A whitepaper

RAN Intelligent Controller (RIC)

A Key Enabler of Total Cost of Ownership (TCO) Reduction for Mobile Network Operators

Introduction

Mobile network operators (MNOs) throughout the world are experiencing an exponential growth of data traffic, mainly driven by an increase in the demand for video streaming services, peer to peer services such as video conferencing, OTT applications, as well as by tremendous proliferation of mobile devices. A recent report from Analysis Mason indicates that between 2021 and 2025 alone, mobile traffic will grow a further 250%. Conventionally, operators tackle the traffic growth and maintain the desired quality of service by acquiring more spectrum, adding carriers or by cell splitting. Since the new spectrum that is being awarded to MNOs are typically at higher frequency bands, their propagation characteristics often require more sites than usual to cover a given geographical area.

Furthermore, it is important to note that the revenue growth from mobile services is not commensurate with the traffic growth. A recent white paper from CMI Corp/VMWare shows that profit per bit for MNOs has been steadily eroding for the past decade, and the trend continues. This means that there is an unprecedented need for MNOs to keep pushing their TCO (total cost of ownership) down as they look for opportunities for revenue growth.

In this paper, we touch upon the TCO reduction aspects of MNOs and how in the O-RAN architecture, RIC (RAN Intelligent Controller) in particular would be helpful in that context.

A rough pictorial representation of the TCO, for a typical MNO is depicted in Figure 1.

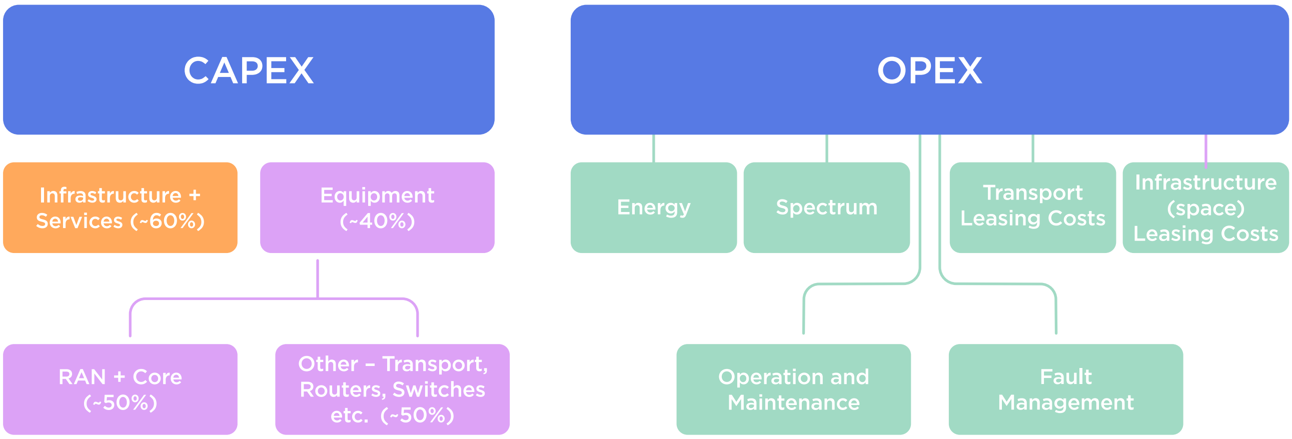

Figure 1 – Major TCO Components of a typical MNO

These are major components of TCO, at a high level. For OPEX, the cost breakdown greatly depends on the geography, operating practices and environment (including regulatory aspects) – hence making an average percentage breakdown estimate may not make much sense.

When it comes to TCO reduction, we need to consider both CAPEX and OPEX reductions. For subsequent discussions, we will concentrate on the radio access network (RAN) only as it constitutes the majority of the mobile network cost. CAPEX reduction could be achieved by improving the spectral efficiency in general – that is making the best use of the deployed capacity such that the MNO can push the deployment of new capacity and/or the need for new spectrum as farther in time as possible. Self Organizing Network (SON) is a mechanism through which operators have traditionally looked to reduce CAPEX and OPEX by enabling automation to increase operational efficiencies of the network. However, SON's effectiveness is limited by the longer response times and the lack of agility in decision-making.

OPEX reduction could be achieved by automating the network deployment and operations wherever possible, arguably with the help of Artificial Intelligent/Machine Learning (AI/ML) techniques, as we can collect and infer from an abundance of data from the RAN. The RIC has emerged to be an efficient platform to achieve these goals.

Open RAN (and O-RAN Alliance) architecture allows for disaggregation, virtualization, and open interfaces between network elements, in addition to driving a cloud-native approach to telecom networks. More importantly, O-RAN architecture defines the RAN Intelligent Controller (RIC) to provide the operator the flexibility and the agility to implement a myriad of use cases and applications including:

- Automate network management and performance
- Facilitate efficient utilization of radio resources
- Enable slicing of logical and physical network resources to support different service level agreements (SLAs)

In other words the RIC is being defined in O-RAN to leverage the open interfaces to drive down TCO.

Not surprisingly, a very recent survey by Analysis Mason shows that MNOs chose RAN TCO reduction as their top commercial driver for deploying the RIC. Analysis Mason also has forecast that network OPEX worldwide will decline at a CAGR of just 0.94% during 2019–2026, which suggests that MNOs must do more to reduce TCO, and the RIC is part of that strategy.

RIC – Promoting an ecosystem for TCO reduction

Traditionally, mobile networks have been built on proprietary vendor technologies. The O-RAN Alliance is poised to change that, bringing multiple vendor options for equipment and services. The new service options, including vendor diversity as well as net new services come from RIC, which is the most radical concept in the O-RAN architecture.

Albeit RIC sounds like a RAN controller, we need to consider it as an open platform where a whole host of third-party developers can create applications for various use cases and then integrate them with a standardized "open" RIC platform. Along this line, RIC presents a great vehicle to achieve the TCO reduction goals efficiently, with best of breed solutions that are closely coordinated and without any vendor lock-ins.

Note that this complements Mobile Edge Cloud (MEC) which enables the development of user plane applications, whereas RIC enables control plane applications.

Fundamentally, RIC enables control loop implementations where RAN events of interest are observed, then analyzed, and subsequently controlled based on certain policies/goals in a closed loop manner – with an option to use AI/ML techniques. The RIC has two flavors, each architected to satisfy specific control loop latency requirements:

- **Near-real-time (Near-RT) RIC:** Carves out some of the radio resource management control loops from the gNodeB to an open environment where they can be customized. Applications (xApps) in the Near-RT RIC are to be designed to respond to events from the RAN in near-real-time (10 ms to 1 second).
- **Non-real-time (Non-RT) RIC:** Supports AI/ML approaches for automation of RAN operations, non-real-time intelligent radio resource management, policy-based optimization of RAN parameters and resources, and provides guidance, parameters, and policies for near-real-time RAN control and operations (via Near-real-time RIC) or by other means (for brownfield). Applications (rApps) in Non-RT RIC support control loops of more than 1 second.

  Figure 2 illustrates this.

A detailed discourse on how O-RAN architecture achieves its goal is beyond the scope of this discussion. However, Figure 2 is a depiction of how the O-RAN compliant architecture RIC which STL is promoting as a service enabling platform helping to create an ecosystem for developing, deploying, and operating third party applications (xApps for Near-RT applications and rApps for Non-RT applications) that would enable TCO reduction. This is one of the key missions of STL Access Solutions.

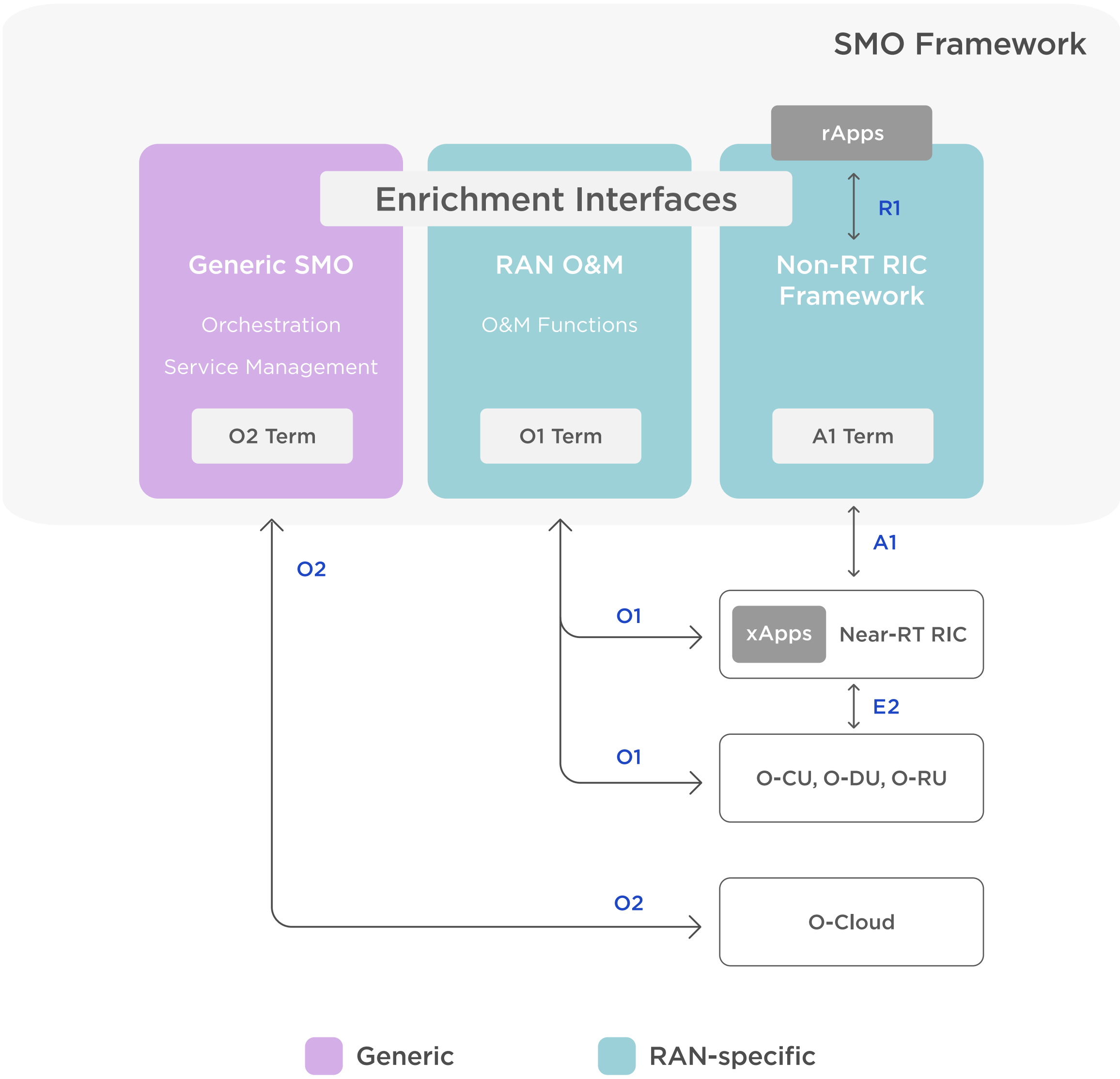

Figure 2 – O-RAN Architecture and RIC: (STL)

One important point to note here is the logical decoupling of Non-RT RIC and RAN O&M from Generic SMO functions. This makes sense as it provides more flexibility for mobile network operators, without violating O-RAN principles. We are working with partners to urge O-RAN to consider defining details of some of the less-specified interfaces across Non-RT RIC and SMO components, as it will greatly help the ecosystem. We are quite pleased to see that the initial reception of this proposal has been very encouraging.

Figure 3 illustrates STL Access Solutions RIC portfolio, where incumbent vendors as well as entrepreneurs can create intelligent RAN applications and services. Utmost care has been taken to make the application development, testing and onboarding process is as simple and straightforward as possible, facilitated by a sandbox environment and synthetic data, enabling close to real-life experience.

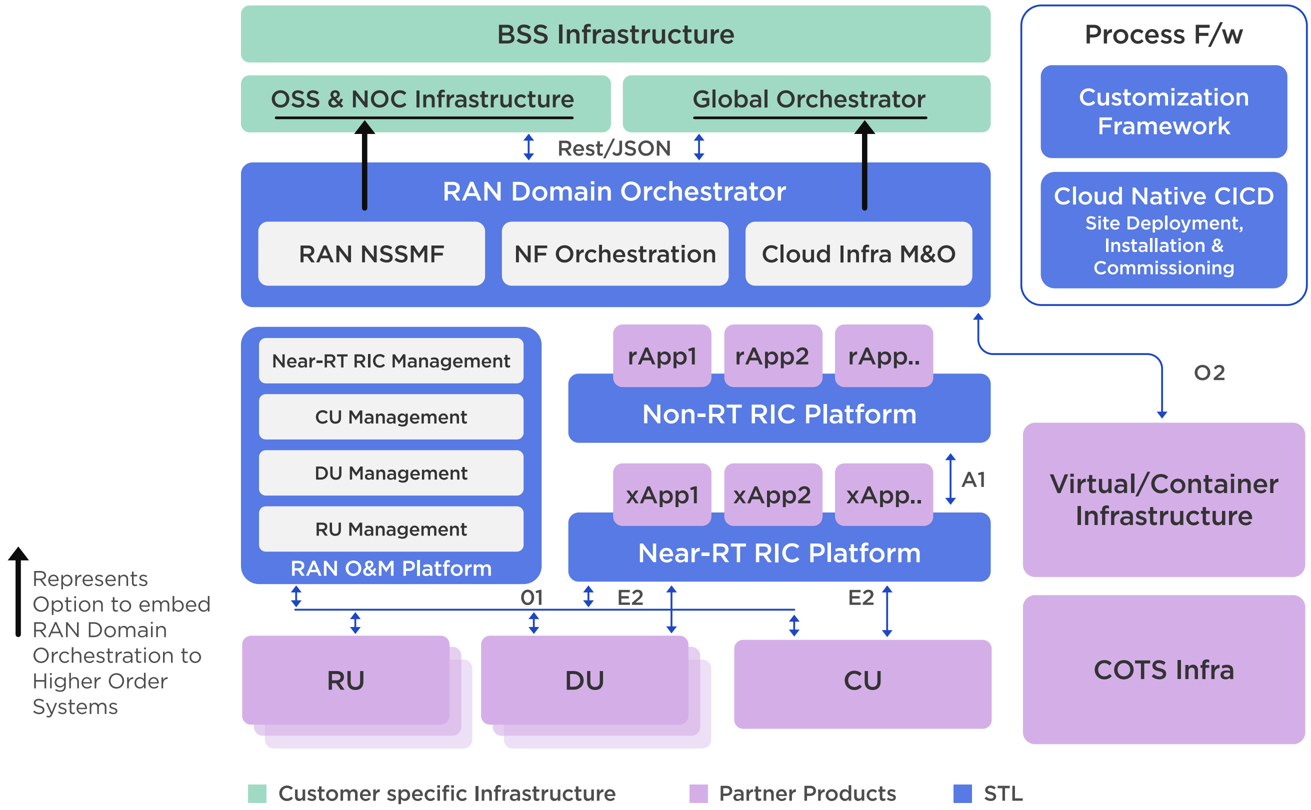

Figure 3 – STL RIC – Creating a RAN service enabling ecosystem

It is natural that full O-RAN architecture would take time to get adopted universally. STL Access Solutions is fully sensitive to this important aspect and we have incorporated that flexibility into our RIC portfolio architecture. There are several TCO reduction applications that do not need CU/DU control or E2 interface support, and MNOs can start using the RIC platform immediately to implement those applications, and progress towards tapping into the full benefits of Open RAN as the adoption level increases. Basically, the intent is to leverage the flexibility and agility of RIC to deliver TCO benefits, even as standardization and adoption are in progress.

RIC – Use Cases for TCO reduction

One can identify dozens of use cases that can be implemented via RIC (especially Non-RT RIC) for TCO reduction, which would need different levels of O-RAN adoption. Also, it is very important to note that TCO reduction applications using RIC can be applied to both 4G and 5G networks – hence MNOs need not wait for 5G rollout or O-RAN adoption to start reaping RIC benefits. One immediate question that would come to MNOs is that how this relates to SON and associated platforms. On this front, RIC is eliminating proprietary SON platforms and associated single vendor solutions.

With RIC, MNOs can implement best of the breed SON like functions at competitive prices, all coordinated in a policy driven environment. In other words, multi-vendor SON solutions can cordially exist and operate on a RIC platform. Now the question is, what does an MNO do with its current SON platform and applications on it? One can address this as follows, without "ripping and replacing" any of the existing software platforms or systems.

- Make RIC the master automation platform and implement all new SON applications on RIC, on-going.
- Interface RIC with the existing SON platform especially with its application coordinating module (which SON platforms generally have to avoid action conflicts between applications).
- Perform all application level coordination from RIC using its policy enforcement module.

Note that RIC is more than a SON platform, however. It can be thought of as a powerful automation platform where AI/ML capabilities can be effectively incorporated. In that respect, RIC supports way more automation use cases, over and beyond the purview of conventional SON.

Though not within the scope of this paper, in addition to RIC-enabled TCO reduction use cases, there is another category of RIC use cases that would foster new revenue generation under the auspices of 5G, by exploiting the full flexibility and power of O-RAN architecture.

Figure 4 is an attempt to capture the above points.

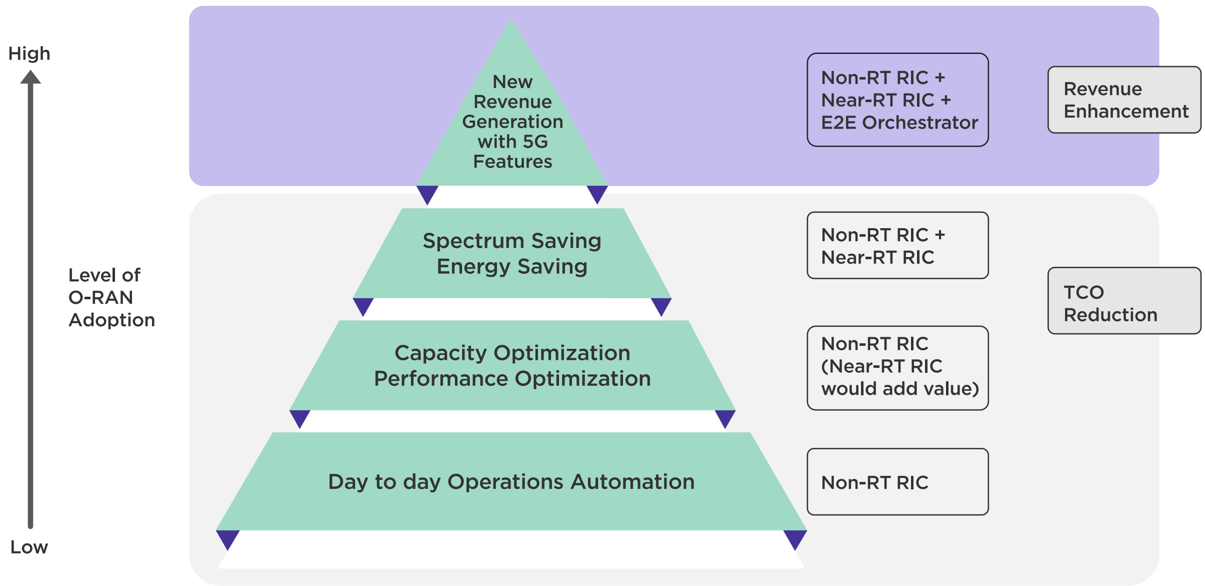

Figure 4 – RIC use cases: A high level classification and calibration with O-RAN adoption

RIC Use cases - Examples of TCO reduction

a) Automating new RF Carrier/Site launches

A very pertinent use case under day-to-day operations is automating the process of adding new carriers (or sites) to the network as many MNOs are in the process of rolling out 5G, and containing the cost of rollout is very important.

Deployment of 5G mobile network infrastructure in a business-as-usual approach would be more expensive and complex as compared to prior mobile technologies. For 5G, the need for using higher radio frequency bands, densification, as well as the proliferation of endpoints would tend to push deployment costs higher.

Several network validation tests need to be conducted after vendors have completed the construction and installation work at a cell site, but before the site is handed over to the Telco's operations team for commercial launch. This is applicable for both new site construction as well as new RF carrier additions on existing sites. Since most of the 5G deployments are not greenfield, the latter is more common.

Needless to say, each MNO has its own specific methods and procedures before a site is accepted for commercial operation. However, at a very high level, it is captured as a five-step process in Figure 5, once the new site (or added carrier) is powered up and running, and all essential connectivity tests are completed.

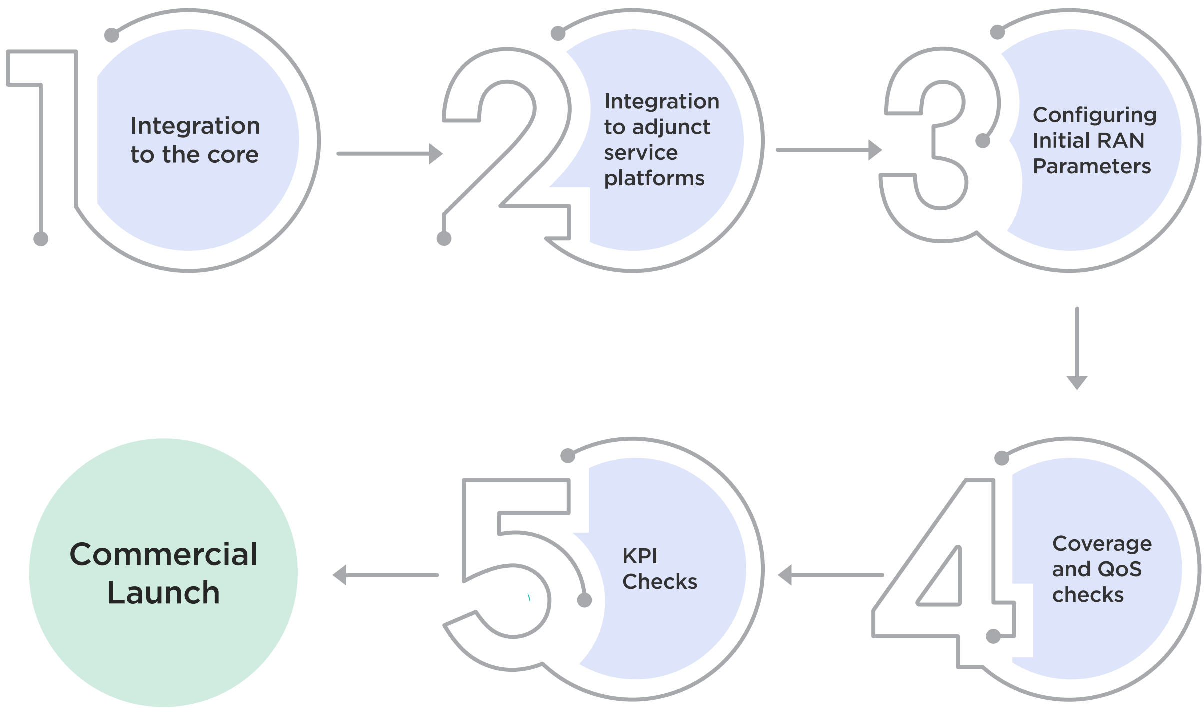

Figure 5 – High level process of new site/RF Carrier launch

With the help of AI/ML and applications on Non-RT RIC, the entire process can be automated, assisted by MDT (minimization of drive test – a standardized mechanism introduced in 3GPP Release 10).

New carrier/site addition Opex costs vary throughout the world, but could add up to tens of millions of dollars depending on the size of the MNO. As a matter of fact, Step #4 is done periodically even after a carrier is commercially launched to ensure coverage and quality, by drive tests. Last year a large US carrier reported that to drive test just 25% of the square kilometers of its nationwide 4G LTE coverage would cost approximately $45 million each year and that drive testing only 10% of its coverage would still cost as much as $18 million/year This is also a candidate for savings as MDT, combined with the closed loop automation using RIC-based applications, can achieve the goals of drive test based coverage and performance optimization.

b) Load Balancing and Traffic offloading (Traffic Steering)

In general, it is well understood that automation around load balancing (among carriers in the same sector) and traffic off-loading (moving traffic between cells on different sites) would yield significant efficiency improvements in making optimal use of available RAN capacity. This has been always in the conventional SON territory, under the traffic steering category of use cases. However, the combined efficacy of various traffic steering approaches would depend on the following:

- How quickly we can detect a traffic imbalance condition and an overload condition?
- Furthermore, can we proactively predict these conditions and invoke control actions?
- How effectively we can coordinate the control actions, in a mutually aiding manner?

The policy-driven, low latency control loop framework of RIC (Non-RT RIC with or without Near-RT RIC) assisted by AI/ML would be an excellent platform for these sorts of automation, to extract the full benefits of traffic steering approaches.

Figure 6 shows where the capital efficiencies come from traffic steering and how that may be quantified.

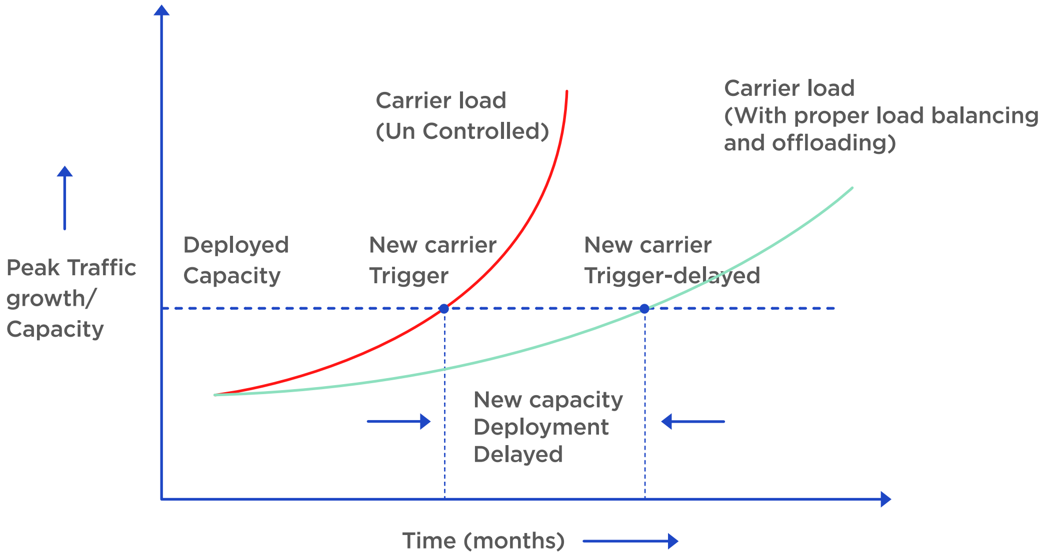

Figure 6 – Automation to improve traffic handling efficiency delays the need for new RAN capacity

In Figure 6, the red curve shows the "organic" peak traffic growth on a cell without any effective traffic steering. Note that RAN capacity is engineered to handle peak loads. The green curve is with traffic steering, which slows down the growth of the peak traffic carried by the cell. When the traffic exceeds the capacity, a new capacity requirement is triggered needing more carrier or site addition and/or more spectrum (more capital deployment) – in the latter (green) case, the triggering is pushed out in time.

As an illustration, if we assume a 10% traffic growth, a modest traffic handling efficiency gain of 10% would delay the new carrier/site trigger by about a year (detailed computations are omitted here). This means the MNO can delay new capital infusion by a year and the cost of that capital for a year is saved. When we consider hundreds of thousands of cells in the network, the savings could add up to a very large number.

c) Energy Saving

Energy costs for telecom operators can be as high as 10% of the the total operating expenditure. This is especially true as 5G is being rolled out by operators. Even though the 5G-NR standards are more energy efficient than its predessors (for example, energy consumption per unit of data (watt/bit) is much less for 5G than 4G) the overall power consumption of 5G base stations can be very high. For example, a report from Huawei indicates that 5G base stations could consume about 1.7 times more energy than their 4G counterpart.

Also, due to propagation charateristics, the new spectrum bands that are used to rollout 5G will require many more mobile sites. As a matter of fact, 5G highly relies upon RAN densification. Along with that, with mobile edge computing, the number of data centers would rise and computationally intense use cases would further intensify energy usages. Thus, overall the energy consumption in mobile networks is expected to go up significantly as 5G is rolled out. In addition to cost concerns, many operators are often required to reduce their carbon footprint as telecom operators already account for 2 to 3 percent of total global energy demand, as per a recent McKinsey report

All these aspects suggest that 5G operators have a huge responsibility on their shoulders to deploy capabilities in their network to optimize energy consumption.

It is also to be noted that only less than 20% of the energy is used for information processing and movement in mobile networks. About 80% of the energy is spent as heat loss in power amplifiers, power converters (AC to DC), HVAC, and battery units. This is especially a waste when the network is under lighter loads, when we can afford to shut down some of the capacity without any performance degradation, by which energy can be conserved.

If one looks at the daily traffic patterns handled by base stations for a typical operator (which can be obtained from multiple publications), it is fairly clear that about a third of the RAN capacity (in this context could be measured as cell-hours, which is the total number of live cells multiplied by 24 – total hours in a day) can be put to sleep at low load hours while ensuring sufficient coverage and capacity such that service quality is not compromised. Note that the RAN capacity is engineered to handle the peak loads (busy hour/busy day load in busy season).

To further look into this, we may classify cells into two categories – coverage cells and capacity cells. We keep coverage cells on all the time, ensuring service availability. We then selectively put capacity cells to sleep or wake them up, based on traffic.

Figure 7 illustrates the overall concept.

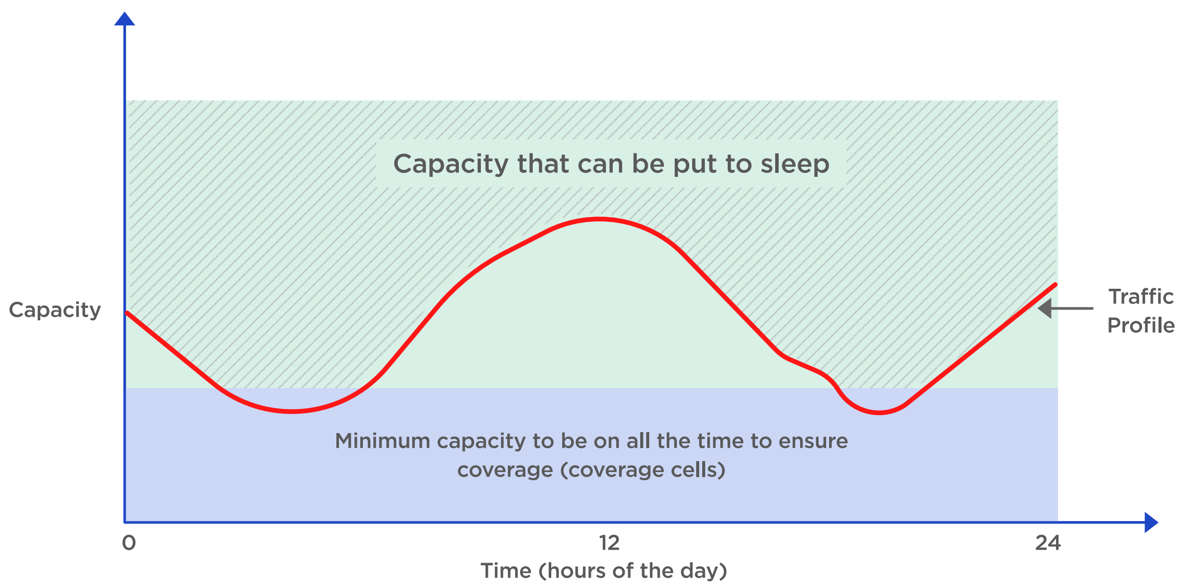

Figure 7 - Energy saving opportunity by allowing cells to go to sleep under low loads

In figure 7, the capacity marked in blue to be on all the time to ensure coverage. The capacity marked green represents the capacity of candidate cells that can be put to sleep, based on the load at a given time. The red line shows the load profile and the shaded area above the red line in the green area represents the capacity that can be put to sleep at various times; hence represents energy saving. Operators often implement overlay/underlay carrier architecture. In that case, they can keep the overlay layer (presumably lower frequency bands – represented by the blue area) on air throughout for coverage and turn off the capacity cells in underlay layer in the shaded area (e.g., n78 band) as appropriate when they are unused.

As a matter of fact, some equipment vendors offer DSON (distribute SON) solutions for LTE and but many operators have been wary of the limitations such solutions pose. Also, augmenting AI capabilities to those solutions have been generally challenging. A detailed discussion on this topic is beyond the scope of this paper, but it is important to note that direct translation of current LTE solutions of energy saving to 5G is not optimal, in general. On the other hand, it is a prime candidate for RIC (Non-RT RIC assisted by Near-RT RIC) application due to the following reasons:

1. RIC has visibility to multiple radio units, hence holistic sleep/wake-up decisions can be made.
2. Policy driven sleep/wake-up strategy makes the management of this capability easier and efficient. Operators can quickly change sleep strategy without changing any software – just need to change the policy specifications.
3. RIC construct provides flexibility and control to ensure that no performance degradation happens.
4. Non-RT RIC is a very effective platform for AI/ML based sleep policy creation. The sleep strategy could be automatically adjusted without any human intervention, in an adaptive mode.
5. Coordination of other automation capabilities can be easily done as policy coordination and conflict resolution. For example, dynamic spectrum sharing (DSS) and energy saving have conflicting goals and coordination between them is well tractable in the policy driven RIC environment.

We can discuss many such TCO reducing applications that can be effectively implemented using RIC.

In summary, RIC provides an excellent platform for MNOs to create applications in a multi-vendor ecosystem for their TCO reduction. STL's modular and fully O-RAN compliant architecture enables reaping the TCO reduction benefits of RIC for brownfield and greenfield operators with varying degrees of O-RAN adoption.

For more information, please contact **Sarat Puthenpura at sarat.puthenpura@stl.tech**

References:

a) https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/microsites/telco/ vmware-transformation-and-5g-o-ran.pdf

b) https://www.lightreading.com/test-and-measurement/verizon-t-mobile-atandt-balk-at-drive- testing-their-networks/d/d-id/763329

c) https://www.mckinsey.com/industries/technology-media-and-telecommunications/our- insights/the-case-for-committing-to-greener-telecom-networks

d) https://www.stl.tech/virtualised-access-products/ric/

e) https://www.fiercewireless.com/tech/5g-base-stations-use-a-lot-more-energy-than-4g-base-stations-saysmtn#:~:text=shift%20to%205G.-,A%20typical%205G%20base%20station%20consumes%20up%20to%20twice%20or,a%20denser%20layer%20of%20small

f) https://www.analysysmason.com/contentassets/62ee68d6f8fd4c91b7318e97036382c9/analysys-mason_near_rt_ric_jun2021_rma07.pdf

Authors

Sarat Puthenpura

Chief Architect – Wireless Access Solutions, STL

sarat.puthenpura@stl.tech

Sarat Puthenpura drives innovative solutions to realize full potentials of wireless technologies at STL, focusing on disaggregated RAN, densification, virtualization, and automation; with significant involvement in ONAP and O-RAN.

Prior to joining STL, Sarat was with AT&T for nearly 35 years, managing multiple R&D teams performing research, development, and fielding a wide range of technologies, mainly in the wireless arena. He was also the head of network planning, development, and operations of the Birla AT&T wireless joint venture (currently Vodafone-Idea), during the initial days of wireless growth in India.

Sarat has been a prolific innovator of several telecommunications technologies with 70 patents, 40 papers, and a graduate-level textbook. He has been a recipient of the AT&T Science and Technology Medal and has been a senior member of IEEE. He also served as an associate editor of the IEEE/ACM Transactions on Networking and as a member of the New York University Wireless Research Board.

Sarat received his B.Tech. degree from the Indian Institute of Technology-Madras and Ph.D. from McMaster University, Canada, both in Electrical Engineering.

Rajesh Gangadhar

CTO - Access Solutions, STL

Rajesh.gangadhar@stl.tech

At STL, Rajesh drives technology transformation and innovation in the open networking domain, leading R&D, Engineering and Product Development of disruptive wireless and wireline access products.

In this role, he is responsible for driving new business growth through development of advanced Access Products leveraging open innovation and customer-centric R&D. He brings over 25 years of experience in Telecommunication industry.

Rajesh has completed his Master of Sciences from Villanova University, Pennsylvania and an MBA from Northeastern University, Massachusetts. In his previous stints, he has worked with companies like CableLabs, Charter Communications, Clearwire, Sprint and Nextel.

beyond tomorrow

About STL - Sterlite Technologies Ltd

STL is an industry-leading integrator of digital networks.

Our fully 5G ready digital network solutions help telcos, cloud companies, citizen networks, and large enterprises deliver enhanced experiences to their customers. STL provides integrated 5G ready end-to-end solutions ranging from wired to wireless, design to deployment, and connectivity to compute. Our core capabilities lie in Optical Interconnect, Virtualised Access Solutions, Network Software, and System Integration.

We believe in harnessing technology to create a world with next generation connected experiences that transform everyday living. With a global patent portfolio of 582 to our credit, 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, cable, and interconnect subsystem manufacturing facilities in India, Italy, China, and Brazil, along with two software-development centers across India and a data centre design facility in the UK.