HEAVY READING WHITE PAPER

Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

Open Mobile Networks: Systems Integration and Automation

A Heavy Reading white paper produced for Tech Mahindra

Tech Mahindra

AUTHOR: GABRIEL BROWN, PRINCIPAL ANALYST, 5G, HEAVY READING

AUTOMATION IN 4G/5G MOBILE NETWORKS

Mobile networks are critical national infrastructure, expanding rapidly in multiple dimensions. On the service side, new 5G capabilities are being introduced to serve increasingly demanding use cases across diverse industries. On the network side, new technologies and architectures enable operators to meet high density demand and the extreme performance requirements of advanced applications.

The challenge is that operators must meet these performance targets using network technologies that are inherently complex to integrate and operate without expanding capital expenditure (capex) and operating expenditure (opex). This paper argues that "open networking" principles can be applied by mobile operators to lower their cost of production and maintain the pace of innovation. Specifically, it addresses mobile radio access networks (RANs) and core networks in the context of 4G to 5G migration. It argues that to capture the benefits of open networks, operators need to automate network service life cycles.

Open and Disaggregated Networks

Inspired by the hyperscale cloud providers, mobile network operators are working to migrate network functions to cloud infrastructure platforms. The goal is to build cloud-native networks that are responsive to changing demand, automated whenever possible, and run with low operating costs. The transition is already reasonably advanced for back-office applications (e.g., business support systems) and for the mobile core network. Disaggregated RAN is running a cycle behind the core network, but work is now underway.

Decoupling virtual network functions (VNFs) from the infrastructure platform requires interoperability, onboarding, testing, optimization, scaling, and life cycle management. Even with well-established standards, this can be onerous for the operator. Consequently, specialist integrators, with experience in multiple applications and platforms, are often used to design and deploy the solution. Because multiple vendors are involved in any deployment, with each vendor's product on a different upgrade cycle, there is a strong requirement to automate these processes. Without automation, complexity and costs trend higher and service quality deteriorates.

5G compounds these challenges. One of the objectives of 5G is to radically expand the number of use cases an operator can serve across many diverse industries. The challenge is that each customer in each industry is likely to have unique configuration, performance, and reporting requirements, all of which the operator must support on optimized network platforms.

5G Strategies Are Defined by Services

5G networks are now live in lead markets, and Heavy Reading expects more than 40 commercial launches in 25 countries before the end of 2019. Initially, these networks will focus on enhanced mobile broadband (eMBB) services; however, operators will rapidly turn their attention to scaling advanced services and applications.

In **Figure 1**, Heavy Reading asked operator respondents to identify the primary drivers for 5G deployment over 2- and 5-year time horizons. On a 2-year view, the largest group (33%) says "faster speeds to end users" is the primary driver for 5G, followed by "system capacity & efficiency" (27%). Over a 5-year view, the finding is very different. The ability to address "new markets & services" climbs from last place to first place in the ranking with 47% of the responses, significantly higher than all other scores. This shows that operators remain committed to the vision of 5G as an enabler of new services and business models across diverse industry sectors.

Figure 1: What Will Be Your Company's Primary Driver for Deploying 5G Networks?

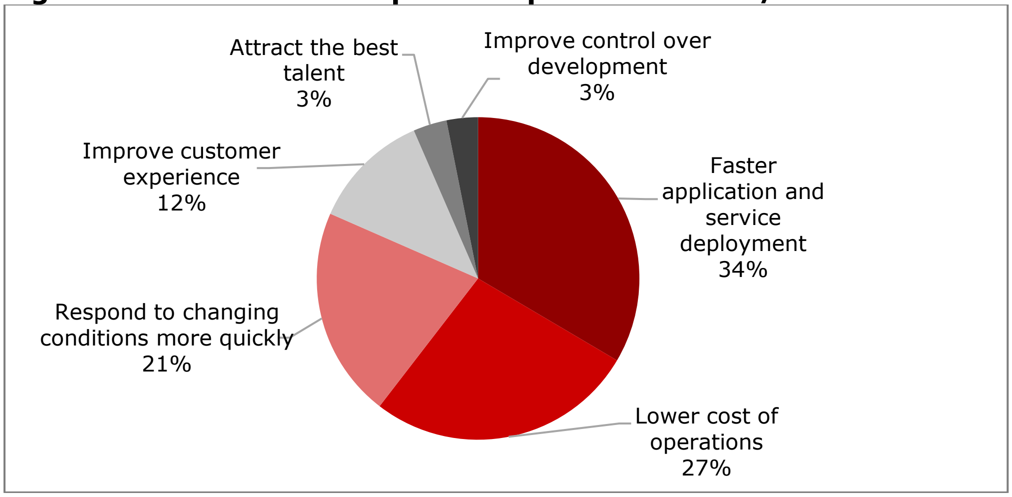
Timeframe	Faster Speeds to End Users	System Capacity & Efficiency	New Markets & Services	Competitive Reasons
Within the next 2 years	**33%**	27%	24%	23%
Over the next 5 years	17%	22%	**47%**	15%

N=142
Source: Heavy Reading's 5G Network & Services Strategy Report, 1Q 2019

Automation, DevOps, and Continuous Integration/Continuous Delivery

Dynamic and diverse services can only be delivered over 5G networks with investment in automation and associated working practices. **Figure 2** and **Figure 3**, excerpted from Heavy Reading's most recent Open Networking Operator Survey, shows most operators agree that such DevOps and continuous integration (CI)/continuous delivery (CD) models are critical. **Figure 2** shows that the primary benefits of DevOps are faster service delivery (34%) and lower cost operations (27%).

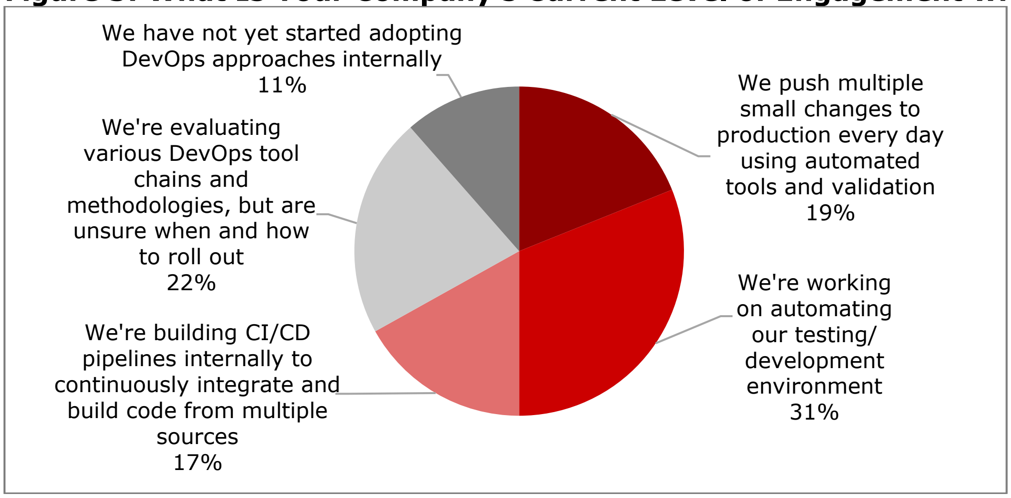
Figure 2: What Are the Top Two Expected Benefits/Outcomes of Adopting DevOps?



N=149
Source: Heavy Reading, Open Networking Report, 4Q 2018

Figure 3 shows the progress operators have made in adopting these techniques. A combined 66% of respondents said their company uses DevOps to some extent and a further 22% said they are evaluating the approach.

Figure 3: What Is Your Company's Current Level of Engagement with DevOps?



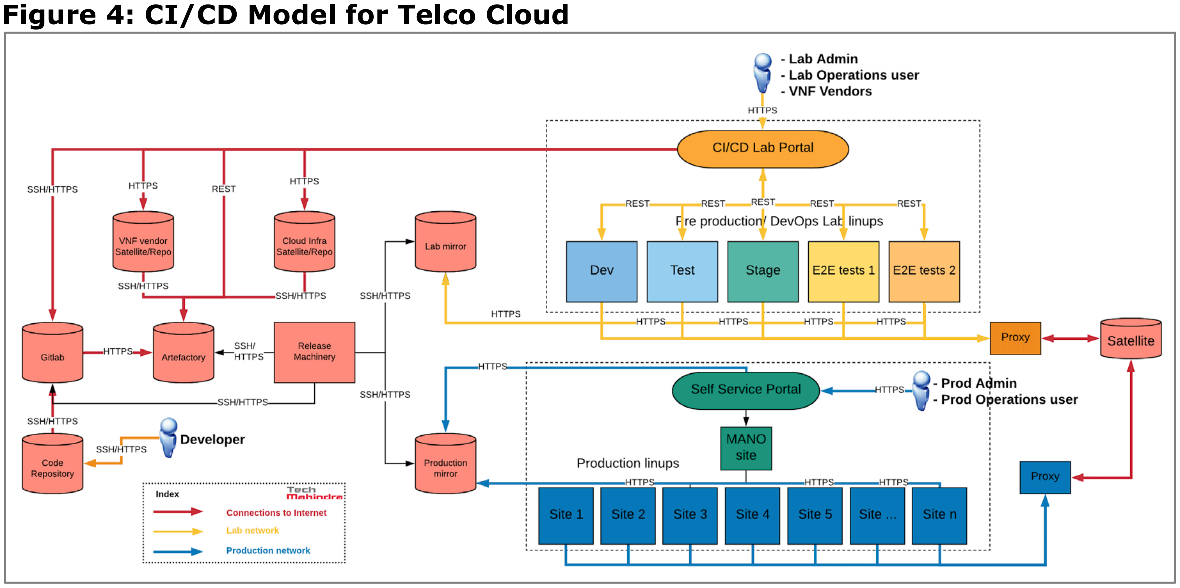
N=149

Source: Heavy Reading, Open Networking Report, 4Q 2018

Figure 4 shows a typical end-to-end CI/CD solution built using open-source components. In a telco cloud context, this involves creating and delivering a "super-package" consisting of the following:

- An infrastructure software release with deployable code and configuration templates.
- Product-specific packages covering hardware/firmware, BIOS, Host OS, Cloud OS, virtualized infrastructure manager (VIM), monitoring tools, automated testing, and VNF software.

Figure 4: CI/CD Model for Telco Cloud



Source: Tech Mahindra's netOps.ai Framework

Cloud infrastructure running telecom applications is, by definition and design, dynamic. To manage risk, the operator must establish procedures such that no release, patch, or change must go to production without validation and certification. Similarly, during the operating phase, in order to automate configuration changes, operators must adopt "shift-left" and "go/no-go" control practices to validate proposed configurations before staging. These CI/CD techniques take time to master and require expert DevOps engineers to implement. Well-staffed, experienced, vendor-neutral systems integrators can be valuable partners for operators on this journey.

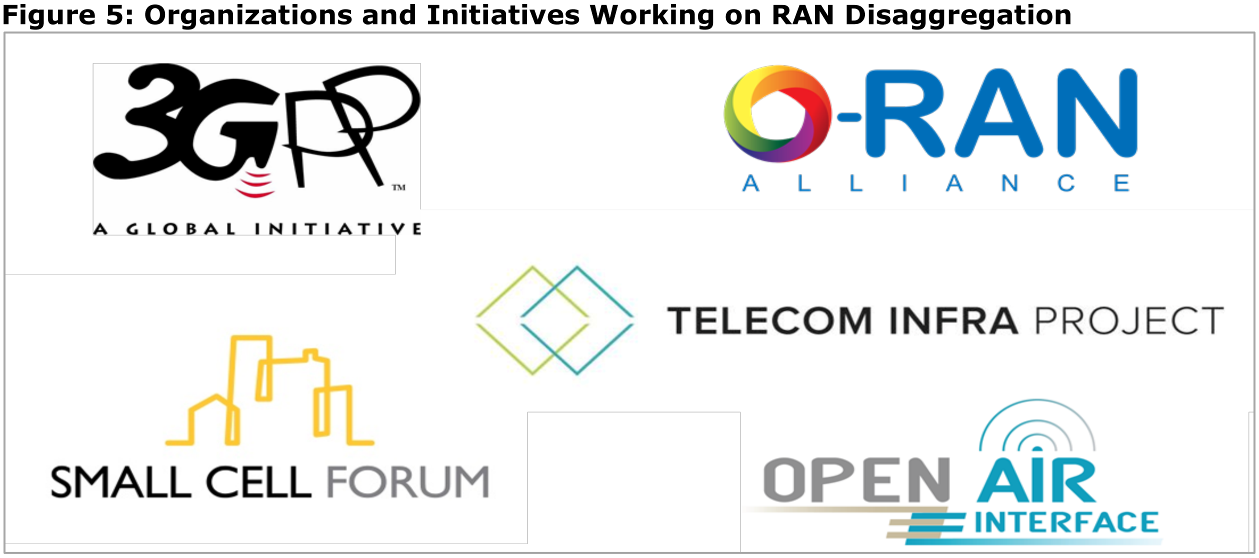
OPEN, DISAGGREGATED RAN

The RAN, as defined by the 3rd Generation Partnership Project (3GPP), is already open in terms of the air interface, which is well standardized and enables devices from different manufacturers to connect to operator networks. It is also open toward the core network. Within the RAN itself, however, the system is relatively closed. It is difficult to deploy base stations from different vendors alongside each other (especially in the same frequency bands). Base station products are made up of subsystems controlled by a single vendor connected over proprietary interfaces. The goal of open RAN technologies is to disaggregate the base station architecture and its functional components to create a more competitive market with more rapid innovation cycles.

There is, as yet, no formal definition of open RAN. In the meantime, Heavy Reading uses the following definition: "The ability to integrate, deploy and operate RANs using components, subsystems, and software sourced from multiple suppliers, connected over open interfaces."

Figure 5 shows several organizations that are working to develop open RAN interfaces and, in some cases, decouple RAN software from hardware. This level of activity from multiple bodies is a mark of the interest in this technology from operators worldwide.

Figure 5: Organizations and Initiatives Working on RAN Disaggregation



Source: Heavy Reading

Common to all these initiatives, there are three major aspect of open RAN to consider:

- **Market structure and competition:** Market forces have concentrated RAN equipment into a handful of vendors. Open RAN is, in part, a response to this.
- **RAN disaggregation:** This includes both hardware and software decoupling (virtualization) and interfaces between RAN subsystems.
- **Network architecture:** A disaggregated RAN is an opportunity for operators to rethink the network architecture; for example, RAN functions and edge cloud may be deployed at the same locations.

Market Structure and Competition

The mobile RAN equipment market was worth about $33 billion in 2018, with the top three vendors having close to 80% market share. RAN market concentration is due to the requirement to sustain high R&D investment, the importance of volume shipments to profitability, and the need for extensive field services and support. This market structure has advantages – it was created by market forces, after all – but consolidation limits vendor choice and, arguably, innovation in the sector. Moreover, it forces operators to move at the same pace as, or slower than, their major vendors.

One of the objectives of open RAN is to disaggregate the RAN so different components of the system can be developed and sold independently. The argument is that this will help stimulate innovation at the component and subsystem level because developers of these technologies and products will have a more direct route to market and will not be gated by the big RAN vendors. There are many companies within the RAN supply chain that develop innovative radio technology that are somewhat "hidden" beneath the dominant suppliers; the opportunity is to surface these companies and enable innovation in radio to flourish.

To change the market structure requires open interfaces between subsystems, open hardware platforms for software developers, and operator customers prepared to invest in these solutions. Progress in each of these areas is reasonably good.

RAN Disaggregation

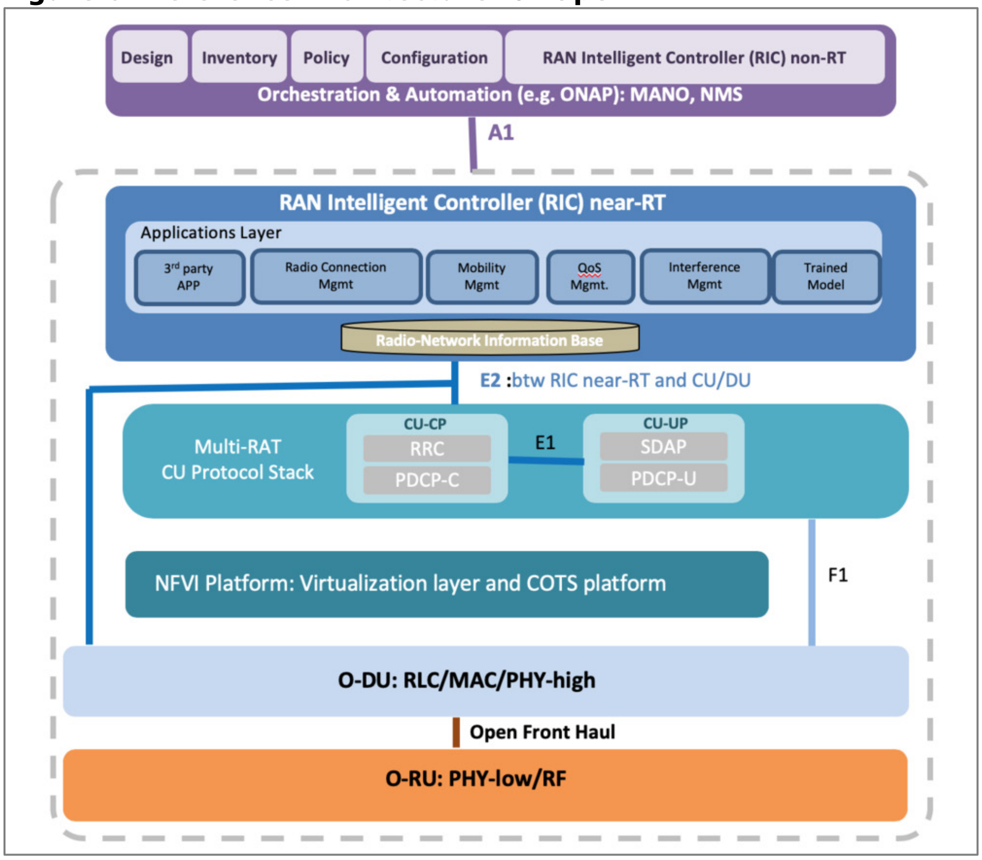
Radio base stations are designed and operated as integrated products, consisting of the entire radio and digital processing chain from antenna to radio, baseband, and an IP/Ethernet transport interface. In 5G terminology, these subsystems are known as the following:

- **Antenna unit (AU)**, which may be passive (AU) or active (AAU).
- The **radio unit (RU)**, which may be integrated with the AU or independent.
- The **digital unit (DU)**, which handles baseband processing.
- The **centralized unit (CU)**, which handles Layer 3 processing.

These functions can be deployed at the cell site, as is the case for the classic distributed RAN, or at different locations. For example, different subsystems can be deployed on the tower, in the cabinet, or in the edge data center.

The 3GPP has started to define open interfaces between these functions (notably F1 and E1) to enable interoperability between vendors. The O-RAN Alliance is contributing to this work by developing a more detailed architecture, shown in **Figure 6**. It is developing interfaces not yet addressed by the 3GPP; for example, an open fronthaul between DU and RU, and between CU and RAN intelligent controller (RIC) functions.

Figure 6: Reference Architecture for Open RAN



Source: O-RAN Alliance

In addition to open interfaces, another aspect of open RAN is the ability to decouple software from hardware using virtual RAN. To date, RAN products have used specialized hardware for radio frequency (RF) and digital processing for performance and power consumption reasons.

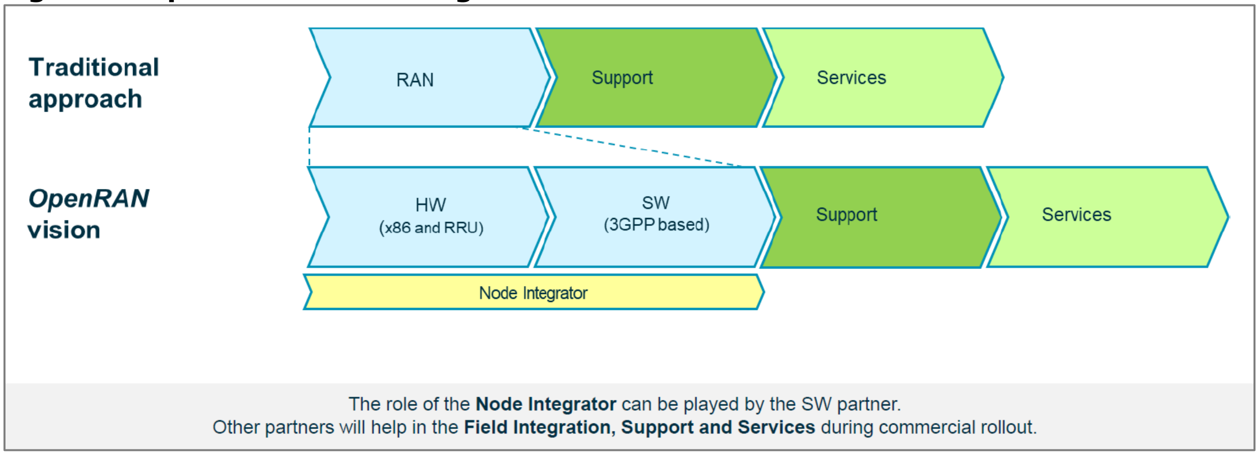
In digital baseband processing (i.e., the DU and CU), there is an opportunity to run a virtual baseband on commercial-of-the-shelf (COTS) server platforms (typically X86 today with hardware accelerators; e.g., using field-programmable gate arrays [FPGA]). The benefits of virtual baseband are the ability to pool capacity in compute clusters; the ability to support elastic scalability so resources can scale up during busy hours or be switched off to save energy during quiet times; and the ability to support in-service maintenance and upgrades. The challenge is to achieve performance and power consumption comparable to dedicated systems.

In the RF chain (i.e., the RU and antenna system), the idea of white box radio is emerging. The RU itself is harder to "software-ize" than baseband because there are many analog components, it is real time, and power consumption is critical. Nevertheless, open fronthaul interfaces between the RU and the CU/DU provide an opportunity to create and grow a competitive market for COTS radio head products.

RAN Systems Integration

The major challenge with disaggregated RAN is systems integration; breaking apart a monolithic system inevitably has costs, as well as issues associated with the responsibility for the overall solution and end-to-end service assurance. It is unrealistic to expect that a majority of operators will integrate disaggregated RANs themselves outside the lab, at production scale, because the costs are likely to be uneconomic relative to classic RAN. Instead, there is an opportunity for specialist providers to assume the role of "node integrator" and to provide ongoing support and maintenance, as shown in **Figure 7**. These integrators can reuse knowledge and processes from multiple customers to increase efficiency. The introduction of virtualization and COTS hardware to the RAN means that a traditional service assurance model will not work. The integrator must, therefore, have both cloud/IT and RAN domain expertise to troubleshoot, debug, and optimize performance.

Figure 7: Open RAN Node Integrator



Source: Telefonica, MWC 2019

Edge Network Architecture

Disaggregated RAN is an opportunity for operators to rethink the network architecture and deploy network functions and services on a unified edge cloud platform. However, this presents many new challenges for systems integration. How the RAN is split between RU, DU, and CU modules deployed at different geographic sites requires balancing the performance gains from pooling and colocation against the capabilities of the available transport network.

It also requires insight into end-user service requirements and the ability to deploy content and applications where appropriate. For example, a cloud gaming service that requires less than 10 ms latency may be deployed at the metro edge, while stadium services (e.g., virtual reality) may be best deployed at the venue and industrial Internet of Things (IoT) applications on the factory floor, at the warehouse, etc.

This is a new world to operators familiar with classic hierarchical architectures, where RAN is distributed to the cell site and core network functions are deployed in central data centers. Specialist integrators, with experience across markets, technologies, and operators, can play a valuable role here.

OPEN 4G/5G PACKET CORE

The mobile core network provides critical services, including session management, packet forwarding, mobility management, authentication, policy control and enforcement, communications services, and more. It is a relatively smaller part of an operator's capex investment compared to RAN; however, it is critical functionality. Outages have very serious consequences because large parts of a network can go offline due to a core network fault. Therefore, operators are risk averse when it comes to the core.

Progress on Virtual and Cloud-Native Core

Mobile core applications are, in principle, well suited to virtualization and cloud infrastructure. In the control plane, the signaling and transaction workload involved in setting up and managing user sessions can be readily processed on general-purpose CPUs. And in the user plane, while significant and growing, mobile network traffic is an order of magnitude lower than fixed broadband networks and can be handled in COTS servers (in 4G at least, but this may change in 5G).

Nearly all vendors, and many operators, have worked to virtualize the EPC, IMS, and policy infrastructure. After several years of development, and even though operators are inherently conservative in their approach to the core network, progress has been good. Heavy Reading estimates over 40% of EPC revenue and the majority of new businesses in 2018 were for virtual systems. In 2019, virtualized EPC will cross the 50% threshold and be worth more than classic EPC. Although vendors are still selling hardware systems, and some continue to invest in line-card upgrades, they are not developing new purpose-built hardware for the mobile core. Instead, most expect general-purpose processors (including CPUs, NPUs, and GPUs) will provide the performance they need for the 5G core. See, for example, AT&T's work with Airship and DevOps for its 5G core development.

One of the major changes for virtual EPC (vEPC), and the mobile core in general, is that vendors need to rewrite their applications to run natively on cloud infrastructure and support the associated scaling and resiliency models. Moreover, the mobile core is inherently stateful and classic designs tie state to software processes and hardware in a way that cloud-native applications do not.

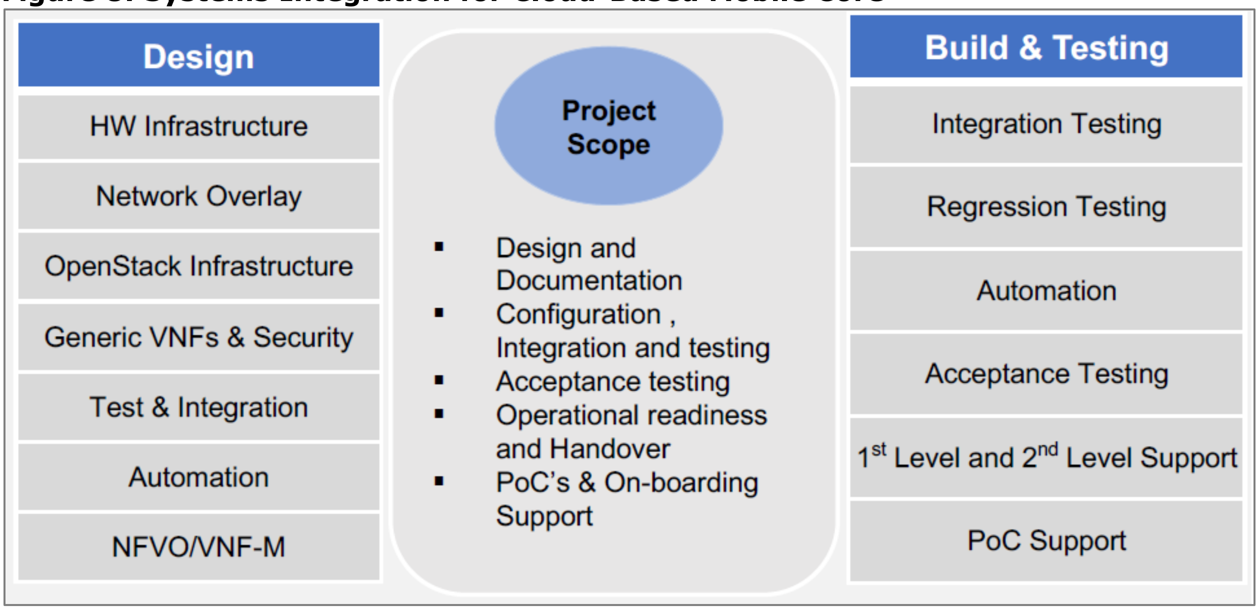
Multi-Vendor Platforms for vEPC

4G core VNFs should run on network function virtualization (NFV) infrastructure from different vendors, which typically means Red Hat (Open Stack) or VMware environments deployed on diverse COTS hardware. It may also include Google Cloud, AWS, Azure, and other public clouds. This interoperability is part of what distinguishes an open mobile core from vendor proprietary solutions.

A major challenge has been, and remains, the availability of VNFs that can be easily deployed across diverse platform environments and still provide consistent performance and reliability. Simply onboarding VNFs has proven difficult; scaling them, ensuring performance, and managing the life cycle have proven even more difficult. Mobile operators typically do not have broad experience in integrating VNFs with cloud platforms at production scale. As a result, many of the vEPC deployments today are effectively "full stack NFV," with a lead vendor providing VNFs, middleware, and integration services – and quite often also specifying server hardware.

Specialist systems integrators, not owned by VNF or platform vendors, can help operators transition to more genuinely open NFV deployments for the mobile core. As **Figure 8** shows, there are many aspects to first designing a system and then deploying and testing it. Operators can do much of this themselves, but the costs involved can easily erode the benefits of the virtualization. An attractive model, therefore, is for the operator to set requirements, define an architecture, and write design guidelines – and then work with specialist integrators on the detailed implementation.

Figure 8: Systems Integration for Cloud-Based Mobile Core



Source: Tech Mahindra's VNF-Xchange

Migration to 5G Core

The first 5G networks will be deployed in non-standalone mode with 5G New Radio (NR) connected to an upgraded EPC known as 5G-EPC. In a second phase, 5G NR can deployed in standalone mode attached to a new 5G core that will enable more advanced services.

In many cases, 5G-EPC will be virtualized. The trend, however, is toward a cloud-native core. Vendors and operators are now involved in plotting a migration from 5G-EPC deployed in virtual machines (VMs) to a cloud-native 5G core, composed of microservices and deployed in containers. In some cases, vendors are rewriting EPC applications as cloud-native; in other cases, the view is that vEPC is now stable and should remain in VMs, while development attention should focus on the cloud-native 5G core. In this scenario, legacy VNFs could, if desired, be deployed in VMs in containers while the transition to the 5G core takes place. It is also thought that a mixture of solutions may make sense; for example, control-plane functions in VMs and user-plane functions in containers.

In each of these cases, there remains significant integration overhead for the operator. And, as the number of options in the 5G-EPC to 5G core transition increases, there is a risk that complexity can multiply beyond the capacity of current operating processes. This underlines the need to move to a more automated CI/CD model in the mobile core.

ABOUT TECH-M'S NETOPS.AI FRAMEWORK

Content provided by Tech Mahindra.

Tech Mahindra's netOps.ai, Network Automation and Managed Services Framework, is geared to accelerate 5G network adoption by automating all the key network life cycle stages. The main features of netOps.ai include the following:

- **Telco Cloud Automation:** Automating the telco cloud deployment of different personas and enabling one-click deployment of a telco cloud.
- **VNF Life Cycle Automation:** Automating VNF onboarding and end-to-end network integration with integration accelerators in its VNF-Xchange.
- **NetDevOps:** Implementing complete DevOps CI/CD pipeline for 5G networks in a multi-vendor environment.
- **Service Orchestration:** Automation of network services with orchestration, including closed-loop control implementation for self-healing and auto-scaling.
- **Network Operation:** AI-driven intelligent network operations to significantly improve customer experience and drive "auto-pilot" mode of operations.

netOps.ai is geared to deliver service agility at scale, not only by automating all network life cycle stages, but also by keeping networks continually updated.