O-RU
- **O2:** Defined between the Service Management and Orchestration framework and the O-Cloud for providing platform resources and virtualized workload management

Importantly, O-RAN has defined the use of Internet Engineering Task Force's (IETF's) NETCONF/YANG (Yet Another Next Generation) standard for programmatically configuring and managing its decomposed RAN architecture. YANG (RFC 7950) [1] is a modelling language that is used by O-RAN to model the configuration and operational state of its Managed Functions, together with defining remote procedure calls (RPCs) for supporting tasks like software management. Because YANG defines syntax, relationships and constraints between the data, it enables operators of O-RAN Managed Functions to validate configuration data against the model before committing the configuration to the specific function. Moreover, the ability to compile and import the YANG models directly into the NETCONF client and server functionality avoids any possible translation errors from specification to implementation, and dramatically simplifies the on-boarding of O-RAN functions into a broader OAM system.

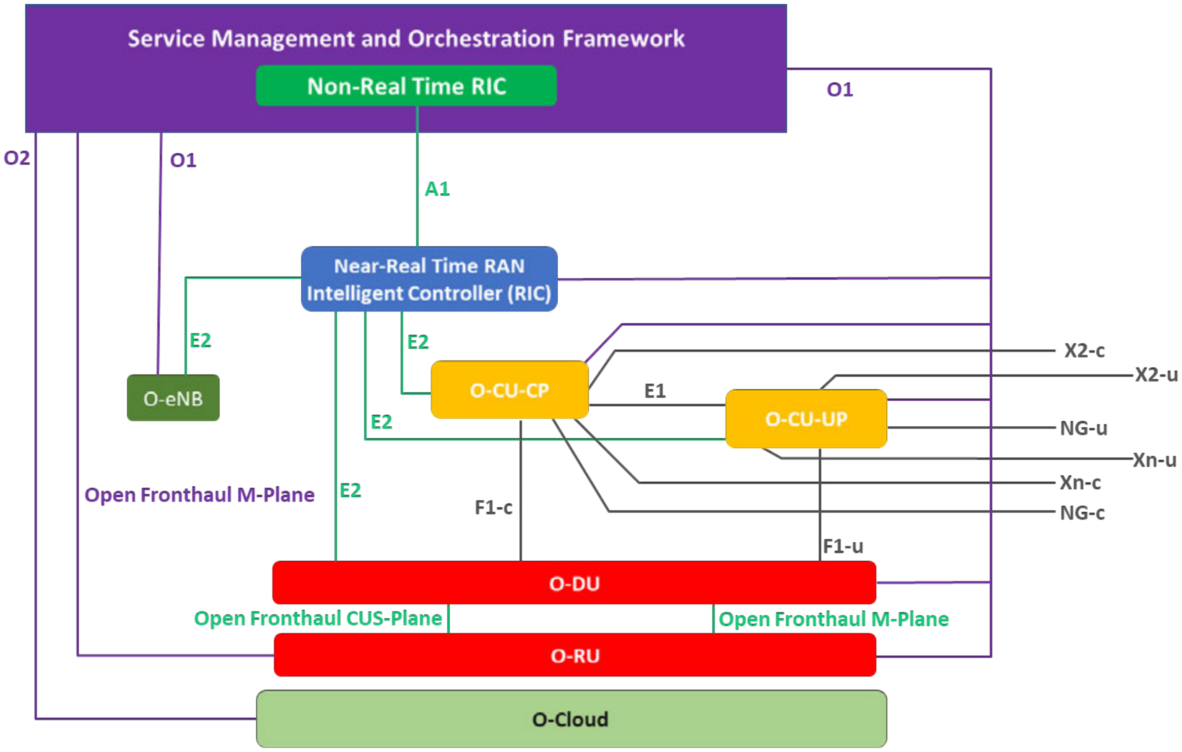


Figure 2-1: Logical O-RAN Architecture

The O-RAN architecture supports both the Higher Layer Split (HLS) being specified by 3GPP, but also a Lower Layer Split (LLS) which was studied by 3GPP in 38.801 [2], but which is not currently being specified by 3GPP.

In particular, the earlier split PHY analysis (called Split 7) highlighted different variations of the Split 7 depending on what functions are located above the split point compared to those located below, with alternatives being referred to as Split 7-1, 7-2 and 7-3. O-RAN alliance has adopted the previous conclusions of the xRAN Forum to standardize on a specific decomposition of functions between the O-RU and O-DU, termed 7-2x.

This split can also be configured to operate in two distinct modes, termed Category A and Category B (shown in Figure 22). When operating in “Category A” mode of operation, the pre-coding and resource element mapping operate in the O-DU, resulting in the fronthaul interface being used to transport different spatial streams. Conversely, when operating in “Category B” mode of operation, the pre-coding functions are moved below the split, allowing the fronthaul interface to transport MIMO layers. In such a configuration, “modulation compression” can be used in the DL to effectively send only the bits equivalent to the constellation points, resulting in the bandwidth approaching that of alternative 7-3 splits. Using such an approach, a converged fronthaul interface can be used to support a variety of use cases, such as outdoor massive MIMO.

2.2.2 Small Cell Forum

The progress of the Small Cell Forum (SCF) on Open RAN predates the work of 3GPP and O-RAN Alliance. In June 2014, the SCF’s Operator Group tasked the Forum with performing a comprehensive analysis into the role of small cell virtualization. The results of this analysis were published a year later [3], highlighting the key benefits of centralization and virtualization of the small cell RAN while broadening the split analysis compared with previous CPRI/ORI approaches.

Specifically, the analysis concluded that the MAC/PHY split shown in Figure 23 delivers most of the benefits of centralization, with only a small increase in transport performance and is well aligned with the current small cell multi-vendor ecosystem approach based on the Functional Application Platform Interface (FAPI). The Forum agreed to use the conclusions to trigger the definition of a “networked” FAPI, or nFAPI, interface for supporting the MAC/PHY split. The nFAPI specification was published in October 2016, describing how to decompose a regular small cell into a Virtual Network Function (VNF) and a Physical Network Function (PNF). From a management perspective, nFAPI looked to blur the lines between small cell and established DAS management models, enabling partitioning of shared PNF resources between multiple service providers. Being based on the same TR-069 framework used for small cells, the Forum has defined a management object specifically for managing the PNF.

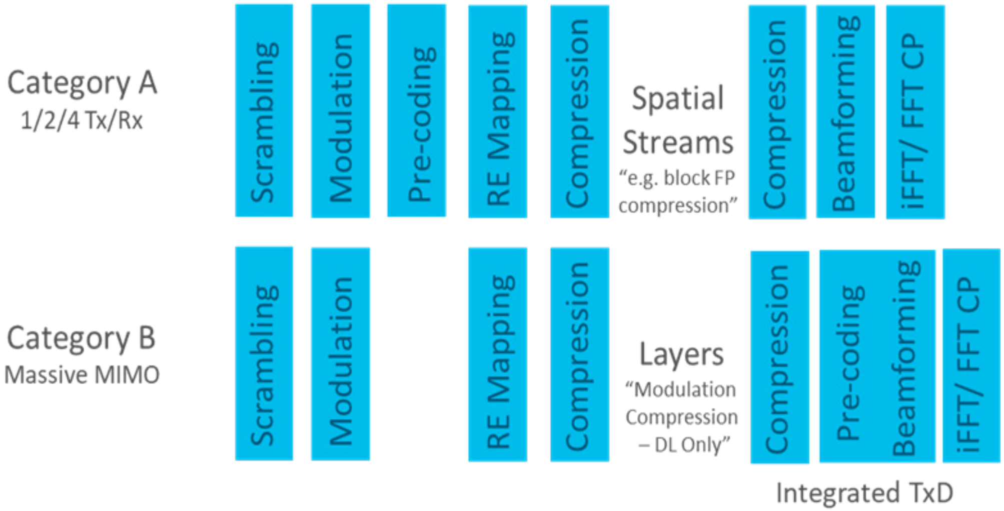


Figure 2-2 O-RAN Split 7-2x modes of operation

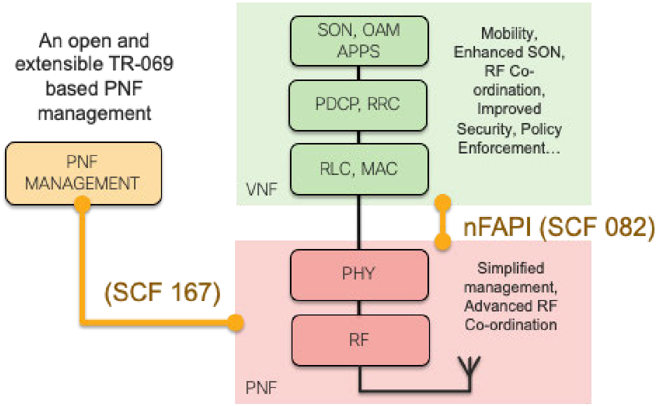


Figure 2-3 Small Cell Forum MAC/PHY split

The split MAC/PHY nFAPI interface has been integrated into the Open Air Interface's (OAI) code [4]. The OAI software alliance is a non-profit consortium fostering a community of industrial as well as academic contributors for open source software and hardware development for 3GPP based core network, access network and user equipment. Specifically, OAI overhauled their MAC-PHY interface, leveraging the nFAPI structures defined in Cisco's open-nFAPI repository for supporting split option 6 [5].

2.2.3 Telecom Infra Project (TIP)

Telecom Infra Project (TIP) was formed in 2016 as an organization that is focused on collaboration and the development of new technologies for building and deploying global telecom network infrastructure to enable access for everyone in the world.

There are over 500 members which include operators, suppliers, developers, integrators, and other entities. The TIP board of directors is composed of individuals from the founding tech and telecom companies. Member companies host TIP community labs, and TIP hosts an annual TIP Summit.

Within TIP, there are project groups working on different network concepts. Below is a list of project groups dedicated to the area of Open RAN platforms:

1. OpenRAN 5G NR

The goal of the OpenRAN 5G NR Project Group is to collaboratively design an open interfaced, multi-vendor interoperable, disaggregated whitebox platform for a 5G NR access point that is easy to configure, scale and deploy. The solution includes a 5G NR compatible baseband unit; antenna and radio, and the provisioning elements.
The focus of this TIP project group is on use cases for outdoor macrocells and small cells as well as indoor small cells. For the outdoor case, technical specifications have been published, with lab testing having been planned as a next step. For the indoor case, a sub-group was kicked off in April 2020 and will go through the process of defining the requirement specifications, lab testing, and other milestones.

2. TIP OpenCellular

The mission of this project group is to connect the unconnected. The goal is to empower communities in underserved areas with tools to build and operate sustainable cellular infrastructure using open-source technologies and an open ecosystem. The aims are to achieve its mission by providing an open-source platform to build, deploy, and operate complete (E2E) cellular networks.

The OpenCellular platform has been deployed by multiple service providers in various African countries and by community networks in Latin America and Africa. Africa Mobile Networks, as part of MTN and Orange, have installed sites covering over 300,000 people in sub-Saharan African countries. The OpenCellular platform will be deployed in Argentina, Rwanda, Tanzania and Uganda as well.

TIP has open-sourced the OpenCellular files including schematic, layout, CAD files and software as part of the project group to help accelerate creation of an open ecosystem where new ideas can be tested.

3. PlugFest

The TIP PlugFest group was launched in 2019. The mission of the project group is to define and accelerate the development of test materials, test plans and other documents that will support TIP-sponsored PlugFest. There are two types of PlugFests. One is focused on proofs of concepts, and the other is focused on product maturity which takes longer than the proof of concept PlugFest.

4. OpenRAN

This project group's main objective is the development of fully programmable RAN solutions based on General Purpose Processing Platforms (GPPP) and disaggregated software. The RAN solutions can benefit from the flexibility and faster pace of innovation capable with software-driven development.

To achieve this, the project will help enable an open ecosystem of complete solutions and solution components that take advantage of the latest capabilities of GPPPs, both at a software level and also using programmable offload mechanisms such as Field-Programmable Gate Arrays (FPGA).

The project will complement existing TIP projects and will focus on disaggregation of virtualized RAN solutions into different components and ensuring each individual component can be efficiently deployed on GPP platforms.

Project group deliverables include:

- reference framework/architecture for implementation of the eNB stack on GPPPs
- reference (and optimized) implementation of the basic building blocks and algorithms, both as software libraries and FPGA register-transfer levels (RTL)
- hardware abstraction Layer, including APIs, to abstract from application vendors the underlying hardware platform capabilities
- defined KPIs and traffic model as part of the reference implementation
- orchestration framework to manage and provide operational capabilities
- carrier-grade lab proof-of-concept evaluation of multi-vendor open solutions

'A PlugFest focused on 4G/5G Open RAN, where designers of electronic equipment or software test the interoperability of their products or designs with those of other manufacturers, was held in September 2020.

Note TIP has recently streamlined the OpenRAN and OpenRAN 5G NR Project Groups into an overarching OpenRAN Project Group.

5. TIP System Integration and Site Optimization

This project group addresses system integration via innovative, cost-effective and efficient end-to-end solutions in order to serve both rural and urban regions in optimal and profitable ways.

It focuses on cost analysis, cost-effective site engineering (site selection and setup), connectivity systems (wireless backhaul, satellite link and efficient antenna technologies), automated maintenance and optimization, system integration and business/revenue model (network infrastructure sharing, revenue-sharing model).

6. TIP Radio Intelligence and Automation (RIA) Workstream

The TIP RIA workstream is a sub-group of the TIP OpenRAN 5G NR Project Group. The goal of this workstream is to provide a platform to develop, test and deploy OpenRAN 5G use-cases that leverage Data Science and AI/ML technologies. Work on the RIA workstream began in April 2020 and was approved in June 2020. See https://telecominfraproject.com/openran/.

2.2.4 3GPP

Open RAN requires open interfaces between the elements of a disaggregated RAN: namely the Central Unit (CU), Distributed Unit (DU) and Remote Radio Unit (RRU).

The 3GPP RAN3 Working Group opened a study item for Release 15 on the "Study on CU-DU lower layer split for New Radio" (RP-180684) [6].

C-RAN, in which the RAN architecture is split into centralized baseband units and distributed radio units, has gained traction, and has proven to be effective in commercial 3G and LTE deployments over previous years. Such centralized architecture has both performance benefits (due to improved inter-cell/frequency coordination at the centralized baseband) and cost benefits (e.g., due to increased hardware/software pooling, reduced site rental and management costs).

With the challenging and diverse requirements for 5G NR systems, the need for split RAN architecture is becoming ever more important. There is high demand from operators to realize multi-vendor interoperability within such split RAN architecture. This was taken into account in the Scenarios and Requirements TR for Next Generation Access Technologies (TR 38.913 [7]) and also in the Study Item Description for NR Access Technology (RP-162469 [8]) as follows:

Excerpt from TR 38.913 [7]

The RAN design for the Next Generation Radio Access Technologies shall be designed to fulfil the following requirements:

- Different options and flexibility for splitting the RAN architecture shall be allowed.
- RAN-CN interfaces and RAN internal interfaces (both between new RAT logical nodes/functions and between new RAT and LTE logical nodes/functions) shall be open for multi-vendor interoperability.

Excerpt from RP-162469 [8]

Detailed objectives of the study item are:

(3) Initial work of the study item should allocate high priority on gaining a common understanding on what is required in terms of radio protocol structure and architecture to fulfil objective 1 and 2, with focus on progressing in the following areas

- *Radio interface protocol architecture and procedures*
- *Radio Access Network architecture, interface protocols and procedures,*

Study on the above 2 bullets shall at least cover:

- *Study the feasibility of different options of splitting the architecture into a "central unit" and a "distributed unit", with potential interface in between, including transport, configuration and other required functional interactions between these nodes [RAN2, RAN3];*
- *Study the alternative solutions with regard to signaling, orchestration, ..., and OAM, where applicable [in co-operation with SA5];*

Accordingly, the study on splitting the RAN architecture into CU (Centralized Unit) and DU (Distributed Unit) was conducted in RAN3 within the NR Access Technology Study Item, where the status is captured in TR 38.801 [2], in which a general need for higher layer splits and lower layer splits were identified as follows:

Excerpt from TR 38.801 [2]

There are transport networks with performances that vary from high transport latency to low transport latency in the real deployment. 3GPP specification should try to cater for these types of transport networks. For transport network with higher transport latency, higher layer splits may be applicable. For transport network with lower transport latency, lower layer splits can also be applicable and preferable to realize enhanced performance (e.g., centralized scheduling).

For the CU-DU higher layer split, the RAN3 study concluded that "*There shall be normative work for a single higher layer split option, i.e. Stage 2 and Stage 3*" (TR 38.801 [2]). The resulting normative work on CU-DU higher layer split will be carried out within the Work Item on New Radio (NR) Access Technology [9].

However, for the CU-DU lower layer split, RAN3 study concluded that "*Further study is required to assess on lower layer splits, their feasibility, the selection of options, and assess the relative technical benefits based on NR, before a decision to go to specification phase can be made. Discussions in the Study Item, favored option 6 and 7 for future study*" (TR 38.801 [2]).

2.2.5 Cisco Multi-Vendor Open vRAN

This Cisco led initiative formed in 2018 is intended to facilitate software-defined networking (SDN) and virtual RAN (vRAN) interoperability. Cisco and the other ecosystem parties decided to focus on assembling viable solutions that build on an open and modular architecture, draw from existing industry efforts, and support a variety of use cases.

The mission for this initiative includes:

- Testing and integration
- Solutions validation
- Transport evolution
- Publishing performance benchmarks
- Running Proof of Concepts (PoC)
- Coordinating roadmaps for end-to-end (E2E) solutions
- Creating and validating network management templates, and more

2.2.6 Open RAN Policy Coalition

In May 2020, the Open RAN Policy Coalition was formed to "promote policies that will advance the adoption of open and interoperable solutions in the RAN and expand the supply chain for advanced wireless technologies". As the name suggests, this group would focus on policy initiatives to complement the work of the more technology-centric bodies such as O-RAN Alliance, TIP etc. in furthering the adoption of Open RAN.

2.2.7 US Government and Open RAN:

Legislation

In January 2020, bill S.3189 was introduced in the US senate. This bill provides financial incentives specifically to develop O-RAN technology with O-RAN defined as the "Open Radio Access Network approach to standardization adopted by the O-RAN Alliance, Telecom Infra Project, or 3GPP, or any similar set of open standards for multi-vendor equipment interoperability." [10]

In April 2020, bill H.R.6624 (USA Telecommunications Act) was introduced in the US House of Representatives. The enactment of this act will "make grants on a competitive basis to support the deployment and use of Open RAN 5G Networks throughout the United States". [11]

DARPA's Open, Programmable, Secure 5G (OPS-5G)

In January 2020, the US military's Defense Advanced Research Projects Agency (DARPA) introduced a program to solicit innovative research proposals called "Open Programmable Secure 5G" (OPS-5G). [12] The goal of the program is to create open source software and systems enabling secure 5G and subsequent mobile networks such as 6G. As stated by DARPA, "*the signature security advantage of open source software is increased code visibility, i.e., the code can be examined, analyzed, and audited, either manually or with automated tools. In addition, the portability of open source serves, as a desired side-effect, to decouple the hardware and software ecosystems. This significantly raises the difficulty of a supply-chain attack and eases the introduction of innovative hardware into the market.*" The program is expected to begin in October 2020 and end four years later.

The program will build on efforts by some U.S. telecom and technology companies to agree on common engineering standards that would allow 5G software developers to run code on machines that come from nearly any hardware manufacturer. In short, the main principal of the OPS-5G program is to initiate innovations for 5G-based open source security architectures. Key strategic outcomes of the OPS-5G program are:

1. spurring new software development and accelerating open source software deployment,
2. establishing new "zero trust" security architectures,
3. mitigating the inherent risks of shared physical hardware such as network slicing and,
4. using NFV and SDN programmability as a defense mechanism.

The contract pre-solicitation process began on January 30, 2020. [13]

2.3 Architectural Considerations

Wireless network architecture evolution has been driven by services, from initial circuit voice and circuit data, to packet data, and then to a single packet data domain to support voice over packet data such as Voice over LTE (VoLTE) or Voice over New Radio (VoNR). From a protocol architecture perspective, the circuit voice and data are tied up with the lower layer implementation. As the network evolved to a single packet data domain, services like voice are moved up to the application layer.

Aside from voice, many other services come up, especially in the vertical industries like IoT etc. The existing 4G/5G network architecture may not be flexible enough to provide the network architecture that can meet the needs of all potential applications. To this end, the RAN architecture needs to be flexible so that it can be configured dynamically on the fly or released when it is not needed. This requires the RAN to be fully software programmable and its functions to be decomposed to the granular level so that they can be flexibly and dynamically allocated. This kind of flexibility can be achieved through virtualization and network slicing in cellular networks, just as we see in the IT world where virtualization, microservices and cloudification have enabled flexibility and scalability for applications.

2.3.1 RAN Function Splits

Traditional Distributed and Centralized RAN systems split RAN functionality between the baseband and radio with a fronthaul interface that requires significant bandwidth and very low latency. To enable deployment flexibility, several RAN split options were proposed in 3GPP, shown in Figure 24.

The RAN function split allows RAN functionality to be divided into centralized and distributed locations with a varying set of capabilities and requirements between the CU, DU and RU. There are different options for the RAN split which go from the high layer RAN split to the lower layer RAN split. In the high layer split, fewer RAN functions are centralized while in the low layer split more RAN functions are centralized. There are trade-offs in terms of complexity, flexibility, transport and latency requirements as well as the overall costs associated with each split. Options 2, 6, and 7 gained traction with the operators and vendor ecosystems for development. For more details on Function splits and differences between traditional fronthaul (CPRI) and Ethernet based fronthaul (eCPRI), see the Appendix.

2.3.2 Disaggregation and Interoperability

The O-RAN Alliance and 3GPP are addressing an industry need for adopting a truly interoperable RAN platform, where multiple vendors can participate in the ecosystem and avoid the over-reliance of operators on a single vendor. In addition, there is a need for providing a disaggregated RAN solution where the baseband can be split into multiple components and these components can

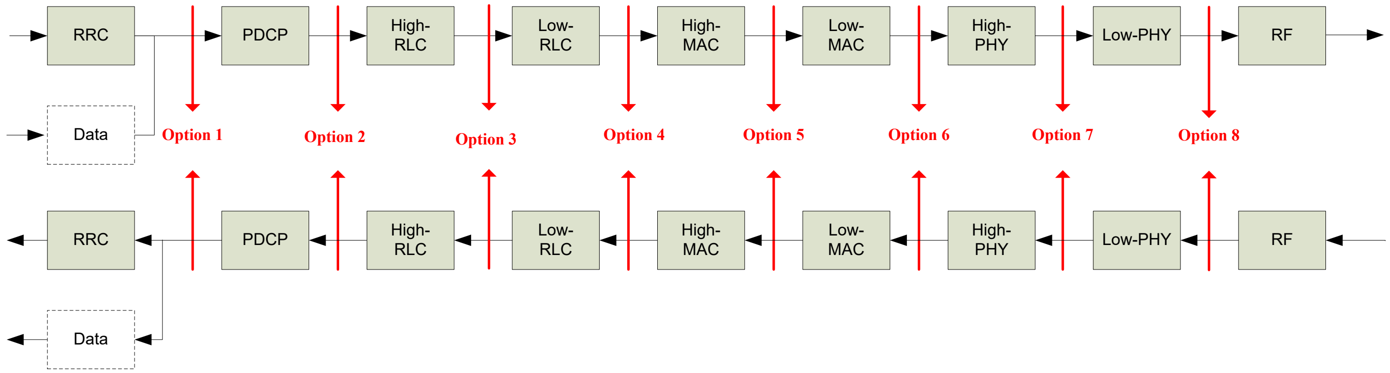


Figure 2-4 3GPP Release 14 study on RAN split with E-UTRA protocol stack [2] inter-operate with those from other RAN vendors. This allows optimization for deployments such as Mobile Edge Computing (MEC) and URLLC applications, where the Centralized Unit-User Plane (CU-UP) may be at the edge while the Centralized Unit-Control Plane (CU-CP) may be centralized, each provided by a different vendor.

2.3.3 RAN Virtualization

Virtualization entails migration from custom-built network nodes to network functionality implemented in software running on a generic hardware compute platform. Virtualization for communications service providers began with the core network, and subsequently cloud technologies, and have been evolving at a rapid rate. In the RAN domain, vendor agnostic commercial off-the-shelf hardware has the potential to enable innovation across a range of software ecosystems.

3GPP Split Architecture and Virtualization Journey

A separation of the upper and lower parts of the RAN was standardized in 3GPP Release 15 (Rel-15), where a higher-layer split was specified with a well-defined interface (F1) between two logical units: The Centralized Unit (CU) and the Distributed Unit (DU) shows the CU and DU functions and interfaces.

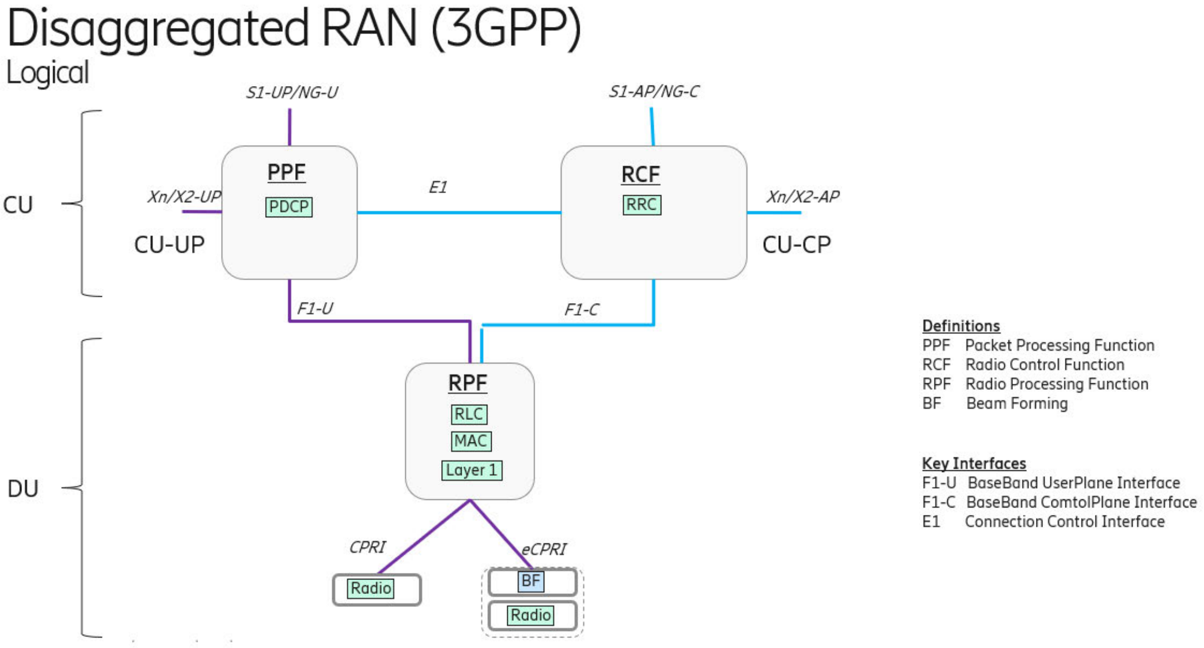


Figure 2-5 3GPP Rel-15 Disaggregated RAN

The Packet Data Convergence Protocol (PDCP) and Radio Resource Control (RRC) layers are the higher layers in the RAN stack and are responsible for the packet processing function and radio control function in the gNodeB (gNB). These higher layer functions are separated from the DU to form the CU functions - namely, Centralized Unit-User Plane (CU-UP) and Centralized Unit- Control Plane (CU-CP) functions.

The 3GPP DU function includes the radio (including beamforming function for M-MIMO) and radio processing functions that consist of physical layer (L1) functions, and higher layer MAC and radio link control functions.

CU-CP and CU-UP are connected over the E1 interface and CU and DU are connected over F1 interface.

The Radio Processing Function (RPF) part of the DU is close to the radio and has strict requirements on latency which creates a tough virtualization challenge. The schedulers work at 1 ms and sub-1ms Transmission Time Interval (TTI) budgets. The CU on the other hand, with its higher latency budget and less stringent processing requirements, was an attractive initial target for virtualization. Centralizing control and user planes allowed pooling, scaling and load-balancing benefits to be realized in the virtualized environment – the first step in realizing cloud scale. Thus, Higher Layer Split (HLS) allowed the virtualization journey to begin in the RAN domain with CU virtualization as the first step.

For full-stack RAN virtualization, the DU is connected to the radio via a packet interface known as enhanced Common Public Radio Interface (eCPRI). There are multiple ways to divide functions between the DU and the radio; in standards discussions these are referred to as "lower-layer split" (LLS) options. One possible alternative specified by the O-RAN Alliance is referred to as the 7-2x split; other functional splits are also being considered.

In this section, we focus on the RPF part of the DU.

2.3.3.1 CU and DU – Virtualization

RAN virtualization involves CU and DU virtualization. In either case, the key decision points are to do with the selection of the right Commercial Off-the-shelf (COTS) server hardware, the right virtualization approach and Cloud OS; and in the case of compute-heavy scenarios the right hardware acceleration approach.

Hardware acceleration approaches are relevant in:

- Acceleration of traffic in input / output path (e.g. virtual Centralized Unit User Plane (vCU-UP))
- Acceleration of individual functions in the L1 pipeline (for a virtual Distributed Unit (vDU))

Finally, the overall system integration, management, orchestration and assurance are significant considerations in the virtualization journey.

To enable scalable service management and orchestration across 5G RAN, open programmability of RAN is an important consideration for virtualized as well as embedded platforms. Open programmable interfaces provide a way to manage different platforms and Virtual Network Function (VNF) workloads in a consistent way. It must be noted that the terms "open interfaces"

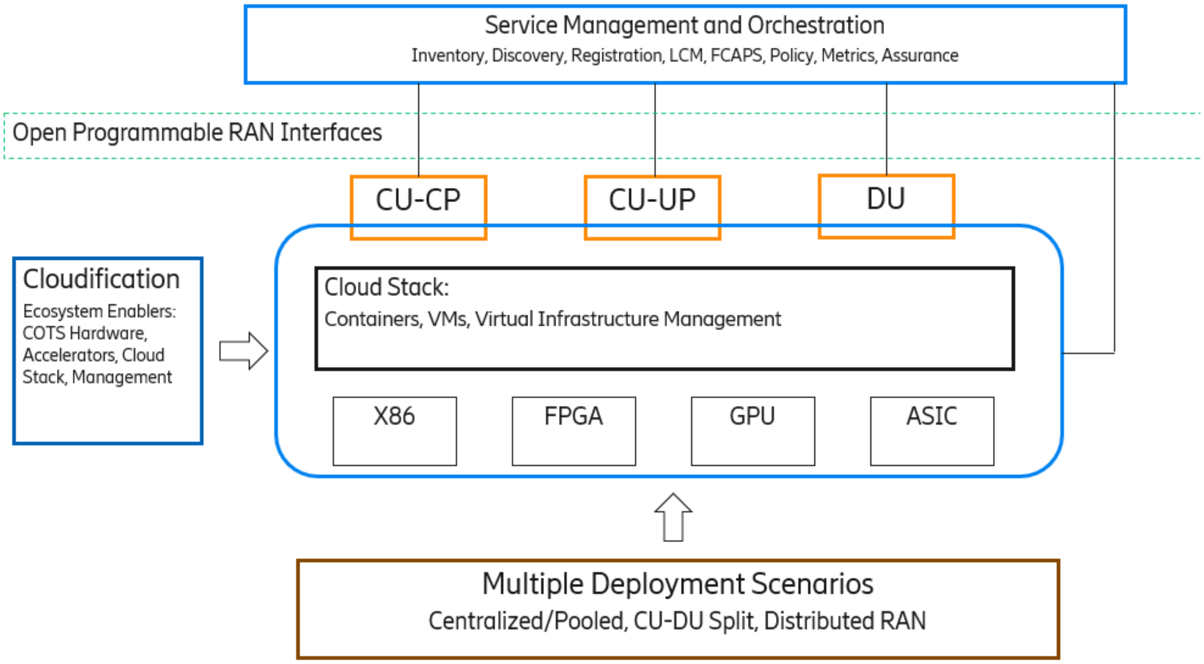


Figure 2-6 CU, DU Virtualization and Open RAN Interfaces

and "virtualization" are not interchangeable. Virtualization involves hardware software disaggregation and programmability while open interfaces involve interface programmability from the Service Management and Orchestration layer, shown in Figure 26. Virtualization can also be achieved with closed interfaces, and embedded platforms can also support open interface programmability

Virtualization of CU includes virtualizing the CU-Control Plane (CU-CP) and CU-User Plane (CU-UP). CU-CP and CU-UP can be virtualized on a COTS server. CU-UP is more demanding than CU-CP in terms of capacity and I/O throughput. A key part of the compute requirement for CU-UP comes from the fact that it handles the high throughput user plane traffic and does flow control over the baseband user-plane interface (F1-U) interface. Depending on the server capabilities and workload demands, acceleration of traffic in the Input / Output path may be required for CU-UP workloads.

The choice of virtualization environment/Cloud OS for CU and DU is an important strategic decision. Most of the virtualization efforts started with a hypervisor environment, allowing VNFs to be realized as Virtual Machines. However, in order to facilitate web scaling, VNFs need to be cloud-native. Thus, even though initial virtualization efforts are in VM environment, the trajectory is towards a microservices-based, Kubernetes-orchestrated container environment.

Virtualizing CU and DU starts with the selection of the hardware platform and the virtualization environment or Cloud OS. The hardware platform is in general a COTS server (e.g., Intel X.86 based server platform) - with NIC and hardware accelerators, where needed.

Hardware acceleration approaches are relevant in the acceleration of traffic in the Input/Output path and in the acceleration of individual functions in the L1 pipeline.

In general, hardware acceleration is required in two areas:

1. **I/O acceleration:** This entails acceleration of Transmit (Tx) and Receive (Rx) data transfers from and towards the Fronthaul interface, Baseband control-plane interface (F1-C) in the case of vCU-CP; and from and towards F1-U in the case of vCU-UP. For I/O acceleration, single-root I/O virtualization (SR-IOV) Network Interface Card (NIC)s and/or software mechanisms such as Data Plane Development Kit (DPDK) are used.
2. **Algorithm acceleration:** This entails acceleration of any specific function or a set of functions within a managed entity. DU virtualization, in highly demanding compute-intensive scenarios, will need algorithm acceleration. The accelerator hardware typically used are FPGA, GPU, and ASIC.

Algorithm acceleration is a key part of L1 acceleration and is especially required in high load/ high compute scenarios, such as in Advanced Antenna Systems (AAS) Massive Multi-In Multi-Out (M-MIMO). There are two approaches in L1 acceleration, namely, the look-aside acceleration approach and inline acceleration approach as seen in Figure 2-7.

Considering the downlink (DL) case, look-aside acceleration approach supports dataflow from the CPU to the accelerator and back to the CPU before being sent to the front-haul interface. Inline acceleration supports data flow from the CPU to the accelerator and directly from the accelerator to the front-haul interface, instead of being sent back to the CPU.

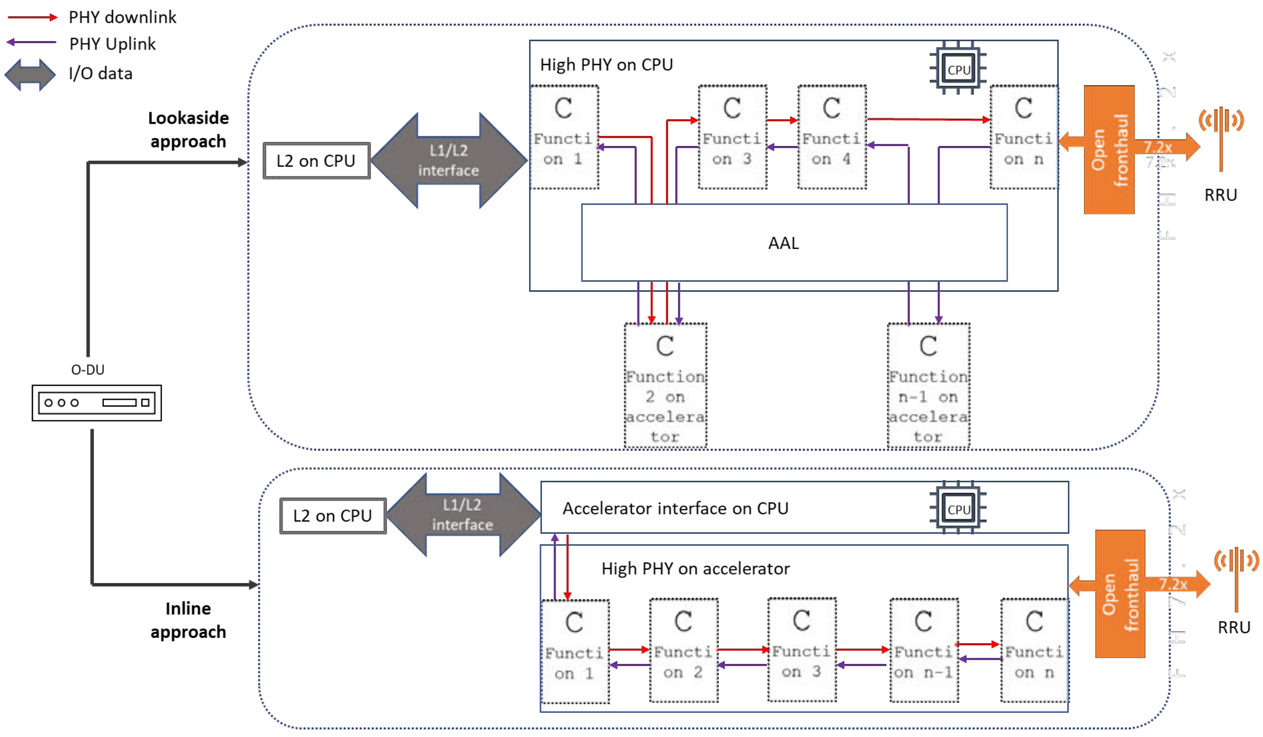


Figure 2-7 Look-aside vs Inline Acceleration. Source: Ericsson

With the look-aside approach, selective functions are accelerated. For instance, in the case of Intel FlexRAN, L1 functions such as Low Density Parity Check (LDPC), rate match/de-match and Cyclic Redundancy Check (CRC) are accelerated. This requires data being moved between CPU and the accelerator (FPGA/eASIC). While the accelerator is working on the data, the CPU is free to use its cycles to process other useful tasks and when the CPU receives it back from the accelerator, it can switch back to the original processing context and continue the pipeline execution till the next function to be accelerated comes up.

In contrast, in the case of Inline approach, a part of or the entire L1 pipeline can be offloaded to the accelerator.

Both approaches can be applied depending on the system vendor implementation and operator cloud infrastructure for specific deployment scenarios, as appropriate.

Regardless of how virtualization is achieved, these systems will necessitate new operational and business models with clearly defined accountability. System integration, whether managed by a RAN software supplier or a cloud infrastructure provider, will be crucial to ensure network performance and reliability. In the disaggregated case where cloud infrastructure is procured separately from RAN application, the system integration complexity, life cycle management and associated costs are factors that cannot be overlooked.

Addressing 5G deployment use cases

The evolution to 5G networks is the result of the continuous improvement of telecommunications technologies by the 3GPP partnership. Each new release has brought enhanced capabilities including supporting more spectrum and additional frequency bands, as well as air interface enhancements in performance and efficiency. One consequence of this evolution is that the processing requirement for the network functions also increases.

With today's radio access technology, RAN workloads can run on a general-purpose computing architecture based on processors such as x86 central processing units (CPUs)—but for full 5G capabilities, more processing power will be needed. The combined CPU and acceleration platform should have the potential to handle even the most demanding 5G use cases. In order to realize the benefits of virtualized RAN, the platform must be open and based on COTS hardware that is fully adopted by the cloud ecosystem with a large developer community supporting it.

Many 5G use-cases run well on pure CPU platforms, but as bandwidths increase and advanced antenna systems are deployed, current x86 cores struggle to keep up and start driving impractical levels of power consumption. Hardware acceleration will be needed for the compute-heavy functions in 5G NR.

In the case of RAN Virtualization, there is an expectation that an operator will be able to run the same 5G software stack on a variety of servers and evolve capacity by swapping out the hardware, as we do with our PCs. The compute capability required for RAN algorithms is an important factor. In COTS based virtualization environments, accelerators become key to realizing use cases with high capacity and stringent latency demands. There is likely a cost to this, not only in lower degrees of system integration, but also in lower performance per watt of power. On the other hand, there is also a possible upside in improving the pace of innovation and the adaptability of 5G to emerging use-cases.

2.3.3.2 RAN Virtualization- Benefits and Challenges

RAN virtualization presents several significant challenges as the processing and timing requirements are very high to implement certain lower layer functions. These functions are critical as many aspects of RAN capacity and coverage are determined by them. Still, the potential benefits that virtualization could bring may very well be significant.

Benefits:

First, a fully virtualized RAN could bring significant benefits of harmonization: one single uniform hardware platform across the core network, RAN and edge. This could simplify the management of the complete network, reducing operations and maintenance costs.

Second, in a full vRAN, the network functions will be separated from the processing hardware. This means that RAN network functions from multiple vendors could run on the same hardware, increasing the flexibility for the service provider. In some cases, the hardware could even be shared between service providers.

Third, vRAN offers an opportunity to embrace established solutions, available in today's public cloud technologies, for non-RAN-specific functions. By agreeing to use industry-established components for common tasks, the need for costly adaptations of vendor-specific solutions would be removed. If this is achieved, it would allow the RAN ecosystem to focus on business-critical components.

Fourth, a vRAN holds the promise of increased flexibility as functionality and capacity could be more easily deployed where and when required. Cloud technologies could facilitate this type of flexibility.

Finally, a widely adopted open platform could also lower barriers for cross-domain innovation, facilitating the development of new use cases and services.

Challenges:

Managing distributed vRAN workloads between far edge, edge/ regional and hyperscale data center hubs is a resource as well as a service orchestration challenge. Workloads may be required to span different cloud environments as demanded by KPI requirements on capacity, scale, resiliency, latency etc. – this applies to both DU and CU virtualization deployment scenarios.

In the case of brownfield operators, they need to manage multi-generation/multi-RAT technologies. This involves the consideration for coexistence of multiple radio access technologies in the network and also for service management across multiple (embedded vs. disaggregated) hardware and software platform infrastructures.

Operational and Management (OAM): This is likely the most significant challenge when RAN is disaggregated between hardware and software, and decomposed into multiple units from different vendors. Consistent management view, regarding event correlation between layers and actions, is important from Life Cycle Management (LCM) and Fault-management, Accounting, Performance and Security (FCAPS) perspective. There are efforts to address this in O-RAN Alliance with new interfaces "O1" managing RAN network functions and "O2" managing O-Cloud / infrastructure. However, this is a key challenge for virtualization scenarios to align multiple platforms with application management, and to manage the services End to End in a consistent fashion.

Deployment Architecture: Typically, centralized deployments could benefit from virtualization by enabling centralized resource pooling over a Common Off the Shelf (COTS) infrastructure and a harmonized cloud platform at the hub location. However, radio networks are very distributed because the antennas and radio unit (RU) need to be near the subscribers to achieve the coverage, strict latency and high performance needed. Virtualizing the real-time components of RUs is very challenging due to the customization needed for purpose-built hardware typically used for the RU. Redundancy in centralized/ pooled deployments must be carefully planned to pre allocate resources. Large failure domains in the case of centralized/ pooled deployments will require active redundancy mechanisms

- **Cloud scaling:** Scaling across servers, especially in the case of DU virtualization, is a real-time challenge. Use of accelerators could exacerbate this problem by introducing additional scaling constraints: A fully virtualized RAN manages several network-state based services, such as a resilient database or session management. It requires resiliency beyond the inherent features of a cloud-native architecture that is built primarily for "stateless" services orchestrations. Operational goals such as high network availability and reliability expected from a telecommunications operator cannot be solved by container orchestration alone and will require state synchronization and data integrity considerations built into the applications themselves. Additionally, protocol services require specific failover and availability mechanisms defined at the protocol level. Operators looking to adopt the cloud-native Open RAN will expect these challenges to be addressed upfront before wider adoption.
- **Security and Trust considerations:** The decentralization and virtualization of many areas of the 5G network will create new trust layers, domains, and functional or exposed weak spots. Virtual Network Function (VNF) security will have to scale both horizontally and vertically to provide adequate security and performance to other VNFs to cope with the velocity and variety of intensive 5G traffic. Secure lifecycle management of RAN VNFs is a key challenge to be addressed.

  Additionally, conducting the trust management amongst NFV hardware and software vendors is challenging. In particular, the maintainability of the trust chain can be problematic.

  Finally, in the case of Container Network Functions (CNFs), although containers provide the convenience of micro-services creation and separation, that does not ensure the creation of security boundaries since they have loose access to kernel resources, rendering them vulnerable to tampering with the container's execution path.

  In summary, virtualization introduces challenges in security and trust areas that must be carefully planned for 5G deployment scenarios and use cases. For a more detailed discussion on this topic, please refer to 5G Americas white paper on "The Evolution of Security in 5G". [14]
- **Timing and Synchronization:** Precision Time Protocol (PTP) accuracy is hardware dependent
- **Site considerations:** Virtualized deployments at the cell site have site-specific constraints in hardware dimensioning, power, and environmental conditions. This means satisfying requirements related to COTS footprint efficiency within the cabinets, power efficiency, HVAC (Heating, Ventilation and Air Conditioning) and Network Equipment Building System (NEBS) compliance as required.

- **Fronthaul:** The Front Haul in centralized architectures will require several tens of gigabits. The bandwidth and stringent latency requirements of Fronthaul can be challenging. The existing transport networks built for backhaul are not dimensioned to handle such high capacity loads and will require major upgrades. In the Americas, for centralized vRANs to succeed, there is a key need for significant growth in fiber deployments to cater to the high fiber bandwidth requirement for fronthaul and midhaul.

2.4 Operator Trials and Deployments

Interest in Open RAN deployments has been steadily growing over the last couple of years and operators around the world have started Open RAN trials and deployments in some capacity. Figure 28 illustrates some of the more notable publicly announced milestones with deployment of various forms of vRAN and Open RAN combinations.

It should be noted that as the standards bodies and the alliances have formed and shaped the technology roadmaps, there is a wide variation in the implementation of the deployments seen to date. Some of the early cases, such as the Rakuten deployment of a 4G vRAN network in Japan, pre-dates the O-RAN Alliance driven specifications, while still embracing several of the underlying principles such as software – hardware dis-aggregation, moving towards multi-vendor RAN and moving towards a more cloud-native application environment.

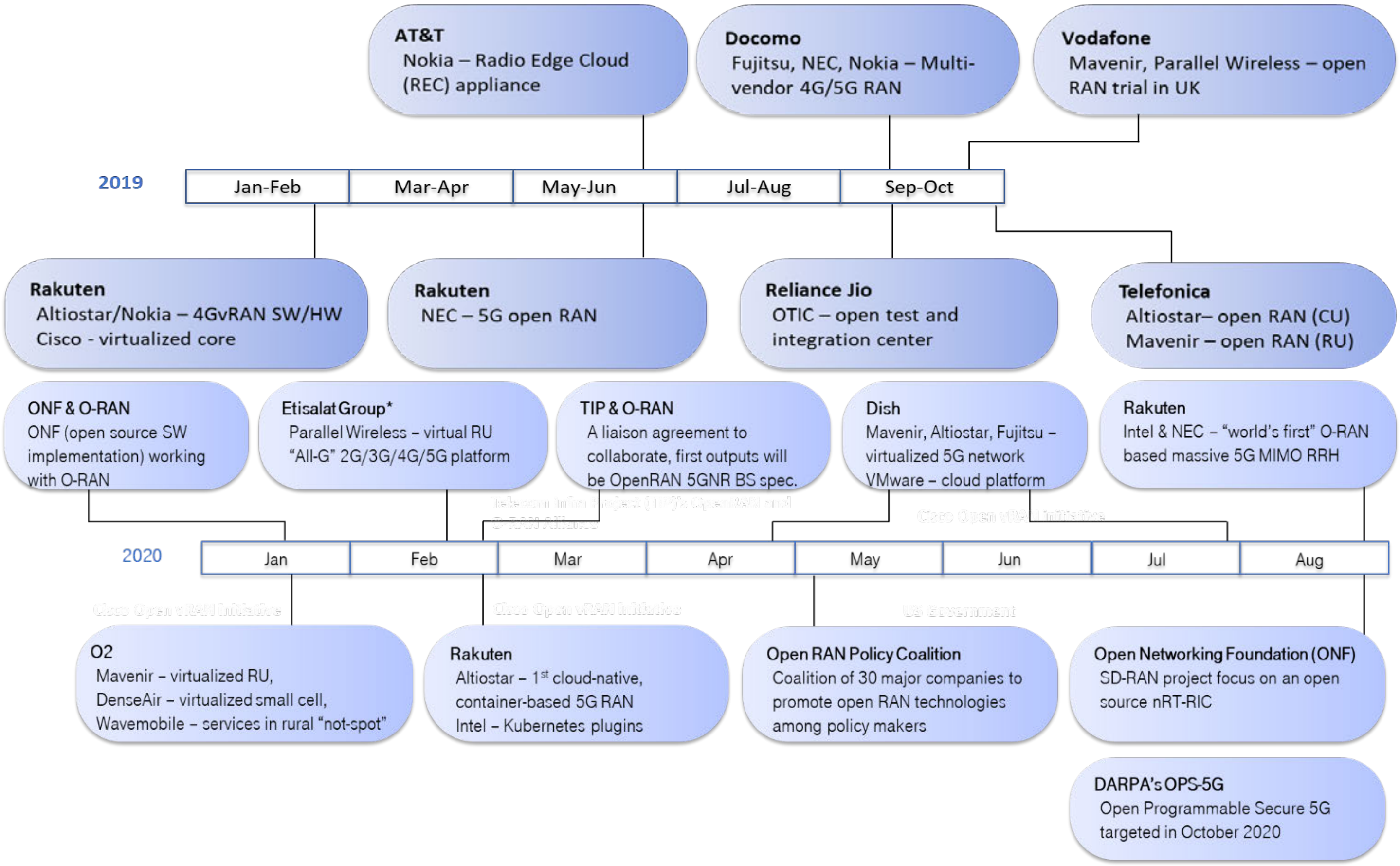


Figure 2-8 illustrating more notable publicly announced milestones with deployment of various forms of vRAN and Open RAN combinations

Below are brief descriptions of some of the notable deployments:

Rakuten Mobile:

Rakuten has built the world's first fully virtualized, end-to-end cloud native mobile network. The innovative network is fully virtualized from radio access network to core and adopts 5G system architecture. Rakuten nurtures an open ecosystem through engaging with industry leaders in crafting solutions. Rakuten Mobile is using equipment, software and services from Intel, Cisco, Nokia, Qualcomm, Altiostar, NEC, Mavenir, and Airspan. The network is also cloud native using COTS servers. Thousands of sites have been built and the LTE network has been commercially launched. 5G will be added in the future. Rakuten will have at least 8,600 base stations deployed within the next year (2021). In addition to the deployments in Japan, Rakuten is also conducting an Open RAN trial in Singapore.

Vodafone:

On October 7, 2019, Vodafone announced that it will trial Open RAN for the first time in the United Kingdom. Before this, Vodafone performed trials in Democratic Republic of Congo and Mozambique, building on its experience in South Africa and Turkey.

Lab trials were with Vodacom South Africa, part of Vodafone group. In Turkey, Open RAN was deployed to deliver 2G and 4G services to customers in both urban and rural parts of the country. Trials in three countries will provide 2G, 3G and 4G services, with 5G possible over Open RAN in the future.

Vendors that supply equipment to Vodafone are Parallel Wireless, Mavenir and UK based Lime Microsystem for Open CrowdCell.

Telefonica:

On March 18, 2020, Telefonica announced that it will deploy Open RAN trials for 4G LTE and 5G in UK, Germany, Spain and Brazil. Telefonica is embracing the O-RAN alliance open interface standard and has reached agreements with Altiostar, Gigatera Communications, Intel, Supermicro and Xilinx to develop and deploy Open RAN trials in its network. Telefonica has built a network under the name Internet para Todos in Peru which covers around 800,000 people and 650 sites. About half of these sites are Open RAN sites using Parallel Wireless products. According to David Del Val Latorre, Telefonica's CEO of research and development, the cost of Open RAN electronic gear was half as much of traditional ones.

Dish:

In February 2020, Dish Chairman Charlie Ergen revealed that the company plans to build a new virtualized and open 5G network.

NTT DoCoMo:

In Sept. 2019, NTT DoCoMo announced that it successfully worked with Fujitsu, NEC and Nokia on multivendor interoperability for its 4G and 5G

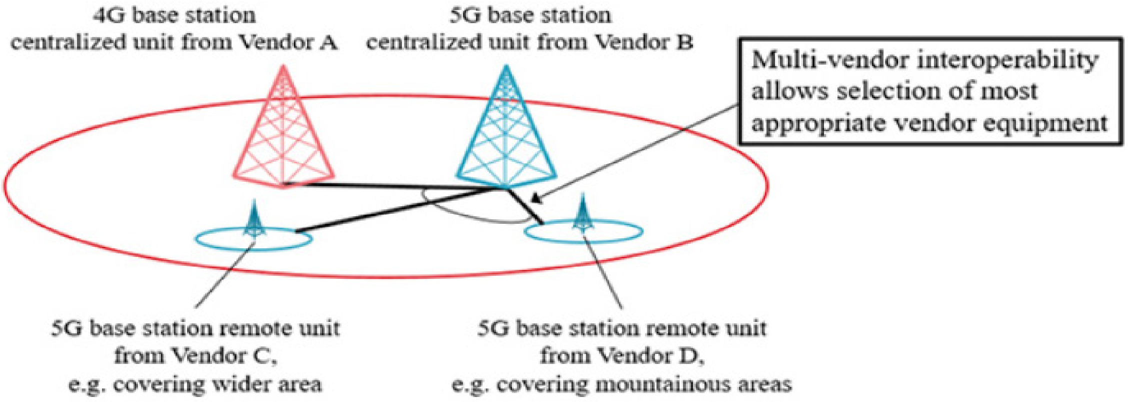


Figure 2-9 NTT DoCoMo 4G and 5G multi-vendor interoperability [15]

base station using O-RAN Alliance specifications. DoCoMo will deploy this in its pre-commercial 5G network.

NTT DoCoMo has adopted O-RAN fronthaul specifications to connect remote radio units with centralized baseband units, and the O-RAN X2 profile specification to connect between 4G base stations and 5G base stations from different vendors, shown in Figure 2-9.

2.5 Advantages and Challenges with Open RAN Architectures

Operators are moving to Open RAN for many reasons. One of the goals of Open RAN is to maximize the use of common off the shelf hardware and minimize proprietary hardware, while making the interfaces between all hardware components standards based with the motive of growing the economies of scale and expanding the ecosystem so that smaller companies can infuse new innovative solutions into the telecom infrastructure space. Open interfaces, open source hardware and software have many benefits, however, the underlying technology and infrastructure design drives the performance and overall value of a RAN product. Open RAN architecture is still evolving and realistic expectations should be considered on the time frame Open RAN can be realized.

Figure 2-10 and Figure 2-11 show the high level benefits and challenges with Open RAN approaches. Some advantages and challenges associated with specific aspects of the Open RAN architecture are further highlighted in the discussion below:

Potential Benefits:

By disaggregating RAN software ecosystem from hardware ecosystem and by standardized open interfaces Open RAN aims to allow multiple vendors to spur the 5G RAN network evolution. It allows for vendor diversity while allowing operators to choose best of breed solutions. By leveraging open software development and by use of cloud technology and tools, Open RAN promises to reduce the Total Cost while potentially improving time to market for new functions and features.

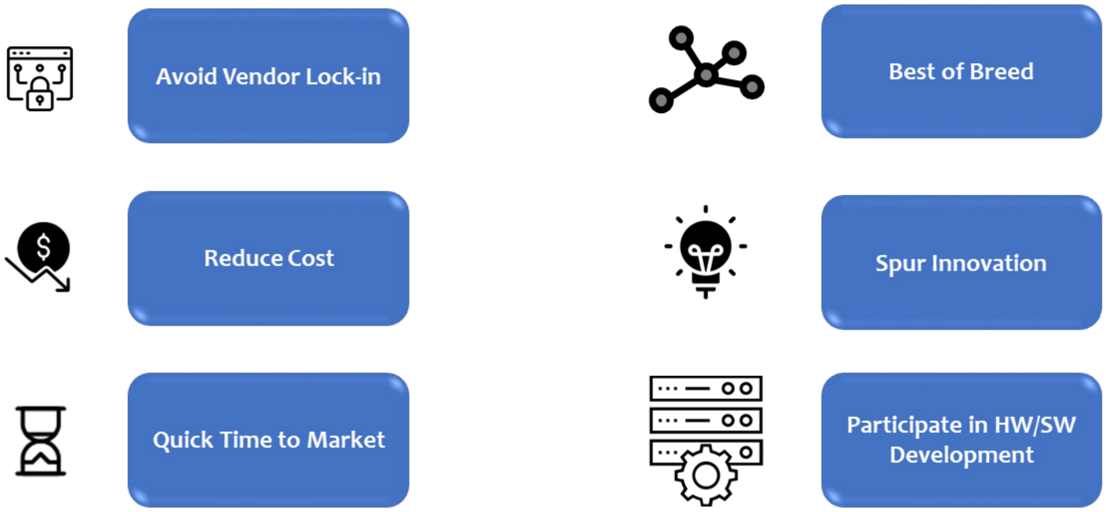


Figure 2-10 Potential Benefits with Open RAN

Potential Challenges:

There are some challenges related to Open RAN in addition to RAN Virtualization challenges detailed earlier. Given the multi-vendor environment, vendor interoperability and system integration are obvious challenges. Automation: The software programmable interfaces for service management and orchestration of RAN workloads such as A1, O1 and O2 are being to be defined in O-RAN alliance. Automation for deployment at scale will need to evolve and mature along with multivendor implementation and roadmap alignment.

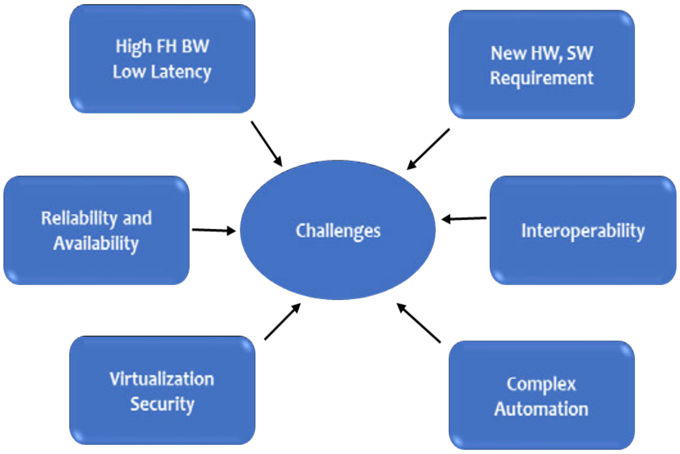


Figure 2-11 Potential Challenges with Open RAN

RIC challenges:

In the AI/ML domain, Open RAN could bring in significant innovations. However, careful consideration is required before adoption of RAN Intelligent Controller (RIC) flavors. This is to address the challenge where competition for the same resources by redundant functions could occur resulting in system instability. However, the O-RAN Alliance is defining arbitrator function to control such conflicts in the RIC.

Security challenges:

Open RAN also adds new security challenges. Open RAN could expand the threat surface due to more disaggregated functions and addition of more interfaces. Open Source code, while benefitting from community scrutiny, is also open for malicious hacking and increases exposure to public exploits. Unsecured/less secure management interfaces could become a point of security vulnerability through the service management and orchestration layer. The need for consistency of implementation of security best practices in an open multi-vendor environment poses a practical challenge for secure 5G deployments. To address some of the above challenges, O-RAN Alliance has set up a Security Task Force to collect security requirements and develop security architecture, framework and guidelines across the entire O-RAN architecture.

While these challenges may exist, many could be resolved over time with the continued maturity of the vendor ecosystem and the experience gained in deploying large systems for the operator community.

2.6 Operational Considerations and Integration Challenges

For an operator to move from a proprietary vendor deployment approach to an Open RAN model, a number of key challenges have to be addressed:

- Increased operational complexity
- Security standards for open interfaces and platforms
- Integration and interoperability challenges
- Shifting operator role in system integration
- Accountability issues with end-to-end network performance
- True TCO savings realizable in brownfield networks

2.6.1 Integration Challenge and Increased Operational Complexity

Operators evaluating Open RAN will choose their own business cases around greenfield, brownfield or mixed deployments, each of which come with their unique set of challenges. For greenfield deployments, timelines may be dictated by the maturity of Open RAN standardization and the development ecosystem in addition to operational challenges, whereas brownfield and mixed environments would additionally be burdened with potential forklift costs and integration challenges because equipment specifications are not open. With significant network assets carrying live 2G, 3G, 4G and 5G traffic, the case for deploying an Open RAN network is also dependent on platform compatibility issues. With 5G deployment planning and implementations underway, some operators may opt for a more measured approach towards mixed deployment of 5G using Open RAN; while others may see 5G as an inflection point or an opportunity to implement Open RAN.

A common cause for concern from network operators is based around multiple vendors / partner solutions replacing what traditionally was a single vendor solution. A potentially significant increase in overall complexity is expected due to the following factors:

- Higher amount of due diligence needed for product and system integration
- Greater validation and troubleshooting issues
- Complex end-to-end performance assessment
- Increase in system maintenance activities
- Complicated upgrades

Product and system integration aspects may include standalone node integration, SW with HW integration (vertical) and integration between different SW modules (horizontal), as well as E2E system integration of a logical node with other network components (core, UE, transport, management systems and deployed radio equipment). Interoperability between multiple parties and components cannot be achieved at the expense of “simplified” interfaces or reduced functionality, as the key requirement from operators is that the end-to-end performance and functionality cannot be compromised.

Likewise, there can be no reduction in overall support and maintenance capabilities, in spite of the disaggregated solution being provided by multiple parties. The process to perform regular operations with the same amount of granularity and efficiency as available with proprietary networks will likely be very challenging. The overall system may be integrated under a common management umbrella and upgrades will have to be carefully planned and coordinated between suppliers. Even after the initial integration, every time a new feature is added in any one of the components, it may require incremental re-integration, re-testing and re-validation. In this scenario, operators may have to step up and take a bigger role to ensure different component suppliers cooperate in aligning roadmaps for new feature support as well as for correcting and solving network issues in accordance with SLAs similar to or better than existing solutions.

Much hinges on the competence of the system integrator (or feasibility of the network operator), to work hand-in-hand with vendors every step of the way. It is not expected that new players in the RAN ecosystem bringing expertise in a specific area (e.g. RIC or Active Antenna) should have the breadth to take on that role, nor is it likely that established vendors with relatively broader scope will have the necessary depth in every functional area as offered by the innovative players that specialize in it. Traditional IT domain integration experts are not an option either, and systems integration players in the mobile telecom sector will likely not have this specific Open RAN integration experience for operators to count on in the near-term. The stark reality is that the model can be extremely challenging to maintain as the overall operation of the network is made significantly more complex with multiple vendors in the mix.

To assist with integration, interoperability and certification, some Open RAN standardization bodies have developed interface test specifications, testing centers and industry plugfests. For example, the O-RAN Alliance has specified Open Front Haul

Interoperability and Conformance Specifications and is deploying several global O-RAN Testing and Integration Centers responsible for system and interface certification and vendor interoperability. Additionally, Open RAN and TIP hold industry plugfests to facilitate vendor interoperability and promote Open RAN to the industry.

2.6.2 Shifting Operator Role

In addition to the role of an integrator that the operator has to play (or rely on outsourcing that activity and still oversee everything with less direct involvement but with full responsibility), the other area of focus is the need for more immediate and urgent consideration given towards the upskilling of network and field operations teams to run a variety of services in a complex Open RAN network, with critical implications for day to day performance variations or security issues. Operators will need training and hands-on experience in every functional block or system component, dealing with a set of known vendors and potentially with implementations from unknown open source contributors.

2.6.3 Realizable TCO Savings

The Open RAN move towards standard COTS and/or white label hardware is expected to drive significant cost savings and supply-chain simplicity with hardware replacements and inventory management, which is a very desirable outcome for most Network Operations teams. On the other hand, with more vendors to deal with, the relationship value (measured in payments) is lower for each vendor compared to a fully sourced single vendor revenue model. While potential benefits from the lower-cost lure of Open RAN may offset some of that, the tradeoffs will likely vary case by case. For instance, some operators and vendors are concerned that the use of a system integrator will potentially come at a steep cost and that it could be a risk to the business given the likely need for complicated business models around support agreements with the component/functionality suppliers resulting in lengthy resolution processes.

As suggested in Figure 2-12 below, the TCO comparison depends on a variety of new business roles and the impact on TCO with large scale operations can go either way. The potential Capex savings from using GPP platforms may be more than offset by the Opex increase due to the resources needed to support the greater operational complexity.

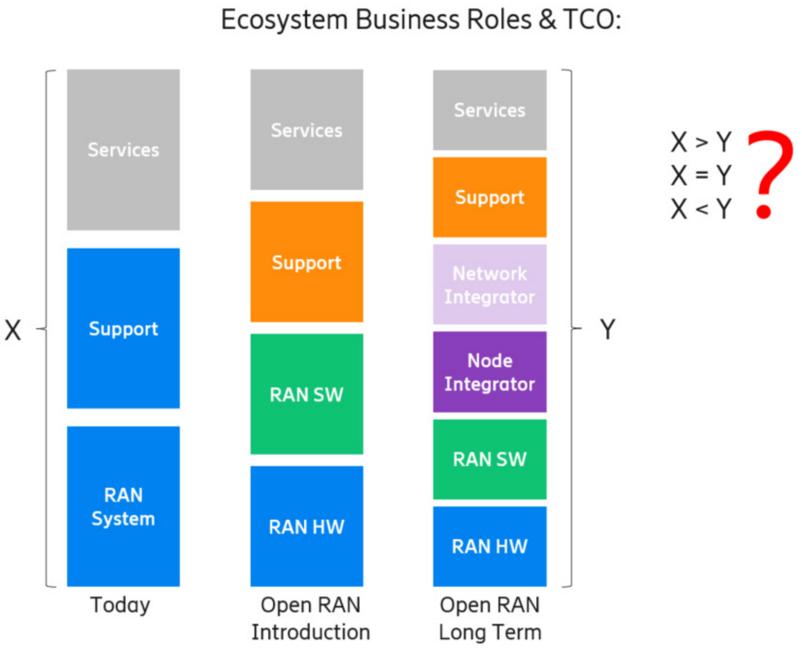


Figure 2-12 Ecosystem Business Roles and TCO

2.6.4 Performance Considerations

The premise of Open RAN includes a greater degree of innovation by leveraging the skillsets of a broader community of designers, engineers, developers etc. As the Open RAN ecosystem is designing novel architectures for next generation technologies, the use cases themselves are evolving and requirements are being investigated. These dynamic aspects are a big challenge for proprietary RANs, and of course apply to Open RAN systems too. It is to be seen whether the novel architectural aspects (e.g. RIC and use of AI/ML) are robust enough to meet the challenging requirements of the upcoming use cases, and if the overall implementation flexibility and resulting performance with the open community based design will be better relative to proprietary RAN systems where vendors provide their special sauce to improve spectral efficiency, manage interference and increase system throughput using components and designs they have full control over. Lack of specialized proprietary implementations of highly advanced functionalities (e.g. digital beamforming, MU-MIMO etc.) might even limit relative performance and flexibility in the near term.

3. Artifical Intelligence (AI)/ Machine Learning (ML) RAN in Open RAN Framework

3. Artificial Intelligence (AI)/Machine Learning (ML) RAN in Open RAN Framework

Enhancing RAN performance with the use of Artificial Intelligence and Machine Learning has many potential benefits and considerations. This chapter introduces the concepts of application of AI and ML to Open RAN networks, identifies architecture requirements with specific use cases, provides a brief survey of related literature, and outlines potential future avenues for AI/ML technologies across 5G and future networks.

3.1 Introduction

5G networks enable operators to provide a vastly expanded range of services across a diverse set of technologies and spectrum. The flexibility and richness of 5G could make it more complex to optimize and manage, with a wider range of performance KPIs parameters to optimize. Some network operators are already dealing with complexities of deploying and managing 5G while maintaining previous generations of wireless networks, and some are also introducing Open RAN networks. The traditional human-intensive means of deploying, optimizing and operating radio access networks may not be able to achieve the level of optimization needed. Artificial Intelligence enabled solutions hold the promise of managing this scale of complexity with capabilities such as auto-configuration, self-driving and self-healing networks that use new learning-based technologies to automate operational network functions and reduce OPEX. This new "intelligent" RAN should be able to sense its environmental and application context, as well as interpret and act on the contextual information in real-time extremely efficiently. Furthermore, device and resource control functionality should be able to take advantage of the de-coupling of the User Plane (UP) and Control Plane (CP) in Open RAN to offer efficient and optimized closed-loop network management capabilities using advanced analytics and data-driven approaches, including advanced artificial Intelligence (AI)/ Machine Learning (ML) enabled applications close to the edge of the RAN networks.

The key benefits of Open RAN with respect to AI/ ML based optimization and automation are:

- Use of interoperable open interfaces to perform data collection, and configuration changes for these tasks.
- Use of open APIs to implement algorithm clusters (such as rApps and xApps in RICs) to allow multiple solutions to be tried and tested for best results for the same use case.
- Allowing the operators to take control of their networks and innovate at their own pace, if they choose to.

O-RAN is expected to have the inherent ability to offer efficient, optimized radio resource management through closed-loop control to enhance network performance and user experience. Further, the platform affords the ability to control, at a per node and/or per-UE level, load balancing, mobility management, multi-connection control, QoS management and network energy saving.

All variants of Open RAN architecture target achieving these goals by embedding intelligence, at component and network levels, to enable dynamic local radio resource allocation and optimize network-wide efficiency. For instance, the TIP Radio Intelligence and Automation (RIA) workstream is discussing topics for achieving a more intelligent network. These topics include ML-based network optimization and planning, massive MIMO estimation and optimization, intelligent energy savings and interference mitigation.

In O-RAN Alliance, "Intelligent RAN" is a key stated objective. In the O-RAN Alliance's reference architecture, the introduction of the hierarchical non-Real Time (non-RT) and near-Real Time (near-RT) RIC with the A1 and E2 interfaces is aimed at enabling an entirely new eco-system of intelligent features and applications residing close to the edge of the RAN network to fulfill the above stated goals. Hence, in this chapter we use the O-RAN architecture from the O-RAN Alliance to discuss the utility and applicability of AI/ML techniques for efficient network operations.

This chapter is organized into five (5) sections:

1. Overview of O-RAN Alliance architecture regarding AI/ML
2. Summary of O-RAN Alliance use cases containing AI/ML elements
3. Brief survey of relevant literature
4. AI/ML use cases and considerations
5. Conclusion

3.2 Overview of O-RAN Alliance architecture regarding AI/ML

O-RAN reference architecture specifies the introduction of a hierarchical RAN Intelligent Controller (RIC) platform and leverages "compute" capabilities of a cloud-native environment to enable AI/ML driven intelligent decisions, and automation in the RAN. Figure 21 pictorially shows the overall structure of the O-RAN logical architecture as given in the O-RAN resource document on Use Cases and Deployment Scenarios [16].

The Non-RT control functionality (> 1s) and near-Real Time (near-RT) control functions (< 1s) are decoupled in the RIC. Non-RT functions include service, configuration and policy management, RAN analytics and model-training for the near-RT RAN functionality. Some of the trained models and real-time control functions produced in the non-RT RIC are distributed to the near-RT RIC over A1 interface for runtime execution. It should be noted, as O-RAN specifications stay within the 3GPP specifications for Radio Resource Management (RRM) and relevant Network Management System (NMS), that these additional capabilities are expected to be compatible with deployed RAN and SON implementations.

The insertion of this new compute platform in the reference architecture also introduces additional "open" interfaces. The interface between non-RT RIC and near-RT RIC is called A1. This interface is aimed at enabling service management function to optimize RAN performance. The A1 interface provides support for Policy-based Guidance Service, AI/ML Model Management Service e.g. training, updating and deploying ML model and Enrichment Information Services, by providing access to RAN-external sources that could be beneficial for RAN optimization performance.

The interface between the O-RAN Managed Element and the management entity defined to support FCAPS management, Software management, File management and related OAM functions is defined as O1.

The interface between the near-RT RIC and the Multi-RAT CU protocol stack and the underlying RAN DU is called "E2". This interface is involved in facilitating functions closer to the edge RAN networks and operates on a granular (per-UE) level for controlled load-balancing, resource management and interference management functions. Of course, there are other significant functions that can also reside in this near-RT RIC which are explored later in the chapter. It should be noted, that near-RT RIC may initiate configuration commands directly to CU/DU, and it is ideally positioned for realization of ML-enabled fast closed-loop actions.

The near-RT RIC is positioned as an "open" compute platform and can host third party applications, an important goal of the O-RAN Alliance to facilitate innovation and openness. These applications, referred to as xAPPS, are expected to be trained ML models working in a cloud-based environment close to the edge of the RAN network, performing near-RT data-based decisions and closed loop actions towards the CU/DU.

Finally, O-Cloud, the O-RAN cloudification and orchestration platform, facilitates flexible deployment options and service provisioning models of O-RAN virtualized network elements in telco clouds. O-Cloud is the cloud computing platform comprising a collection of physical infrastructure nodes that can host the relevant O-RAN functions, the supporting software components and the appropriate management and orchestration functions.

3.3 O-RAN AI/ML Use Cases

In the O-RAN Alliance paper titled "O-RAN Use Cases and Deployment Scenarios White Paper" [16], an initial set of key O-RAN use-cases are outlined and can potentially serve as candidates for using AI/ML techniques via rApps and/or xApps to embed network intelligence in the RAN network.

The AI/ML use cases defined in this white paper are summarized below:

Phase I:

- Traffic steering
- QOE optimization
- QOS based resource optimization
- Massive MIMO optimization

Phase II:

- RAN Slice SLA assurance
- Context Based Dynamic Handover Management for V2X
- Flight Path Based Dynamic UAV Resource Allocation
- Radio Resource Allocation for UAV Applications

3.3.1 Traffic Steering

Traditional network traffic controls are often cell-centric and hence miss the opportunity to tune their behavior to the different radio conditions of the cell by treating most UEs in the same way using average values from that cell. O-RAN architecture in comparison, is designed to improve this situation by customizing UE-centric strategies and providing proactive optimization by predicting the network condition. Finally, the RAN intelligent module enabled by machine learning is well positioned to be able to control the adaptation of diverse scenarios and objectives. This may be accomplished by the application of non-RT RIC and near-RT RIC control traffic steering strategies through AI/ML learning from the data collected via the O1 interface from the O-CUs and O-DUs.

3.3.2 QoE and QoS Optimization

The highly demanding 5G native applications like Cloud VR are both bandwidth consuming and latency sensitive. To allocate the correct bandwidth, a method to perform closed loop optimization in real time would be helpful. In this way radio resources could be allocated to the UE before the QoE is degraded. In O-RAN multi-dimensional data can be acquired and processed via ML algorithms to support traffic recognition, QoE prediction, and finally guiding close-loop QoS enforcement decisions. A similar method may also be used for QoS based decisions.

3.3.3 Massive MIMO Optimization

The benefits of Massive MIMO antennas in 5G is well acknowledged. The beam and panel control requirements lend themselves inherently to be controlled by AI/ML based optimization algorithms. The target would be to proactively and continuously improve cell-centric network QoS and/or user (group)-centric QoE in a multi-cell and, possibly, multi-vendor massive MIMO deployment area with multiple transmission/reception points, depending on specific operator-defined objectives. The advantages the O-RAN architecture provides to this use case include the possibility to apply and combine both non- and near-real-time analytics, machine-learning, and decision making for various sub-tasks of this use case.

3.3.4 RAN Slice SLA Assurance

In the 5G era, network slicing is a prominent feature that provides end-to-end connectivity and data processing tailored to specific business requirements, and consequently the requirement for relevant supporting SLAs and KPIs is born. Ensuring and optimizing these KPIs is of renewed interest. The O-RAN architecture enables such challenging mechanisms to be implemented, which could help pave the way for operators to realize the opportunities of network slicing in an efficient manner as well as potentially change the way network operators do their business.

3.3.5 Context Based Dynamic Handover Management for V2X

V2X (Vehicle to anything) communications promises numerous benefits such as increased road safety, reducing emissions, and saving journey time. The technology is based upon Cooperative Awareness Messages (CAMs) [17], which contain radio cell IDs and basic radio measurements. As vehicles traverse along a highway, suboptimal HO sequences and anomalies might substantially impair the connectivity and hence the performance of the V2X system. Thus, their mitigation is of prime importance. The O-RAN architecture allows for the collection and maintenance of the radio/HO data and the deployment, continuous retraining, and evaluation of AI/ML based applications that detect, predict HO anomalies on a UE level.

3.3.6 Flight Path Based Dynamic Unmanned Aerial Vehicle (UAV) Resource Allocation

The application of UAV has played a great role in civil applications including agricultural plant protection, power inspection, police enforcement, geological exploration and environmental monitoring. With normal cellular masts and antennas, the coverage can be patchy and interference unmanaged which could lead to the UAV landing or returning to base. In O-RAN architecture, multi-dimensional data can be acquired, for example, non-RT RIC can retrieve the necessary Aerial Vehicles-related measurement metrics from the network based on a UE's measurement report and flight path information, along with other measurements. Near-RT RIC can support the deployment and execution of the AI/ML models from the non-RT RIC. Based on this, the near-RT RIC can perform the radio resource allocation for on-demand coverage for UAV considering the radio channel condition, flight path information and other application information.

3.4 Brief Survey of Relevant Literature

3.4.1 5G Americas – 5G At The Edge

The 5G Americas white paper "5G at the Edge" [18] outlines the implications of the cloud-native 5G architecture and the wide variety of use-cases it is designed to support. Specifically, it explores the Edge Compute use-cases highlighting the need for AI/ML techniques to facilitate the deployment, orchestration and management of these use-cases in the context of 5G.

Figure 3-1 reproduced from the document clearly highlights that AI/ML tools are expected to permeate all aspects of edge systems, including the RAN, driven by the same well voiced requirements of data intensive, real time applications.

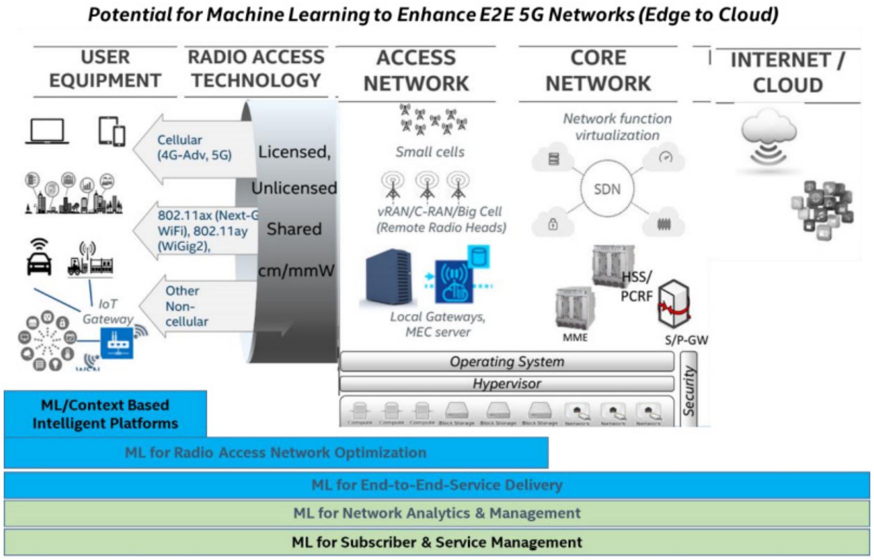


Figure 3-1 ML and AI based techniques permeate all aspects of E2E wireless system design, service management and delivery

With some analysis, each of the following ML capabilities foreseen in this paper may be facilitated with the O-RAN reference architecture via the hierarchical RIC or AI/E2/O1/O2 interfaces for collecting the relevant data. This provides the required cloud-based compute platform and enables the closed-loop actioning to perform or manage:

- Context base intelligent platforms
- Radio Access network optimization
- End to end application delivery
- Network analytic and management (enabler/partial)
- Subscriber and service management (enabler/partial)

The paper goes on to state in section 4.3

"AI and ML approaches will play an important role in enhancing and automating next generation wireless edge networks. They will also enable a ubiquitously available edge learning engine to facilitate the real-time learning required for emerging autonomous/immersive services. Enabling reliable and real-time learning over wireless edge networks, will require a cross-disciplinary approach, capable of understanding the fundamental theory of AI/ML techniques, adapting ML approaches for wireless applications and comprehending the uncertain, dynamic nature of learning over wireless channels. Also of importance are understanding the sensing and storage limitations and their impact on available data sets, as well as addressing compute and privacy concerns in moving the compute resources and data sets across the network. We expect that synergistic and integrated design of wireless networking with edge AI/ML will be key to addressing these challenges."

3.4.2 5G Americas – 5G Evolution: 3GPP Releases 16-17

"The 5G Evolution: 3GPP Releases 16-17" [19] reviews the enablers for Network Automation (eNA) for 5G architecture compared to Release 15 data collection and network analytics exposure features. It defines the output analytics information based on statistics and prediction from the data collected. Some examples of Analytics ID include: Slice Load Level information, Network Performance Information, UE Mobility Information and QoS sustainability. This will clearly help in 5G architectures to improve the possibility of AI/ML algorithms

3.4.3 5G Americas – Precision Planning for 5G Era Networks with Small Cells

"Precision Planning for 5G Era Networks with Small Cells" [20] is most relevant as it gives solid examples of where AI/ML and other advanced algorithms have been used. The positioning use cases are not in the context of O-RAN, so it is recommended that this functionality be replicated in the O-RAN use cases. The case of planning small cell networks with the use of AI/ML from O-RAN should also be considered, as proper placement considerably improves ROI of deployed cells.

3.4.4 5G Americas – Management, Orchestration and Automation

5G Americas White Paper: Management, Orchestration and Automation [21], is very relevant in setting the direction for the overall picture at a high level. The paper does mention O-RAN specifically and replicates similar architectures to those presented in this document. Interestingly, it tempers the fully automated view with this comment "*Due to developments in artificial intelligence, network providers are now able to rethink their operations to achieve end-to-end automation. However, most operators do not want to fully cede control to networks that decide their own direction and remove humans from the equation. They want their networks to become more 'adaptive' to respond to an ever-changing competitive landscape and consumer demands, which requires a coherent combination of human-controlled and –supervised automated operational processes, analytics-driven intelligence, and an underlying programmable infrastructure*".

The white paper provides an outline of the new and enhanced OSS solutions that are being proposed for the management, orchestration and automation of 5G networks. Specifically, it discusses the OSS/BSS system architecture in 5G, SON implementation in 5G and mentions AI/ML however only at a high level in more abstract ways. Reproduced below, Figure 32 from the paper highlights how the learning and decision-making environment for AI/ML solutions is embedded in each layer of aggregation for a large distributed network such as a 5G network.

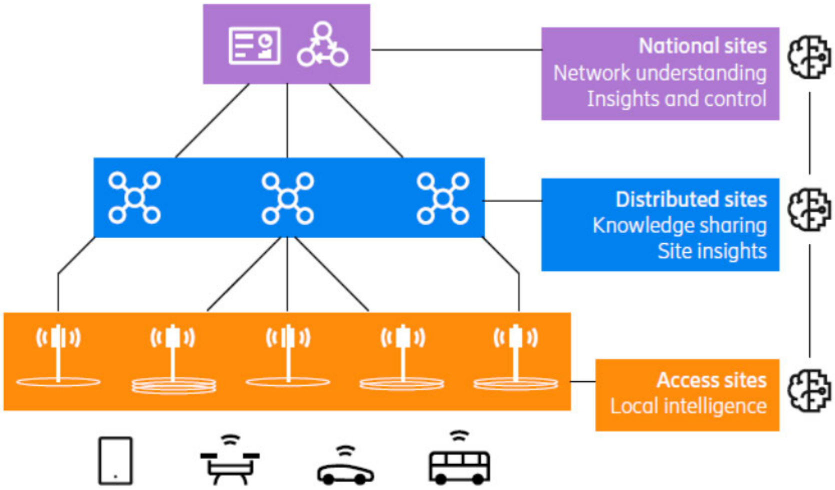


Figure 3-2 Local and global learning and decision making in large distributed networks

The paper also gives indications on the key aspects for 4G, including:

- Provide for coordination between Centralized SON and Distributed SON with the goal to improve efficiency of combined (Hybrid) SON
- Provide for coordination between Distributed SON implementations provided by different vendors
- Provide for joint operations of RAN SON and self-organization functions in the Core Network
- Provide for efficient coordination between RATs: Evolved LTE and 5G

3.4.5 Small Cells Forum: SON Features for 5G NR

Another context explored here is the SON framework evolution expected with the 4G to 5G transition. Within [22] there is a good summary of how the well-known 4G SON functions will be transitioned into 3GPP Release 16 and an overview of how SON and orchestration technologies must evolve to meet the needs of 5G-Era enablers, such as virtualization, Edge Computing, Network Slicing etc. This can certainly benefit from O-RAN AI/ML enabled closed-loop implementation that can enhance data-driven decisions.

The chapter "SON features for 5G NR" contains well written summaries of the key changes to signaling messages in the 5G NR specifications, at a digestible level. These include:

- Automatic Neighbor Relation (ANR)
- Inter-Cell Interference Coordination (ICIC)
- Coverage and Capacity Optimization (CCO)
- Physical Cell Identity (PCI) optimization
- Mobility Robustness Optimization (MRO)

Here are some key takeaways for 5G, identified in the paper:

- *"Though the underlying techniques are likely to be similar, there are key new elements of the 5G standard that enable better measurement and reporting – and hence hold the promise of better ANR"*
- *"Experience with the real problems of 5G networks in the years to come will reveal if the higher densities of these networks and the overlapping of bandwidth parts due to varying numerologies across neighbor cells really leads to performance degradation of the 4G-era SFR-based ICIC algorithms. In such cases, new approaches to ICIC will need to be standardized."*
- *"The existence of highly disparate frequency bands, each with their unique propagation characteristics, will require a diverse set of tools for capacity and coverage planning."*

For the implementation of Small Cell networks, the paper [22] explains a large number of essential enablers of dense HetNets all of which need to be coordinated to achieve optimal usage of network assets. See Figure 3-3.

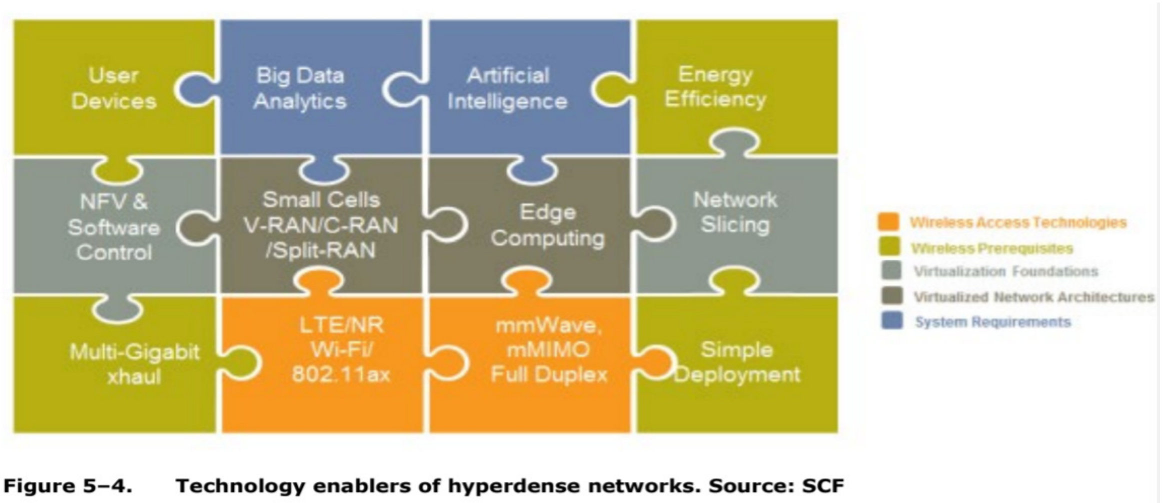
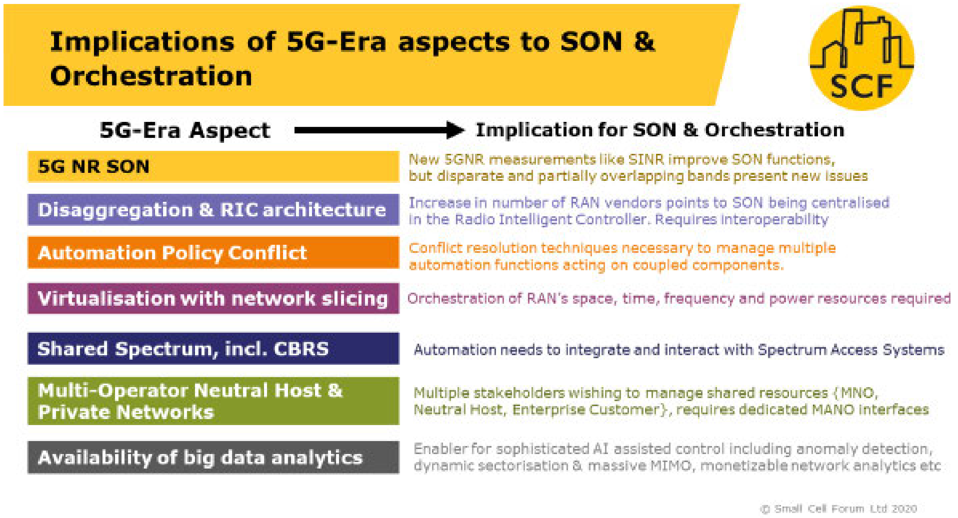


Figure 5–4. **Technology enablers of hyperdense networks. Source: SCF**

Figure 3-3 Technology enablers for hyperdense networks. Source: SCF

SCF also analyzes the implications of various aspects of 5G-Era technologies to SON and Orchestration. A summary is shown in Figure 3-4:



Implications of 5G-Era aspects to SON & Orchestration

SCF

5G-Era Aspect	Implication for SON & Orchestration
5G NR SON	New 5GNR measurements like SINR improve SON functions, but disparate and partially overlapping bands present new issues
Disaggregation & RIC architecture	Increase in number of RAN vendors points to SON being centralised in the Radio Intelligent Controller. Requires interoperability
Automation Policy Conflict	Conflict resolution techniques necessary to manage multiple automation functions acting on coupled components.
Virtualisation with network slicing	Orchestration of RAN's space, time, frequency and power resources required
Shared Spectrum, incl. CBRS	Automation needs to integrate and interact with Spectrum Access Systems
Multi-Operator Neutral Host & Private Networks	Multiple stakeholders wishing to manage shared resources (MNO, Neutral Host, Enterprise Customer), requires dedicated MANO interfaces
Availability of big data analytics	Enabler for sophisticated AI assisted control including anomaly detection, dynamic sectorisation & massive MIMO, monetizable network analytics etc

© Small Cell Forum Ltd 2020

Figure 34 SON and Orchestration in the 5G Era

While SCF's stated purpose isn't about focusing on all these aspects, the paper suggests the role of open management and orchestration (MANO) functions, when underpinned by the goal of full automation, will naturally lead to *"eventually leverage machine learning and artificial intelligence techniques to evolve beyond just automating the optimization of the network, and to add a hefty dose of intelligence to those automated decisions".*

With respect to the O-RAN solutions specifically, the paper summarizes the SON and Orchestration functionalities addressed in O-RAN into the following two areas:

Non-real-time functionalities:

- QoE cross-layer guarantee
- Load balancing
- Multi-cell massive MIMO beamforming (BF) optimization
- Data collection for AI/ML analytics
- Customized algorithm deployment and enablement through A1 interface
- Alarm and troubleshooting

Near-real-time RIC functionalities:

- Admission control
- Bearer admission and modification
- Mobility management
- Load balancing
- Inter-cell interference coordination
- Multi-DU/cell radio resource management

The paper also highlights the *'significant'* challenges with the O-RAN approach and notes that *"this transformation to a true multi-vendor framework will happen gradually. Initially, it is likely that there will be one dominant SON implementation, with some additional third-party applications provided as extensions. Such integration of 'native SON' with third-party SON presents its own set of challenges."*

In short, the key take-away is *"Gradual movement from native SON/RRM to RIC architecture is likely. The transition will present challenges – for example, conflict between the native and third-party solutions and parameter matching. The industry needs to come together to smooth this transition. The success of the RIC architecture will depend on this aspect".*

Next, the paper discusses four AI/ML use cases:

- Dynamic sectorization in a virtualized RAN to get a significant improvement in user experience by adapting the sector boundaries to the changing conditions through the day, hence reducing inter cell interface and the number of handovers
- Location prediction based upon the relative power from each UE, the channel impulse response and the last few sets of measurements that have a correlation to users' locations. This would be very useful for value added services to venue owners for example.
- Anomaly detection of KPI levels from the norm to trigger further manual or a self-healing process
- Elastic scale of virtualized RAN during busy times of the day from the pool of baseband units, in conjunction with NFV

Finally, the paper recognizes that *"O-RAN Alliance's management models are being developed mainly for macrocells and also assuming significant data volumes per cell. The scale of small cell deployments requires leaner models in order that the management traffic remains a small overhead in the network. SCF is looking at which management settings and reports might be superfluous for small cells as well as compression techniques, in work that will be highly complementary to that of O-RAN."*

3.5 What's Next?

O-RAN architecture enables Intelligent RAN control using AI/ML. RIC based AI/ML demos have been proven and performed in numerous venues [23]. There are some exciting future applications for AI/ML technologies in the RAN, in creating an intelligent network. This section outlines several application possibilities, including ensuring quality of experience with application-aware RANs, diving into the role of AI/ML in total cost of ownership optimization, highlighting the role of AI/ML on network reliability management, improving the automation and AI capability gap and identifying issues surrounding trusted AI adoption.

3.5.1 QoE AI/ML Use Case Example – Application Aware RAN

Most of the current AI/ML use-cases, specified in Section 3.3 as Phase I and Phase II are still in an early definition and benefits evaluation phase. There is a clear need for a much richer set of use-cases, especially targeting the growing 5G adoption happening worldwide. 5G technology brings tangible benefits in terms of latency and throughput, however, significant effort and resources need to be expended to measure and manage a subscriber level Quality of Experience. Additionally, as many applications in 5G can be very sensitive to latency, traditional optimization techniques may not suffice to address the new real time requirements. Detection of traffic type in real time becomes more essential to identify the right set of actions to improve the network quality (see Figure 3-5).

This challenge requires introduction of intelligent AI/ML techniques which need to operate in a high volume and high velocity Big-Data environment. An O-RAN near-RT RIC platform becomes an ideal enabler.

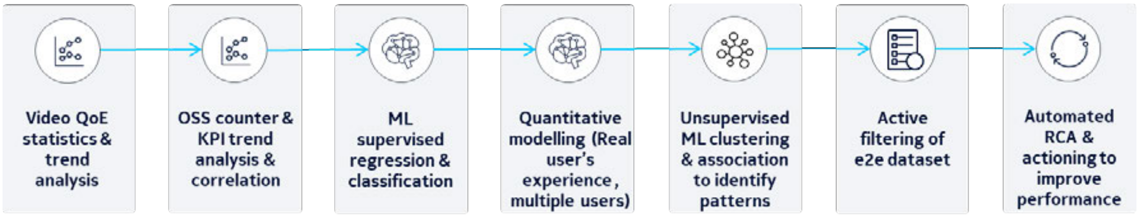


Figure 3-5 Illustrates Video QoE related AI/ML use-case process steps

This is accomplished by moving real time end to end data collection and correlation away from traditional OSS and probing systems directly to the edge of the RAN (at near-RT RIC) where the correlation of the user experience data set with radio signaling (L3) and MAC (L2) can be done on the fly. These insights can then be fed to sophisticated AI/ML algorithms that first detect the traffic type QoE, especially for encrypted video and gaming users, and then provide the end to end root cause analysis.

3.5.2 Role of AI/ML in Network Planning and TCO Optimization

As 5G networks proliferate, "5G-era Mobile Network Cost Evolution" study performed by GSMA [24] points to a subtle shift in 5G focus from mobile operators, with attention moving towards fine-tuning 5G deployments and optimizing 5G-era costs. As shown in the graphic reproduced from the study, RAN infrastructure and energy are the two largest Total Cost of Ownership (TCO) accelerators – implying the largest impact area for operators cost considerations (see Figure 3-6).

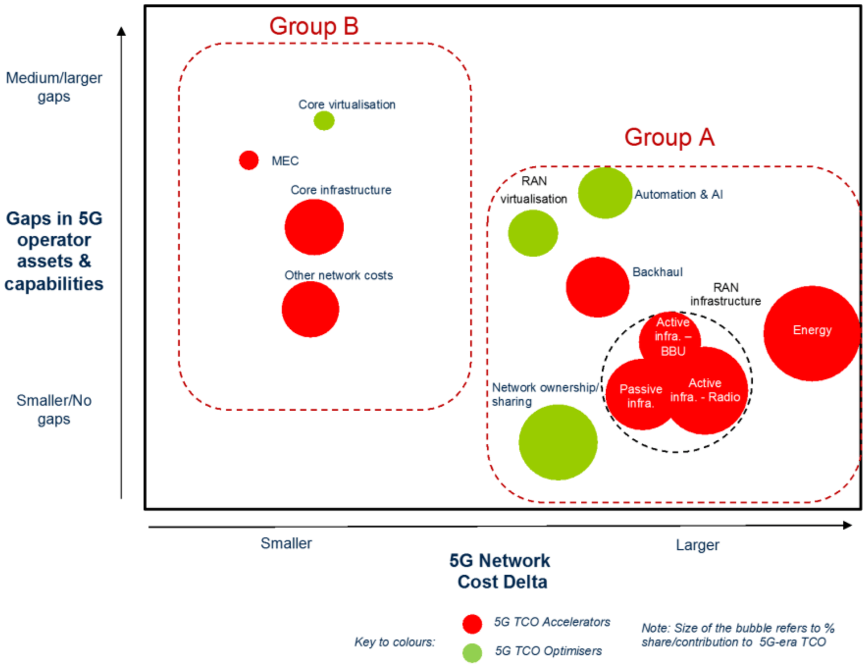


Figure 3-6 5G Total Cost of Ownership Accelerators and Optimizers

This underscores the need for Open RAN architecture to bring in additional savings levers into the transition from 4G to 5G. Motivated by the impact, we anticipate significant opportunities in utilizing AI/ML based rApps and xApps developed to improve the resource management, resource utilization and performance assurance areas. Several of these areas, such as Traffic Steering, QoE Optimization, QoS based resource optimization are already defined as Phase I targets in O-RAN development.

We identify four additional areas where focus from Open RAN community and the industry is likely to be required to meet the future challenges:

- Energy Savings use-cases are conspicuous by their absence on the currently defined O-RAN areas of focus in either Phase I or Phase II
- Network Assurance and Reliability use-cases while addressed may require additional focus for further development
- Automation and AI capability gap, as noted in the study referenced above, can be the biggest impediment in unlocking further benefits from Open RAN AI/ML capabilities
- Trusted AI adoption

3.5.2.1 AI/ML Power Management

With the advent of 5G and technologies such as massive MIMO, energy consumption in 5G is expected to be much higher. In addition, with concerns around global warming, it is no surprise that a focus on energy usage and how to improve power management through AI/ML should be considered in O-RAN. GSMA Intelligence [25] explains how the cost of energy is responsible for 20-40% of OPEX (and 2-3% of world energy usage). Combined with this a predicted x4 growth in data traffic by 2025 is expected to put additional pressure on the energy consumption of many networks. A key requirement would be to understand what protocols or procedures are built into the O-RAN environment to allow AI/ML algorithms to take the optimization of power beyond the 5-15% typical values outlined in the GSMA document

3.5.3 AI/ML Network Reliability Management

Today’s network infrastructure is designed to be over-provisioned for redundancy and to ensure hardware failures do not cause service outages. This ties-up capital in the infrastructure. This way of achieving network “Reliability” can potentially be improved by analytics driven by AI/ML providing a smarter mechanism to predictably handle such failures. The vast amount of network telemetry data already available can be ingested by algorithms to detect outliers and anomalies before the occurrence of the failure. The stream of telemetry can be analyzed in real-time by trained models to generate valuable inference of the health of the system and predict the time-window available to take corrective action.

In the era of 5G the sheer number of CU/DU instances, number of RRUs and distributed hardware assets is expected to overwhelm the traditional network monitoring and service reliability efforts that are in-use at major networks today. Without an AI/ML-driven smart management system, service providers won’t be able to improve reliability of their network service, while containing cost.

3.5.4 Challenges: Scalable Automation and AI Capability Gap

As highlighted by the study above, cloud technologies such as Containers, Microservices and Kubernetes, together with AI/ML techniques and algorithms are key to unlocking the TCO benefits. However, these are relatively newer technologies for the telecom context. Enabling the capabilities and upskilling the workforce for operating and hardening the solutions for telco-grade operations are each a steep hill to climb on its own. The operators would be well advised to take these challenges seriously upfront with a plan for success, or these technologies and investments may not yield the returns. While the focus on interface standardization makes node level interworking feasible, lack of standardized network configuration and performance data exchange makes Data Access, Data Pipelines and Data Validation unable to scale fully yet. Consequently, while there are several tens of operators who have trialed and piloted AI/ML operationally in live networks, large scale deployment and operations of AI/ML assets are still rare and elusive. To overcome the challenge of scaling and operationalizing these AI/ML use-cases will require a combination of deep telco domain expertise, experienced Data Scientists and Cloud/big-Data technology fluency. However, the benefit of adopting the AI/ML based automation & optimization framework should outweigh these perceived challenges for the operators. The alternative to not moving in this direction is to stay on the old and manual process of running the next gen network infrastructure which is likely to become increasingly dense (think small cells) and dynamic (think micro-services).

3.5.5 Challenges: Trusted AI Adoption

With the concerns of user privacy and security widely reported in many areas of the press it would be worth considering how the newly formed AI algorithms would be deemed to be 'safe' for use. In this case there is a need to encapsulate a set of very disparate requirements ranging from user/governmental (General Data Protection Regulation - GDPR as an example) at one extreme to network functional integrity at the other. There are many security models such as International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC) TS 27008 [26] and its associated documents which focus on the security within an organization and this may help to form relevant thinking to tackle such issues.

Besides, the security concerns, another significant deterrent to adoption can be simply lack of confidence in the outcomes from an AI/ML algorithm before it can be trusted with active closed -loop actions. This will require a strong testing, learning and validation environment together with data and metric driven continuous accuracy measurements.

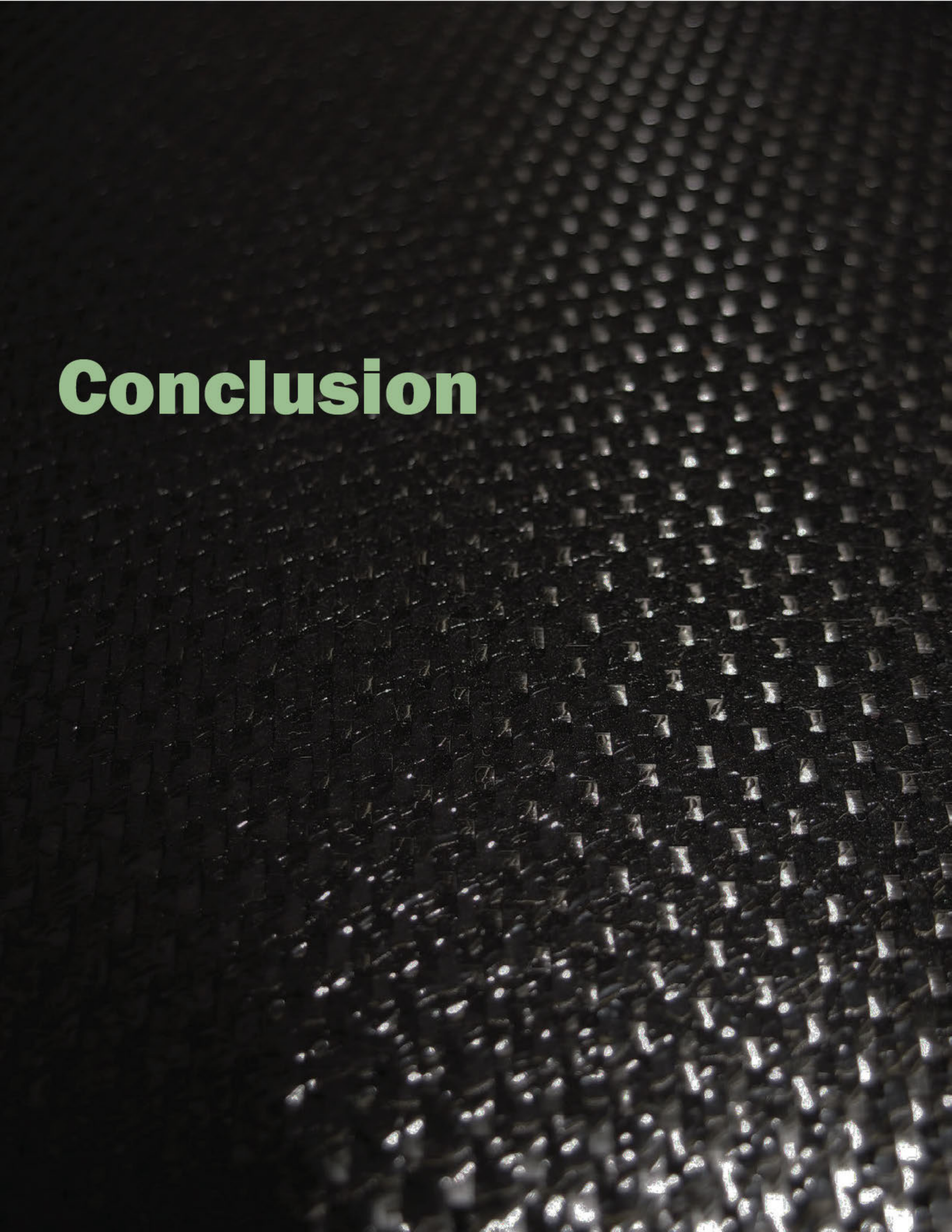
3.6 Closing Thoughts on AI/ML enabled Open RAN

Using AI/ML algorithms and running multiple models on the dataset allows tweaking of a large set of configuration parameters that were not humanly possible to fine tune so far. To make the matter even more interesting, the AI/ML algorithms can run continuously to monitor performance drifts and auto correct for it with differential compensations in configuration values.

The use of O-RAN architecture with its non-RT, near-RT RICs and with its open interfaces such as O1, A1, E2 can enable operators to efficiently deploy 5G and leverage the power of AI/ML to solve the RAN Performance and Operations conundrum. The architectural support for additional cloud-based compute near the edge of the RAN network coupled with open eco-system for rApps and xApps like solutions are ideal for overcoming the TCO challenge of running an increasingly complex and expensive network. Moreover, AI/ML has become a core-requirement for telco operators given the promise of additional revenues from 5G tied to Slicing and SLA driven services in the near future.

As with any new concept, there is a set of challenges to overcome. These range from how to 'trust' deployed AI/ML algorithms and how to manage the integration of many different third party vendors to developing newer use-cases that target additional areas from automation and optimization perspectives. The AI/ML use-cases are still in early definition and benefits evaluation phase. However, the destination is clear and enabled by Open RAN.

Conclusion



Conclusion

The Open RAN approach disaggregates software from hardware, allowing RAN software to run on common hardware platforms. Open, standard interfaces allow best of breed vendor components to be integrated. An open RAN approach enables innovation, competition and drives down the cost with the global vendor supply chain.

Architectural considerations include flexibility for dynamic configuration (through virtualization and network slicing); support for multiple RAN function splits between CU, DU and RU; development of open interfaces for interoperability between multiple vendors; support for RAN virtualization with associated accelerator options for handling processing needed for the full suite of 5G use cases.

The Open RAN ecosystem is quite active with multiple organizations including 3GPP, Cisco Multi-Vendor Open vRAN, O-RAN Alliance, Small Cell Forum, Telecom Infra project (TIP), and the recently formed Open RAN Policy Coalition. Each of these groups with their own distinct perspectives have enriched the Open RAN movement raising interest and awareness across the industry. It is encouraging to see all of these organizations working together by coordinating with each other with liaison activities, leveraging existing work to avoid potential duplication and streamlining the move towards the common goal of establishing Open RAN architectures. Collectively, these bodies have defined a rich set of specifications and approaches to add to the 3GPP baseline with new open interfaces and architectures leveraging AI/ML capabilities, spurring great interest with global operators and vendors leading to multiple trials and deployments.

The benefits of disaggregation and virtualization, open innovation, faster feature development and supply chain advantages are hard to ignore, however operators are also conscious of the need for end to end management, troubleshooting, end-to end integration for the system as well as with any deployed network etc.

It is no surprise that with this ever-increasing complexity of mobile networks, the use of Artificial Intelligence (AI) and specifically Machine Learning (ML) algorithms to help manage the network is an inevitability. The manual interventions required in 2G/3G and to a larger extent within 4G cannot possibly keep up with the scale and scope of what needs to be achieved in multiple spectrum bands with highly sliced and multiservice networks for 5G and beyond.

Appendix

Appendix

This section provides additional information not otherwise covered at length in chapters above.

Terminology

It is important to align the terminology used with reference to the idea of having an open RAN. This whitepaper utilizes the background terminology proposed by the Open RAN Policy Consortium with several terms being used in the industry by various stake holders.

Open RAN - disaggregated RAN functionality built using open interface specifications and interoperability between elements. Can be implemented in vendor-neutral hardware and software-defined technology based on open interfaces and community-developed standards as opposed to closed proprietary interfaces.

O-RAN – O-RAN Alliance or designated specifications. A specification group is defining next generation RAN infrastructures, empowered by principles of intelligence and openness. https://www.o-ran.org

TIP – Telecom Infra Project. A community of more than 500 network operators, technology companies, telecom equipment vendors, standards organizations and Internet companies developing open and interoperable technologies using open interfaces. https://telecominfraproject.com/

vRAN – an implementation of the RAN in a more open and flexible architecture which virtualizes network functions in software platforms based on general purpose processors

3GPP - 3rd Generation Partnership Project, composed of seven telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC) which adopt 3GPP specifications as standards. The project covers cellular telecommunications technologies, including radio access, core network and service capabilities, which provide a complete system description for mobile telecommunications. https://www.3gpp.org/

SCF - Small Cell Forum, defines specifications to accelerate deployments, and works to remove commercial & technical barriers to network densification using open small cell architectures. https://www.smallcellforum.org/

RAN Functional Splits

RAN Split Option 2

3GPP has standardized the RAN split option 2 with Radio Resource Control (RRC) and Packet Data Convergence Protocol (PDCP) functions hosted in the CU while the rest of the RAN functions remain within the DU and stay at the cell site. The interface between the CU and DU is the F1 interface, which does not have as stringent latency and bandwidth requirements as a lower layer split. 3GPP defines the control and user plane interfaces as F1-C and F1-U. Furthermore, 3GPP standardizes the interface between the user plane and control plane as the E1 interface.

RAN Split Option 6

The Small Cell Forum (SCF) has standardized the RAN split option 6. This is the split between the MAC and PHY layer which requires more fronthaul bandwidth than option 2. Also, its latency requirement in the range of a couple of hundred microseconds is stringent. SCF defines the control, user, synchronization, and management planes of this interface.

RAN Split Option 7

The Split-7 (low-PHY/high-PHY split) fronthaul solution provides compression capabilities in fronthaul while virtualizing L2 and parts of L1 in the DU and supports advanced receiver functionality such as COMP for performance. Multiple types of compression are supported for standardization.

This split is Ethernet based allowing wide deployment and ecosystem possibilities. It enables many-to-many connections, allowing aspects such as geo-redundancy, edge computing and having the RRUs talk to a separate server for authentication.

Split 7 implementations can be based on Commercial Off-the-shelf (COTS) hardware. Accelerators for the DU, where needed (e.g., for eMBB) are based on FPGAs which can be programmed in high level languages such as OpenCL and P4 and can also be virtualized.

CPRI and eCPRI

CPRI was initially designed providing an interface between the RF and the baseband using fiber optic resulting in replacement of the copper cable that traverses the antenna tower and recovering the copper loss and associated power savings. The RF could be distributed among various sites and the baseband centralized. This would allow pooling of resources at the central location and also provide support for advanced receivers such as COMP and interference management for performance benefits. This maps to the interface between the RF and PHY (L1) – also known as Option 8 in 3GPP discussions.
CPRI has the following limitations:

- It requires 2.4 Gbps data rate per antenna for a 20 MHz system. As we approach massive MIMO systems or larger bandwidth systems for NR with 100's of MHz of bandwidth, this clearly is not a scalable solution.
- It is a point-to-point interface. This implies that all communication must happen with a single entity. There is no way to have some traffic going to one entity (e.g., OAM) while other traffic goes to another entity (e.g., UP traffic)
- It is a streaming interface. This implies there is constant data at the interface at a fixed rate at all times. For example, this means there is no way to reduce the transport data when the eNB is idle, or during low traffic conditions.
- Most vendors use proprietary extensions on the interface, making it difficult to interoperate.
- CPRI is not a widely deployed solution across industries (compared to Ethernet, for example), making the eco-system small and thereby expensive.

Limitations of eCPRI

eCPRI was developed as an extension to CPRI to address some of the limitations of CPRI, for example to provide up to 10x compression in terms of fronthaul rates as well as provide transport over Ethernet. However, a large portion of the messages is still vendor-specific, making the interface very difficult to provide interoperable solutions.

The specification recommends multiple splits across the PHY, MAC and upper layers without really defining the protocol and the messaging on the splits – everything is open and customizable – which seemingly provides significant flexibility – but is really an interoperability and test nightmare. For example four split points for DL PHY and three split points for UL PHY have been defined with no guidance on how to use and what messages to send to support this and how to provide interoperability.

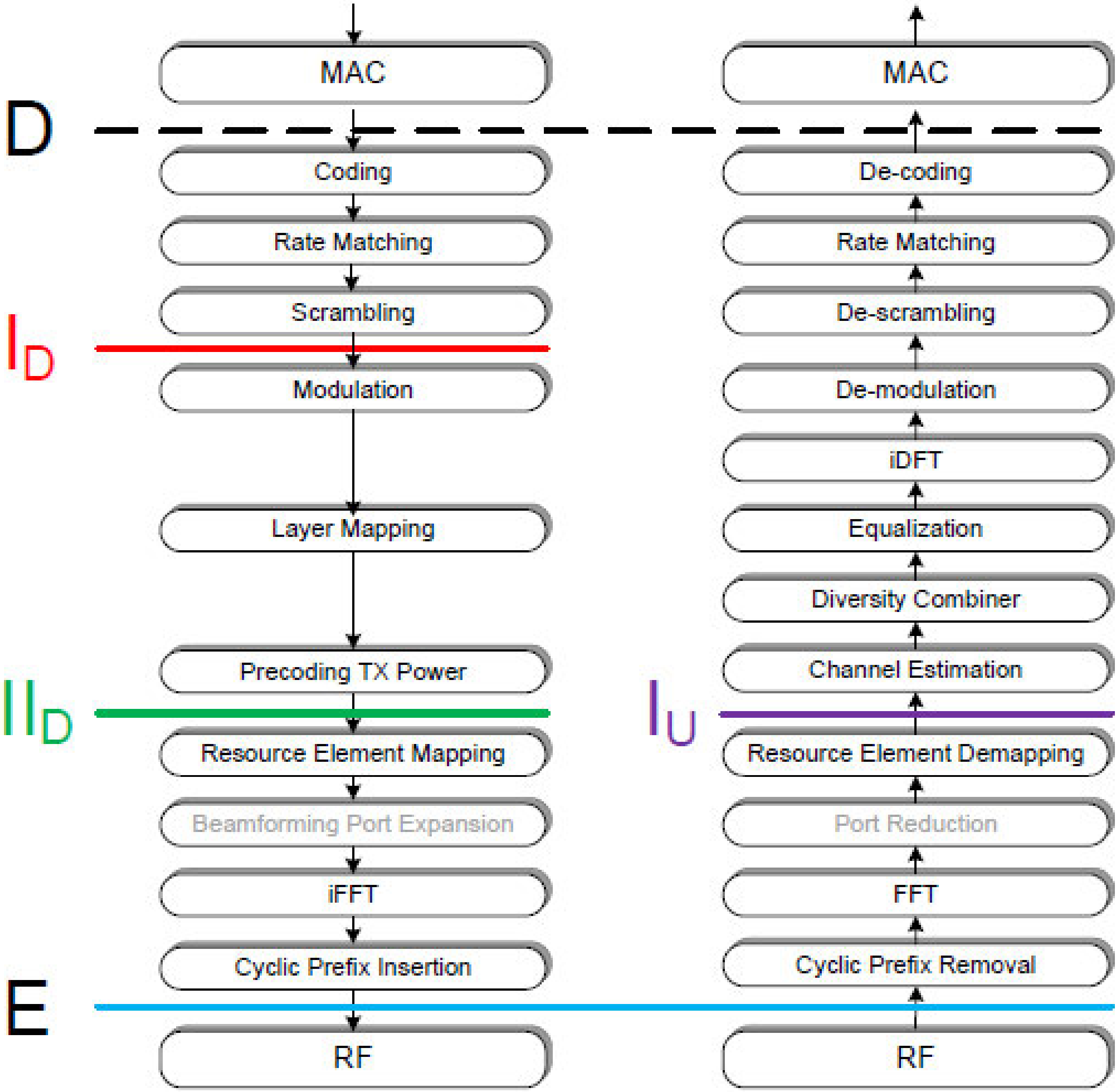


Figure Appendix- 1 CPRI stack and split points

Acronyms

3GPP	3rd Generation Partnership Project
A1	O-RAN interface
AAS	Advanced Antenna Systems
AI	Artificial Intelligence
ANR	Automatic Neighbor Relation
ARIB	The Association of Radio Industries and Businesses, Japan
ARM	processors from ARM Holdings
ASIC	Application Specific Integrated Circuit
ATIS	The Alliance for Telecommunications Industry Solutions, USA
BF	Beamforming
CAM	Cooperative Awareness Messages
CCO	Coverage and Capacity Optimization
CCSA	China Communications Standards Association
CNF	Container Network Function(s)
COTS	Commercial Off-the-Shelf, also Common Off-the-Shelf
CP	Control Plane
CPRI	Common Public Radio Interface
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
CU	Centralized Unit
CU-CP	Centralized Unit-Control Plane
CU-UP	Centralized Unit-User Plane
DARPA	Defense Advanced Research Projects Agency
DL	Downlink
DOCSIS	Data Over Cable Service Interface Specification
DPDK	Data Plane Development Kit
DSP	Digital Signal Processor
DU	Distributed Unit
E1	O-RAN interface: Connection Control Interface between PPF and RCF
E2E	End to End
eASIC	Fabless semiconductor company acquired by Intel in 2018
eCPRI	enhanced Common Public Radio Interface
eMBB	Enhanced Mobile Broadband
EMS	Element Management System in LTE
eNA	Enablers for Network Automation
eNB	see eNodeB
EN-DC	eNB to NR Dual Connectivity
eNodeB	4G LTE Base Station
ETSI	The European Telecommunications Standards Institute
F1	Baseband interface between CU and DU
F1-C	Baseband control-plane interface
F1-U	Baseband user-plane interface
FAPI	Functional Application Platform Interface

FCAPS	Fault-management, Accounting, Performance and Security
FD.IO	Fast Data - Input/Output project
FPGA	Field-programmable Gate Array
FRAND	Fair, reasonable and non-discriminatory licensing
GDPR	General Data Protection Regulation
gNB	5G NR Base Station
GPPP	General Purpose Processing Platforms
GPU	Graphics Processing Unit
HLS	Higher Layer Split
HVAC	Heating, Ventilation and Air Conditioning
ICIC	Inter-Cell Interference Coordination
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet of Things
ISO	International Organization for Standardization
ITS	Intelligent Transport Systems
ITU-T	The Study Groups of ITU's Telecommunication Standardization Sector
JSON/REST	JavaScript Object Notation representational state transfer
KPI	Key Performance Indicator
L1	see PHY
L2	Layer 2 of protocol stack - see MAC
L3	Radio Signaling Layer
Layer 1	see PHY
LCM	Life Cycle Management
LDPC	Low Density Parity Check
LLS	Low Layer Split
LTE	Long Term Evolution (4G)
MAC	Medium Access Control (3GPP NR protocol stack)
MANO	Management and Orchestration
MEC	Mobile Edge Computing
MIMO	Multiple In, Multiple Out
ML	Machine Learning
M-MIMO	massive MIMO
mMTC	massive machine-type-communications
MNO	Mobile Network Operator
M-Plane	Open Fronthaul Management Plane
MRO	Mobility Robustness Optimization
multiRAT	multiple RATs
near-RT	near Real-Time
near-RT RIC	near Real-Time RIC
NEBS	Network Equipment Building System
NETCONF	Network Configuration Protocol
nFAPI	networked FAPI

NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NIC	Network Interface Card
NMS	Network Management System
non-RT RIC	non-Real-Time RIC
NR	5G New Radio, i.e. 5G radio access technology
nRT	near Real-Time
nRT RIC	near real-time RIC
NRT RIC	non real-time RIC
NSA	Non-Stand Alone
O&M	See OAM
O1	O-RAN interface
O2	O-RAN interface
OAI	Open Air Interface
OAM	Operations, Administration and Maintenance
OCP	Open Compute Project
O-CU	open CU
ODP	Open Data Plane project
O-DU	open DU, the virtualization of the RPF
ONAP	Open Networking Automation Platform
OPS-5G	Open, Programmable, Secure 5G
O-RU	O-RAN Radio, Open RAN Remote Unit
OS	operating system, e.g. Cloud OS
OSC	O-RAN Software Community
OTIC	O-RAN Testing and Integration Centers
PCI	Physical Cell Identity
PDCP	Packet Data Convergence Protocol (3GPP NR protocol stack)
PHY	Physical Layer (3GPP NR protocol stack)
PNF	Physical Network Function(s)
POC	Proof of Concept
PON	Passive Optical Network
PPF	Packet Processing Function
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
rApp	An application designed to run on the Non-RT RIC. The A1 & O1 interfaces enable a direct association between the rApp and Near-RT RIC and the RAN functionality.
RAT	Radio Access Technology
RCF	Radio Control Function
RIA	Radio Intelligence and Automation
RIC	Radio Intelligent Controller
RLC	Radio Link Control (3GPP NR protocol stack)
RPC	Remote Procedure Call
RPF	Radio Processing Function

RRC	Radio Resource Control (3GPP NR protocol stack)
RRH	Remove Radio Head
RRM	Radio Resource Management
RRU	Remote Radio Unit
RT	Real Time
RTL	register-transfer levels
RT-RIC	Real-Time RIC
RU	Remote Unit
Rx	Receive
SCF	Small Cell Forum
SDAP	Service Data Adaption Protocol (3GPP NR protocol stack)
SDN	Software Defined Network
SDO	standards development organization
SLA	Service Level Agreement
SON	Self-Optimizing Network
SR-IOV	Single Root Input/Output Virtualization
TCO	Total Cost of Ownership
TIFG	Testing Integration Focus Group
TIP	Telecom Infra Project
TSDSI	Telecommunications Standards Development Society, India
TTA	Telecommunications Technology Association, Korea
TTC	Telecommunication Technology Committee, Japan
TTI	Transmission Time Interval
Tx	Transmit
UAV	Unmanned Aerial Vehicle
UE	User Equipment
UL	Uplink
UP	User Plane
URLLC	Ultra-Reliable Low-Latency Communication
V2X	Communication between vehicles and other devices, Vehicle to Anything
vCU-CP	Virtualized CU-CP
vCU-UP	Virtualized CU-UP
vDU	Virtualized DU
VES	VNF Event Stream
VM	Virtual Machine
VNF	Virtual Network Function(s)
VoLTE	Voice Over NR
VoNR	Voice Over LTE
VPP	Vector Packet Procession (see FD.IO)
VR	Virtual Reality
vRAN	Virtualized RAN
WG	Working Group
x86	Intel processor family

xApps	Independent software plug-in/application to the Near-RT RIC platform to provide functional extensibility to the RAN by third parties. The E2 interface enables a direct association between the xApp and the RAN functionality.
xWDM	wavelength-division multiplexing technology
YANG	Yet Another Next Generation

References

[1] IETF, “The YANG 1.1 Data Modeling Language,” Internet Engineering Task Force, August 2016. [Online]. Available: https://tools.ietf.org/html/rfc7950.

[2] 3GPP, “TR38.801 Study on new radio access technology: Radio access architecture and interfaces,” 03 04 2017. [Online]. Available: https://www.3gpp.org/DynaReport/38801.htm.

[3] Small Cell Forum, “SCF106 Virtualization for small cells: Overview,” 02 06 2015. [Online]. Available: https://scf.io/en/documents/106__Virtualization_for_small_cells_Overview.php.

[4] OpenAirInterface5G, “nFAPI howto,” 12 2017. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/-/wikis/nFAPI-howto.

[5] CISCO GitHub, “open-nFAPI,” 10 August 2018. [Online]. Available: https://github.com/cisco/open-nFAPI.

[6] 3GPP, “RP-180684: Revised SID on CU-DU lower layer split for New Radio,” 04 06 2018. [Online]. Available: https://www.3gpp.org/ftp/TSG_RAN/TSG_RAN/TSGR_80/Docs/RP-180684.zip.

[7] 3GPP, “TR38.913: Study on scenarios and requirements for next generation access technologies,” 19 July 2018. [Online]. Available: https://www.3gpp.org/DynaReport/38913.htm.

[8] 3GPP, “RP-162469: Study on New Radio (NR) Access Technology,” 24 February 2017. [Online]. Available: https://portal.3gpp.org/ngppapp/CreateTDoc.aspx?mode=view&contributionUid=RP-162469.

[9] 3GPP, “WI 750067: New Radio Access Technology,” 26 March 2019. [Online]. Available: https://www.3gpp.org/DynaReport/GanttChart-Level-2.htm#bm750067.

[10] M. R. Warner and R. Burr, “S.3189 - A bill to use proceeds from spectrum auctions to support supply chain innovation and multilateral security.,” 14 January 2020. [Online]. Available: https://www.congress.gov/bill/116th-congress/senate-bill/3189/text. [Accessed 14 January 2020].

[11] Pallone, Guthrie, Matsui and Walden, “H.R.6624 - USA Telecommunications Act,” US Congress, US House of representatives, 04 April 2020. [Online]. Available: https://www.congress.gov/bill/116th-congress/house-bill/6624. [Accessed 04 April 2020].

[12] D. Dr. Jonathan M. Smith, “Open, Programmable, Secure 5G (OPS-5G),” DARPA, 30 January 2020. [Online]. Available: https://www.darpa.mil/program/open-programmable-secure-5g. [Accessed 30 January 2020].

[13] DARPA, “Contract Opportunity: Open Programmable Secure 5G (OPS-5G),” DARPA, 30 January 2020. [Online]. Available: https://beta.sam.gov/opp/c07b83f8b69244f6a78f0d92560690a5/view. [Accessed 16 April 2020].

[14] 5G Americas, “The Evolution of Security in 5G,” July 2019. [Online]. Available: https://www.5gamericas.org/the-evolution-of-security-in-5g-2/.

[15] NTT DOCOMO, INC., “DOCOMO to Commence Deployment of World’s First 4G/5G Multi-vendor Radio Access Network Conforming to O-RAN Specifications,” NTT DOCOMO, INC., 18 September 2019. [Online]. Available: https://www.nttdocomo.co.jp/english/info/media_center/pr/2019/0918_00.html.

[16] O-RAN, “O-RAN Use Cases and Deployment Scenarios WhitePaper,” February 2020. [Online]. Available: https://www.o-ran.org/resources.

[17] ETSI, “ETSI EN 302 637-2: Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service,” April 2019. [Online]. Available: https://www.etsi.org/deliver/etsi_en/302600_302699/30263702/01.04.01_60/en_30263702v010401p.pdf.

[18] 5G Americas, “5G At The Edge,” October 2019. [Online]. Available: https://www.5gamericas.org/5g-at-the-edge/.

[19] 5G Americas, “The 5G Evolution: 3GPP Releases 16-17,” January 2020. [Online]. Available: https://

www.5gamericas.org/5g-evolution-3gpp-releases-16-17/.

[20] 5G Americas, “Precision Planning for 5G Era Networks with Small Cells,” October 2019. [Online]. Available: https://www.5gamericas.org/precision-planning-for-5g-era-networks-with-small-cells/.

[21] 5G Americas, “Management, Orchestration & Automation,” November 2019. [Online]. Available: https://www.5gamericas.org/management-orchestration-automation/.

[22] Small Cell Forum, “SCF233: Small Cell SON and Orchestration from 4G to 5G,” February 2020. [Online]. Available: https://scf.io/en/documents/233_Small_Cell_SON_and_Orchestration_from_4G_to_5G.php.

[23] [Online]. Available: reference www.virtualexhibition.o-ran.org/index.html .

[24] GSMA, “5G-era Mobile Network Cost Evolution,” GSMA, 28 August 2019. [Online]. Available: https://www.gsma.com/futurenetworks/wiki/5g-era-mobile-network-cost-evolution/.

[25] GSMA Intelligence, “Networks unwrapped: green network transformation is no longer just talk,” March 2020. [Online]. Available: https://data.gsmaintelligence.com/research/research/research-2020/networks-unwrapped-green-network-transformation-is-no-longer-just-talk.

[26] ISO, “ISO/IEC TS 27008:2019 Information technology — Security techniques — Guidelines for the assessment of information security controls,” January 2019. [Online]. Available: https://www.iso.org/standard/67397.html.

[27] Mavenir, “MAVENIR AND MTI ANNOUNCE FIRST OpenRAN O-RAN INTERFACE REMOTE RADIO UNIT (RRU) PORTFOLIO,” Mavenir, 26 February 2020. [Online]. Available: https://mavenir.com/press-releases/mavenir-and-mti-announce-first-openran-o-ran-interface/.

[28] C. V. Rajeev Madhavan, “Software eats telco: The coming disruption,” venturebeat.com, 25 February 2020. [Online]. Available: https://venturebeat.com/2020/02/25/software-eats-telco-the-coming-disruption/.

[29] Small Cell Forum, “SCF059: X2 interoperability in multi-vendor X2 HetNets,” 05 06 2014. [Online]. Available: https://scf.io/en/documents/059_-_X2_interoperability_in_multi-vendor_X2_HetNets.php.

[30] ITU-T, “Technical Report,” 9 February 2018. [Online]. Available: https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-HOME-2018-PDF-E.pdf.

Acknowledgments

5G Americas facilitates and advocates for the advancement and transformation of LTE, 5G and beyond throughout the Americas.

5G Americas' Board of Governors members include AT&T, Cable & Wireless, Ciena, Cisco, CommScope, Crown Castle, Ericsson, Intel, Mavenir, Nokia, Qualcomm Incorporated, Samsung, Shaw Communications Inc., T-Mobile USA, Inc., Telefónica, VMWare, and WOM.

5G Americas would like to recognize the significant project leadership and important contributions of group leaders Charlie Michaelis and Paul Smith Jr. of AT&T, Amit Mehrotra of Nokia, and Durga Satapathy of T-Mobile, along with many representatives from member companies on 5G Americas' Board of Governors who participated in the development of this white paper.