HEAVY READING

WHITE PAPER

The Role of the RAN Intelligent Controller in Open RAN Systems

A Heavy Reading white paper produced for Sterlite Technologies Limited

STL

AUTHOR: GABRIEL BROWN, PRINCIPAL ANALYST, HEAVY READING

The RIC comes in two forms, each adapted to specific control loop and latency requirements:

- **The near real-time RIC (near-RT RIC):** Provides programmatic control of open centralized units (O-CUs) and open distributed units (O-DUs) on time cycles of 10ms to 1 second.

- **The non-real-time RIC (non-RT RIC):** Provides higher layer policies that can be implemented in the RAN either via the near-RT RIC or via a direct connection to RAN nodes. Can be used in conjunction with artificial intelligence/machine learning (AI/ML) model training. It is specified for control loops of more than 1 second.

The RIC is more than a RAN controller, however. It is also an open platform that can host RAN control applications developed by specialist software suppliers that are external to the RIC vendor itself. These "xApps" (near-RT) and "rApps" (non-RT) are among the primary reasons why the RIC is a compelling component of the O-RAN architecture. XApps and rApps will enable innovation—in the form of RAN control algorithms—to enter the RAN at a pace that is orders of magnitude faster than is the case with today's vendor-proprietary systems or centralized SON methods. Such innovation will help operators create differentiated network experiences that offer performance adapted to particular service types, user groups, or locations.

Operators that have said publicly that they are running trials of RIC solutions include AT&T, Deutsche Telekom, KDDI, and China Mobile.

O-RAN Alliance and the ONF 5G SD-RAN

There are several organizations involved in O-RAN specification and development. These include the 3GPP, the O-RAN Alliance, the Telecom Infra Project (TIP), the Small Cell Forum, the Open Networking Foundation (ONF), and the OpenAirInterface Software Alliance (OSA). Each of these groups is making a useful contribution to O-RAN development.

The 3GPP is the preeminent RAN standards organization, and this will remain the case. However, where the RIC is concerned, the O-RAN Alliance is the critical organization. The RIC is specified by the alliance as an integral part of the O-RAN architecture and is not (yet) part of the 3GPP RAN standards. The O-RAN architecture dovetails with the 3GPP RAN architecture and, generally speaking, should be seen as additive to existing standards. There are two working groups focused on the RIC in the O-RAN Alliance:

- **WG2:** The Non-real-time RAN Intelligent Controller and A1 Interface Workgroup
- **WG3:** The Near-real-time RIC and E2 Interface Workgroup

The O-RAN Alliance supports the O-RAN Software Community in a cooperation with the Linux Foundation. In the summer of 2020, the alliance also announced it would partner with the ONF to launch the 5G Software-Defined Radio Access Network (5G SD-RAN) project to encourage and facilitate the creation of open source software for mobile 4G and 5G RAN deployments. Initially, the project will "focus on building an open source near-RT RAN Intelligent Controller compatible with the O-RAN architecture." It is backed by major operators and tech companies, including hyperscale public cloud companies.

The existence of the SD-RAN group does not mean that all RICs will be open source—or even that they will be based on the open source code developed by the project group. However, the software community that is expected to grow up around this open source project will influence RIC development—particularly the wider xApps/rApps ecosystem. An example of this is the involvement of the TIP RAN Intelligence and Automation (RIA) subgroup with the 5G SD-RAN program. The RIA aims to develop and deploy AI/ML-based applications (as xApps) for a variety of RAN use cases, including SON, radio resource management (RRM), and massive multiple input and multiple output (MIMO).

What are the RIC use cases?

The RIC customizes RAN functionality by optimizing radio resources according to operator policies. To use an SDN analogy, the RIC is a controller/orchestrator and the base station a type of specialized forwarder. The RIC has an impact on RAN performance in three main areas:

- **Network intelligence:** For example, measuring and reporting how the RAN is performing. This generates data, in standardized formats, that can be analyzed (e.g., using AI/ML techniques) to create new algorithms and polices.

- **Resource assurance:** To ensure that devices/users and services receive the required performance (e.g., via radio link control optimization, handover optimization, or prioritization).

- **Resource control:** To ensure the RAN system is operating efficiently when multiple user groups may be competing for adequate resources.

xApps/rApps and RIC use cases

The services delivered by a RIC are composed of xApps or rApps, or a combination of both. There is no hard limit on the types of xApps or rApps that can be built, and it is anticipated that more than one xApp or rApp will be executed in the RAN at a time. It is also expected that some RIC services will use a combination of xApps and rApps.

Sample xApps/rApps suggested to date include the following:

- Context-based dynamic handover management for vehicle-to-everything (V2X)

- Dynamic radio resource allocation for unmanned aerial vehicles

- Traffic steering

- Quality-of-service/quality-of-experience (QoS/QoE) optimization

- Massive MIMO beamforming optimization

- RAN sharing

- QoS-based resource optimization

- RAN slice service assurance

- Multi-vendor slice performance management

- Dynamic spectrum sharing

- Network slice subnet instance (NSSI) resource allocation optimization

- Local indoor positioning in RAN

Further detail on these can be found in the O-RAN Alliance's ***O-RAN Use Cases and Deployment Scenarios White Paper*** (February 2020), which can be found here: www.o-ran.org/s/O-RAN-Use-Cases-and-Deployment-Scenarios-Whitepaper-February-2020.pdf.

Initial xApps will be focused on health check functions. For example, are the RAN nodes operational and functioning as normal? In a second phase, xApps will extend deeper into observability by gathering more granular data from RAN nodes for analysis. xApps that make close to real-time changes (i.e., less than 1 second time cycles) will come in later phases. In terms of control plane decisions, an initial approach will be to complement the current RRM functions implemented in the CU and DU, wherein the policies from the RIC could change or override the local RRM logic. A more aggressive approach of offloading the RRM function completely to the RIC will be a longer-term event.

Initial rApps, hosted in the non-RT RIC, will initially resemble today's centralized SON (C-SON) applications. They have the potential to evolve quickly as RAN data collection is paired with ML techniques to create algorithms that bring new forms of optimization to the RAN.

THE RIC IN THE MOBILE NETWORK ARCHITECTURE

The RIC is specified in the O-RAN architecture, which is overlaid on the 3GPP RAN architecture. **Figure 1** below shows a simplified view of the combined architecture that highlights the role of the RIC. The 3GPP-defined RAN functions that interface with the RIC—the DU and CU—are shown in dark red. The near-RT RIC, shown in bright red, connects over the O-RAN Alliance's E2 interface to the DU and CU. The non-RT RIC, shown in light red, connects directly to the radio nodes and to the near-RT RIC over the O-RAN Alliance's O1 and A1 interfaces.
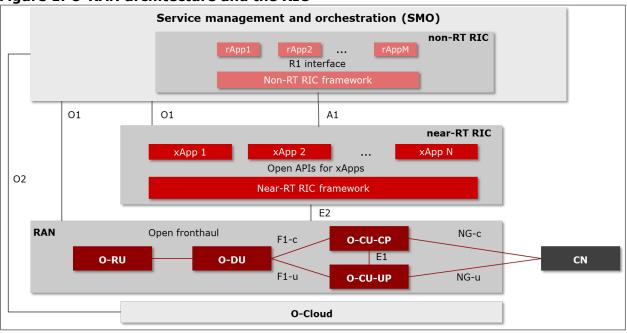
Figure 1: O-RAN architecture and the RIC

Source: Sterlite Technologies Limited

The DU is the baseband unit in a 5G RAN. It performs Layer 1 and 2 processing and executes critical functions such as encoding/decoding, scheduling, MIMO processing, and beamforming. It is the heart of a RAN system and makes sub-millisecond decisions on how to allocate radio resources within a cell based on factors such as user equipment type and distribution, interference conditions, and higher level policies. The near-RT RIC does not take over these real-time RRM functions, which must remain on the DU due to stringent latency requirements. However, what the near-RT RIC can do (in principle, in the future) is programmatically configure the DU to improve how these functions work. For example, the near-RT RIC could be used to change scheduler behavior on the DU.

The CU is a new node in 5G that does not exist as a discrete function in 4G. In the 3GPP RAN architecture, the CU provides Layer 3 functions such as session management and mobility management. It is split into the CU-UP for user plane processing and the CU-CP for control plane processing. The near-RT RIC again interfaces with the CU-UP and CU-CP over the E2 interface to deploy a policy or apply dynamic control. For instance, in a scenario where devices have high speed mobility requirements (e.g., motor vehicles), a different handover management algorithm may be superior.
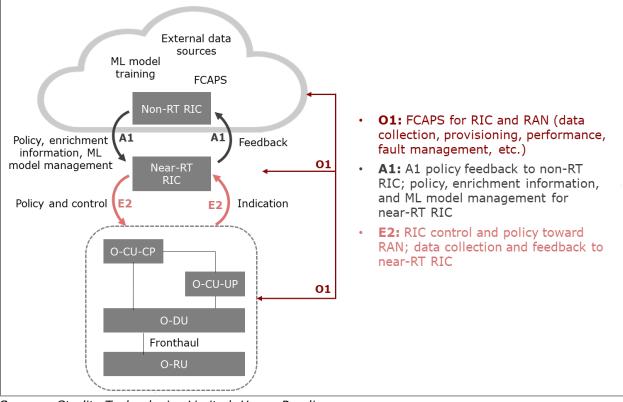
In this architecture, the RIC is therefore largely additive and complementary to the existing 3GPP-defined architecture. By bringing new capabilities, it can enhance how these RAN nodes perform and, as an open platform, it can attract software innovators to the radio domain. There may be instances, particularly with incumbent 5G vendors, where the new control loops introduced by the RIC conflict with their own procedures for updating policies on the DU and CU functions. This should be resolvable, and it is hoped that incumbent vendors will also support the E1 and O1 interfaces.

A near-RT and non-RT RIC model

The near-RT RIC interfaces directly with the radio nodes, operates on sub-second time scales (typically 100s of milliseconds), and hosts xApps. It is therefore sometimes considered the more important of the two RIC nodes. However, near-RT and non-RT RICs work together in a hierarchical model, as shown in **Figure 2**.

Figure 2: Interworking RAN and RIC nodes

Sources: Sterlite Technologies Limited, Heavy Reading

The non-RT RIC is used for tasks with longer time cycles, including functions traditionally associated with C-SON, as well as emerging ML approaches to RAN optimization. It connects to the RAN over the O1 interface to collect data (in standardized formats) and to deploy policies (in the form of rApps) where longer time cycles are appropriate. The non-RT RIC can also deploy policies over the A1 interface to inform the behavior of xApps in the near-RT RIC, which then are deployed into the RAN over the E2 interface.

With its placement higher up in the architecture, the non-RT RIC can access larger RAN datasets, generated over longer time periods, to gain deeper insight into performance and identify potential optimizations that are not apparent in sub-second processing. The non-RT RIC can interface with other network data sources (e.g., to co-optimize radio, IP networking, and edge cloud performance). It can also access datasets external to the networks itself, such as those related to traffic, emergency services, weather, mass public events, etc., so the network can prepare or respond to real world events. Traditional C-SON can do some of this, but the process is typically reactive, and it can be slower to implement changes than the new RIC model.

The non-RT RIC is particularly interesting in the area of ML for RAN optimization. RANs generate vast amounts of data that can be processed to provide insight into network behavior. This data is classically analyzed offline using RAN analysis tools, with the results used to inform how a network could be better optimized. The use of ML to analyze this data and identify improvements is showing great promise. The non-RT RIC could become the platform on which new policies, developed using ML insight, can be created and hosted prior to being pushed down into the network. This may result in faster training of models that are specific to the network in question and faster implementation in the network than is currently achieved with C-SON platforms.

The importance of the E2 interface

The E2 interface has a lot of work to do in the O-RAN architecture. It is the critical interface between the near-RT RIC and radio nodes (radio nodes are sometimes called E2 nodes by RIC developers), and it is imperative that this interface is unambiguously open. A key test of O-RAN will be the how well E2 is supported by RAN equipment companies.

The "E2 manager" is an integrated part of the near-RT RIC platform and is sometimes referred to as "xApp zero." It is used to initiate E2 connections with the RAN nodes and then store RAN configuration information learned at connection setup (and kept updated over time). The first release of the RIC specifications is now available, and O-RAN Work Group 3 will continue to be developed to add functionality; for example, E2SM (E2 Service Model) specifications are now in development.

Two key work items currently underway are to standardize E2SMs that enable traffic steering and QoS/QoE optimization through the RIC. Traffic steering targets idle mode mobility load balancing (MLB), inter-intra frequency MLB, carrier aggregation, and dual connectivity. QoS/QoE optimization allows the RIC to control network functions related to QoS control, radio resource allocation, radio access control, connection management, and mobility functions.

It is anticipated that operators will deploy RICs on a one-to-many basis, with each RIC connecting to multiple O-DUs and O-CUs. The E2 must therefore support failure handling for improved resiliency. However, according to the specification, RAN nodes must also be able to function normally independently of the RIC in case the E2 interface and/or the RIC fails.

RIC DEPLOYMENT ARCHITECTURE

The 5G RAN is hierarchical. Multiple RUs connect to one (or more) DUs, multiple DUs connect to a CU, and multiple CUs connect to the core network. At each stage, traffic is aggregated and control is centralized. The RIC will be deployed at some point between the cell site and the core network. The O-RAN Alliance Working Group 6 (WG6), which is focused on cloud architecture and deployment scenarios for virtualized O-RAN, distinguishes between the core cloud, the regional cloud, and the edge cloud at the cell site. The RIC will be deployed in one of these locations.

If the nature of the use case is not latency sensitive, it is possible to deploy the RIC in the central cloud at same locations as core network functions are deployed today. A deployment of this type would be similar to a C-SON deployment and would be capable of making changes on longer time cycles, such as performed by the non-RT RIC. The core cloud would not, however, be suitable for xApps that target near-real-time tasks.

In most cases, it would also not make sense to deploy the near-RT RIC at the cell site because the value of this function is that it coordinates across multiple radio nodes to improve system performance. An exception to this would be private network deployments, or venue networks, where a RIC could be deployed onsite to monitor and control the local area RAN.

Regional and edge cloud deployment

The right place for near-RT RIC deployment, in most cases, will be the regional cloud (to use WG6 terminology). WG6 identifies several O-RAN deployment scenarios. From a RIC perspective, the following three appear most likely:
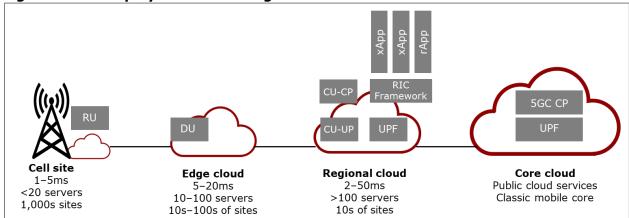
- **Scenario A:** The near-RT RIC, O-CU, and O-DU functions are all virtualized on the same edge cloud platform, and the interfaces between those functions are within the platform.

- **Scenario B:** The near-RT RIC is deployed on a regional cloud platform, while the O-CU and O-DU functions are virtualized on an edge cloud platform.

- **Scenario C:** The near-RT RIC and O-CU are deployed on a regional cloud platform, while the O-DU is virtualized on an edge cloud platform.

Scenario C is particularly interesting because the near-RT RIC covers a geographical area that is comparable to the number of sites that would be aggregated on a CU node. It also has the same kind of latency requirements as the CU in relation to the DU. The CU-CP, for example, carries out tasks such radio link control and handover messaging that can tolerate milliseconds of delay, while the CU-UP processes user plane traffic and forwards it into the IP layer. It both cases, it makes sense to deploy these nodes at locations where it is possible to aggregate traffic from relatively large areas (e.g., aggregating 10s–100s of cell sites).

Figure 3 below shows that CU placement and RIC placement are topologically aligned in Scenario C. In this case, the near-RT and non-RT RICs are deployed on the same regional cloud cluster as several other functions, including the CU and perhaps user plane functions (UPFs), SDN controllers, security applications, and perhaps even some end-user applications. This type of deployment offers operators the potential to integrate RAN, packet networking, cloud, and end-user applications into a high performance, efficient end-user service.

Figure 3: RIC deployment at the regional cloud location

Source: Heavy Reading

Figure 4 shows more detail on the RIC deployment scenarios and maps logical functions to central, regional, and edge cloud locations according to operator requirements.
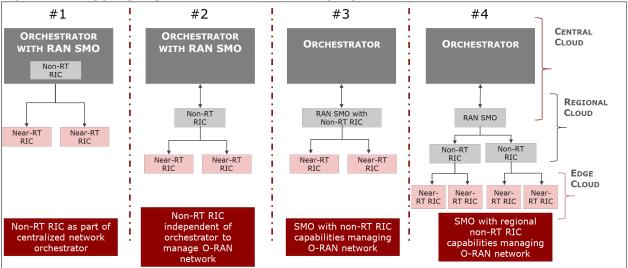
Figure 4: Mapping logical architecture to deployment scenarios

Source: Sterlite Technologies Limited

- **Deployment #1:** The non-RT RIC is integrated into the service management and orchestration (SMO) framework, which addresses the overall operation and maintenance of O-RAN networks. The near-RT RIC is distributed to the regional cloud. This is likely to be a common scenario in the initial phases of RIC-driven RAN operations.

- **Deployment #2:** In this case, the non-RT RIC is extracted from the centralized SMO and deployed in the regional cloud, as in **Figure 3**, alongside the near-RT RIC. This is also likely to be a common deployment choice.

- **Deployment #3:** In this case, the RAN-specific SMO framework is also distributed to a regional cloud, where it is hosted alongside the RIC functions. This is more likely in larger networks and/or when O-RAN networks are deployed at larger scale.

- **Deployment #4:** The most distributed model in which near-RT RIC is deployed to the edge cloud (i.e., the cell site of baseband hub in a C-RAN deployment). This appears to be most suited to larger networks or venue networks.

SUMMARY AND NEXT STEPS

O-RAN is a new way to build mobile RANs based on open interfaces, modularization, off-the-shelf hardware, and innovation in software. Operator trials are underway in all major regions, with the first commercial networks to be deployed and put into service from 2021 onwards. Heavy Reading believes there is potential for O-RAN deployments to scale as performance is shown to be competitive and as integration of O-RAN into operator planning and management tools improves.

Software control of the RAN is a key breakthrough of O-RAN. Programmatic control of the network enables operators to adapt their networks for different service types and user groups. This, in turn, introduces software-driven operating models to a part of the network that is classically hardware-oriented and where change occurs over long time cycles. Programmatic control is vital to automation and therefore significantly affects the cost and efficiency of the network.

The key messages of this paper are the following:

- The RIC is a core component of the O-RAN architecture; it is what makes O-RAN dynamic and software-centric, which ultimately enables higher performance 5G RANs.

- The RIC is more than a RAN controller. It is also an open platform that can host xApps and rApps developed by specialist software suppliers. This has the potential to stimulate investment RAN control applications and make it faster to deploy innovation into the live network.

- It is early days, however. The baseline specifications are now available and there is a software community growing up around the RIC, including open source development. The first RIC use cases are emerging focused on reporting and observability.

- Heavy Reading expects the RIC and xApps/rApps to become progressively more sophisticated and, in later phases, able to make active changes to live networks. This will require careful testing and monitoring to ensure the stability and reliability of the RAN nodes when under the control of the RIC. This is a big change in how RANs operate, and it will take time for the industry to be comfortable with this.

- In terms of deployment, a regional cloud model is probably optimal in most cases. The near-RT RIC covers a geographical area similar to that served by a regional cloud point-of-presence. This deployment offers the potential to co-optimize the RAN, core, transport, and cloud to deliver superior end-user services than would otherwise be possible.

ABOUT STL – STERLITE TECHNOLOGIES LIMITED

STL is an industry-leading integrator of digital networks.

We design and integrate these digital networks for our customers. With core capabilities in Optical Interconnect, Virtualised Access Solutions, Network Software, and System Integration, we are the industry's leading end-to-end solutions provider for global digital networks. We partner with global telecom companies, cloud companies, citizen networks, and large enterprises to deliver solutions for their fixed and wireless networks for current and future needs.

We believe in harnessing technology to create a world with next-generation connected experiences that transform everyday living. With an intense focus on end-to-end network solutions development, we conduct fundamental research in next-generation network applications at our Centre of Excellence. STL has a strong global presence with next-gen optical preform, fibre, and cable manufacturing facilities in India, Italy, China, and Brazil, along with two software development centres across India and one data centre design facility in the UK.

To realize the gigabit connectivity for all, STL has been focusing more on the access part of the network infrastructure. We are offering futuristic 5G and FTTx solutions that are based on the concept of openness, disaggregation, and virtualization. In the wired connectivity realm, we are offering a platform that is entirely virtualized OLT and ONT software stack integrated with white box solutions to support fiber to the home or enterprise. On the wireless side, we have strategically invested in and partnered with leading companies in the virtualization domain to offer a COTS-based vRAN solution that can host both 4G and 5G concurrently, as well as radio units for different deployment configurations. This includes small cells and macro RUs to meet various deployment archetypes that operators have for both 4G and 5G. To complement these, we are also offering software platforms such as the RAN intelligent controller (RIC) and orchestrator—essential platforms for a cloud centric solution.

With expertise in all these capabilities, STL is well positioned to address the needs of telcos and private enterprises through our Access Solutions portfolio offerings, which are fundamentally open, disaggregated, and virtualized platforms. To learn more about them in detail, you may visit https://www.stl.tech/virtualised-access-products.