The Future is Virtual:
Routing in 5G Transport Networks

Volta Networks

Published by

Mobile World Live

Introduction

5G opens up a broad range of new services to Mobile Network Operators (MNOs). These new services will focus more on enterprises connecting things than end users connecting devices. However, new network architectures and significant investment will be needed to realize these opportunities.

Software and the cloud will be integral to deploy these 5G services economically and at scale. Network Function Virtualization (NFV) provides the basis for network operators to move away from proprietary hardware to software implementations. There are a number of NFV initiatives such as the European Telecommunications Standards Institute (ETSI), which released an NFV standard in October 2013 as well as cooperative projects like OPNFV. Like enterprises, operators can see the benefits of virtualization using commodity hardware like x86 servers. NFV improves scalability and agility by allowing service providers to deliver new network services and applications on demand, without requiring additional hardware resources.

NFV will require 5G operators to invest in computing resources closer to cell towers and the network edge to run their virtualized network equipment in order to meet latency requirements in 5G. Those facilities can also be used for a wider edge computing play. Given the benefits of the tight integration of edge computing with 5G, industry standards are important. ETSI's **Multi-Access Edge Computing (MEC)** defines an edge computing ecosystem. It is an open framework for applications and services that are tightly coupled with the Radio Access Network (RAN) via open interfaces to integrate software services into wireless networks.

For many new 5G applications like industrial, medical, drone and transportation, reliability and latency requirements surpass bandwidth needs. Most of these services will rely on some form of cloud computing. Conventional cloud data centers are concentrated in a few areas and are often quite distant from the user. Since signals cannot travel faster than the speed of light, distance translates into latency. The resulting latency will not meet the requirements for these emerging services.

These services will only work if the computing resources are much closer to the end user or device. The enterprise market will want support for both Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). They will also want the range of the development environments they currently use for PaaS. Finally, they will want multicloud implementations to avoid vendor lock-in. This suggests a distributed, cloud-friendly infrastructure where MNOs will play a key role.

However, 5G services cannot be delivered without a robust and modern transport network with integrated routing. The transport network is the glue that holds together the disaggregated RAN components. Routing is integral to these networks so that traffic can be managed to meet service level requirements. As end users take advantage of the greater bandwidth in 5G, the transport network will also need to scale to handle the increased aggregate traffic. In addition, many operators plan to converge their various networks. With 5G, both the core and access network can be converged so the same fiber plant that connects cell towers can also handle business and residential wireline services.

To understand how 5G determines the requirements for the transport network, there are seven drivers and trends that will determine how transport networks will be built:

1: 5G extends trends started in 4G

4G introduced virtualized software for functions that traditionally had been done in hardware. An individual function delivered as virtualized software is a Virtual Network Function (VNF). This disaggregated approach has a number of benefits. MNOs have realized that costs needed to be reduced given the significant decline in revenue per bit. Cloud radio access network (C-RAN) showed that significant operational expenditure (opex) and capital expenditure (capex) reductions can be achieved with virtualization compared to traditional equipment deployments. In fact, a trial from China Mobile showed a 53 per cent reduction in opex and 30 per cent savings in capex[1].

In 4G LTE, eNodeB, the functions were disaggregated into the Remote Radio Head (RRH) and the Baseband Unit (BBU) for C-RAN implementations. Moving the BBU to a BBU hotel (such as a central office) eliminates a significant amount of equipment from the base station. This eliminates the principal source of heat generation inside the base station, making it feasible for the remaining equipment to be cooled passively. Japan-based NTT Docomo's goal is to reduce a single site's power consumption by over 75 per cent[2]. A study by ACG Research compared the use of virtual infrastructures to using purpose-built platforms in the mobile packet core. It found the cumulative total cost of ownership (TCO) to be 67 per cent lower than a purpose-built solution[3].

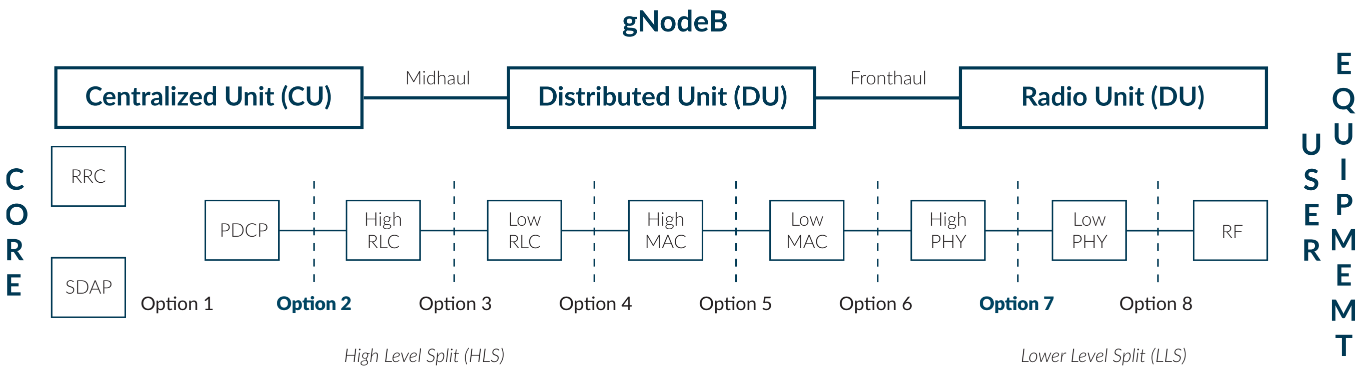
According to Rethink Research, operators will deploy centralized and virtualized macro cells and microcells at a compound annual growth rate (CAGR) of 23 per cent between 2017 and 2025. These deployments will overtake new deployments of conventional cells by 2022.

The CPRI specification enables the communication between the RU and the BBU in C-RAN. CPRI provides an interface for the interconnection of remote radio heads with a baseband unit by breaking up traditional backhaul into fronthaul and backhaul networks. However, CPRI is constant bit rate and does not scale well for 5G.

5G Next Generation RAN (NG-RAN) takes the cell site functions and breaks them down into multiple functions. 5G gNodeB functions are the Radio Unit (RU), the Distributed Unit (DU and the Centralized Unit (CU). These interconnect via the transport networks and then connect to the core network. The DU and CU can be implemented as VNFs and centrally pooled for saving by not overprovisioning the cell site and keeping costs low. It is important to note that the 5G core is also virtualized and the location of some elements, like the User Plane function (UPF), is application specific.

3GPP Release 14 addressed how the disaggregated RAN functions can be connected using the concept of functional splits. These define the possible ways that the functionality can be disaggregated and then connected though standardized interfaces.

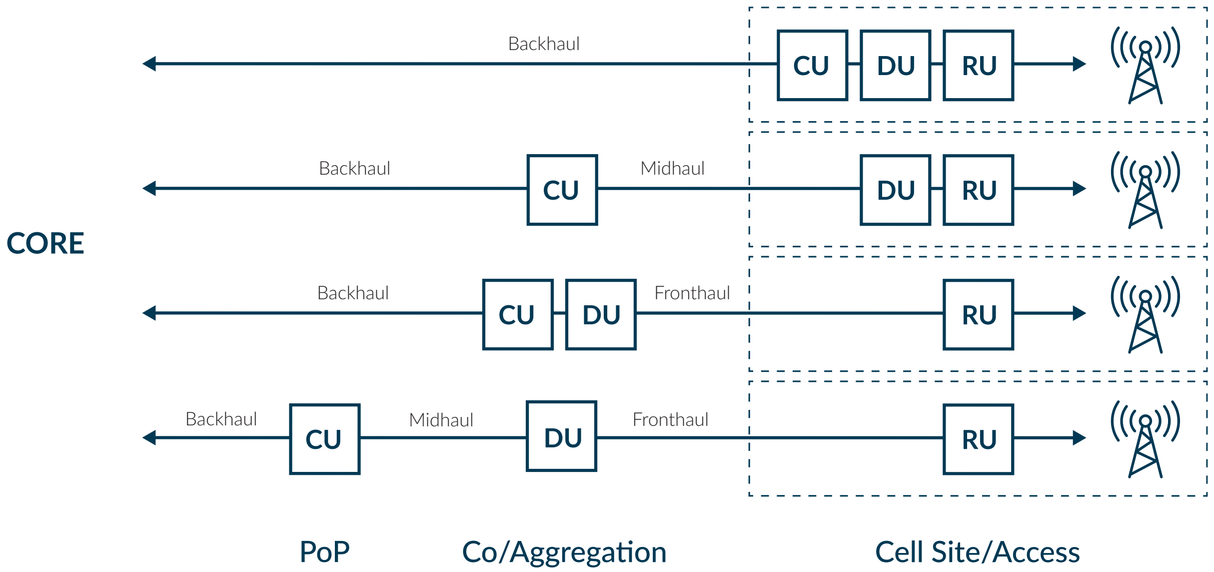
Functional splits in 5G enable a range of approaches to the transport network. While the radio unit is at the cell tower, there is a great deal of flexibility in placing the Distributed Unit (DU) and the Control Unit (CU) which can operate as VNFs.



Lower Level Splits are constrained by latency budgets. The Option 7.X splits offers the most support for advanced RAN features while maximizing TCO gains from reduction in radio complexity and centralized NFV[4]. According to a study[5], the 7.X functional split is the most popular of functional splits and is also adopted by industry groups like the O-RAN Alliance.

Configurable splits offer the potential to adapt a network architecture to fit particular use cases and requirements so that it will deliver the desired speed latency and throughput. Split option 7 (particularly 7.2) keeps the cost of the RU low. That may be best suited where there are large numbers of small cell sites. Split option 2 places complexity at the RU so it increases the RU costs. As we noted earlier, CPRI is not well suited to 5G which drives the interest in eCPRI. Option 7 splits enable roughly ten-times bandwidth compression at peak rate relative to Option 8 with CPRI. Further, Option 7 splits allow fronthaul bandwidth to vary with the user plane traffic load which offers significantly more compression relative to Option 8 (which is constant bit rate) when cell load is less than peak. High Level splits (like Option 2) are much less latency sensitive.

Figure 1: The functional split will determine the placement of the disaggregated network functions and the requirements for the network that ties them together



The consensus is that the combination of split 7.2 and 2 will be the dominant implementation. As a result, transport networks allowing fronthaul, midhaul and backhaul to be supported based on the specific needs for a given area. However, the DU and CU can be implemented as VNFs which optimizes flexibility. Routing must be equally flexible, which is driving a high interest in virtual routing.

The fronthaul bandwidth depends on the exact split point between the DU and the RU. As we noted, we need to make assumptions about the RU to determine bandwidth. The following compares the fronthaul bandwidth needs for a 64 Transmit-64 Receive Massive MIMO installation with 100MHz system bandwidth:

- **Split 6:** PHY is completely implemented in the RU. This option requires a 3Gb/s link.
- **Split 7:** PHY is split between the DU and the RU. The bandwidth need varies between ~10Gb/sand 140Gb/s. The 7.2 and 7.3 splits seem more realistic as they keep the Massive MIMO beamforming at the Radio Unit.
- **Split 8:** PHY is moved completely to the DU. This split seems impractical as it requires 236Gb/s link.

Fronthaul	• Connects the DU with the RU. • Fronthaul latency is constrained to 100 microseconds. • A DU may be serving RUs up to 10 km away.
Midhaul	• Connects the CU with the DU. • The latency on the link should be around 1ms. • A centralized CU may be controlling DUs in an 80km radius.
Backhaul	• Connects the 4G/5G core to the CU. • A latency of ~40ms may be tolerable on this link. • The 5G core may be up to 200km away from the CU.

Latency is a factor in transport networks. As is noted in the chart, there are latency budgets to connect the RU, DU and CU. Since these can be software VNFs running on commodity hardware (i.e. x86 servers), the servers, like MEC resources, will have to be relatively close to the edge. This was a driver for the Central Office Re-architected as a Data center (CORD).

Implication: The network will be virtualized and disaggregated.

2: More Bandwidth per Cell Site

There is a great deal of excitement about what can be done with the kind of bandwidth available in 5G. Higher frequency spectrum, coupled with antenna technology like MIMO, make this possible. According to TECHnalysis Research, "at the present time, we should see 5G speeds in the 400-500 Megabit per second (Mb/s) range for sub-6 5G service and over 1.5 Gigabits per second (Gb/s) for mmWave, when the services are available. That's significantly faster than the 35Mb/s average across the US right now for 4G services[6]."

Figure 2: Different spectrum ranges will support different services in 5G and will offer different capabilities than 4G LTE.

	RAT/BAND	ILLUSTRATIVE CONVERGAGE COMPARISON	SCENARIO	5G COVERAGE	5G DATA RATE
<6 GHz	NR mmWave	NR	Local coverage Peek data rate 10Gbps	Less than 1 mile	1 to 3 Gbps
<6 GHz	NR 3.5GHz mMIMO LTE 1800	NR LTE	Resuses of 1800 grid possible for Downlink Peak data rate 1Gbps	Radius of several miles	100 to 900 Mbps
<1 GHz	NR 700MHz LTE 800 MHz	NR LTE	Deep indoor penetration Peek data rate: 100MBps	Hundreds of square miles	30 to 250 Mbps

How much bandwidth will come from each 5G cell site is complex and involves a number of assumptions. However, there are resources available that describe this. NGMN did a comparison of 4G to 5G NR sites. They found that where a 1Gb/s interface from the cell site would suffice in 4G, the 5G NR would require at least a 10 or more likely 25Gb/s interface[7]. However, services like eMBB offer up to 10Gb/s per location so the likelihood is that the bandwidth will be much higher for some cell sites. In fact, IEEE 1914.1 assumes that the network interface from a fronthaul node will be between 100 and 400Gb/s.

High-speed user equipment connections in the gigabit range get a lot of attention because the existing transport networks were not designed to handle the cumulative impact of this much bandwidth from the user equipment. As end users take advantage of the greater bandwidth in 5G, the transport network will need to scale to handle the increased aggregate traffic. In addition, many operators plan to converge their various networks. With 5G, convergence in both the core and access network can be converged. The same fiber plant that connects cell towers can also handle business and residential wireline services.

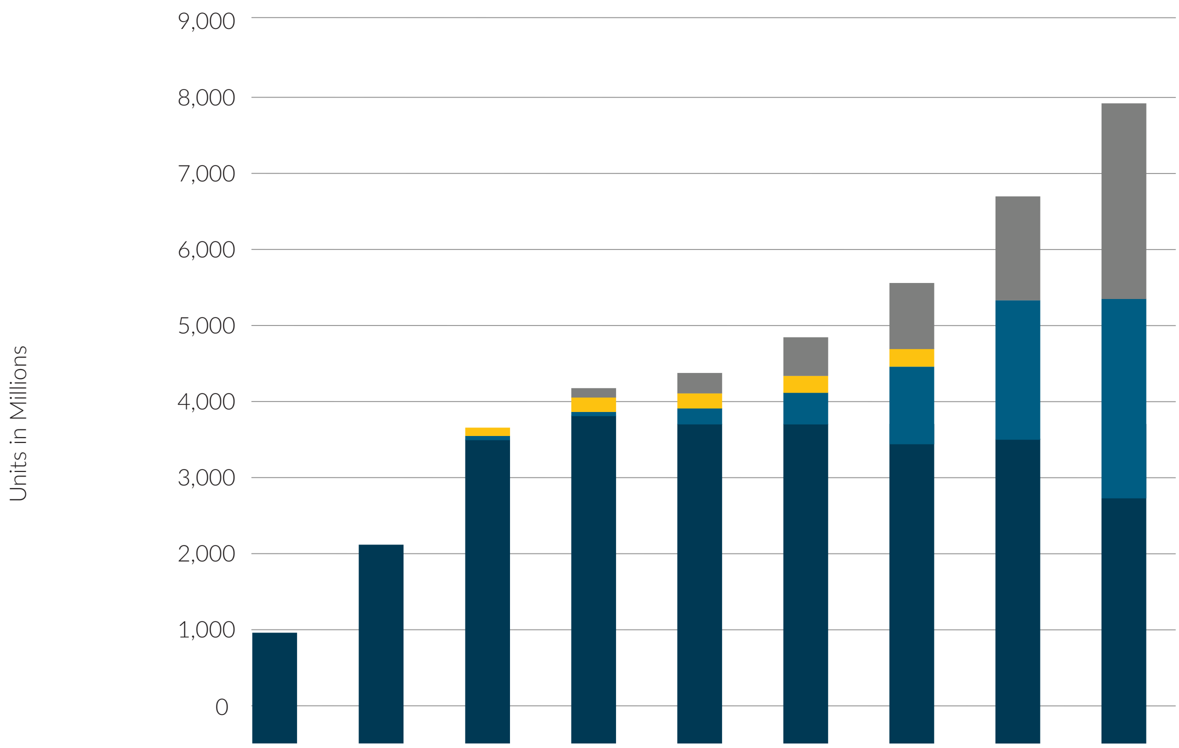
Implication: The transport network will need to handle much more bandwidth from each cell site.

3: Densification means more cell sites

5G's use of higher spectrum and small cell sites will drive densification so there will be many more cell sites to connect. The spectrum determines the distance covered by a cell site. One low band (600-700MHz) tower can cover hundreds of square miles with 5G service that ranges in speed from 30 to 250Mb/s. A mid-band (2.5/3.5GHz) tower covers a several-mile radius with 5G that currently ranges from 100 to 900Mb/s. Lastly, a high-band (millimeter wave/24-39GHz) tower covers a one-mile or lower radius while delivering roughly 1-3Gb/s speeds. The performance of each of these tiers can improve over time.

Thus, there will be many more cell sites than currently exist, although they will be much smaller. The greenfield buildout of small cell sites is significant. The Small Cell Forum estimates that the number of cell sites will grow from 4.1 million in 2019 to 8.4 million in 2020:

Figure 3: The number of small cell sites will grow dramatically with 5G NR growing in importance.



	2017	2018	2019	2020	2021	2022	2023	2024	2025
5G NR Standalone				125	241	443	829	1,353	2,562
5G NR Non-standalone			118	167	227	249	245		
5G/4G combined BTS			34	92	222	416	1,035	1,843	2,621
LTE or 3G/4G	1,462	2,602	3,986	4,279	4,185	4,197	3,928	3,982	3,220

Implication: The number of cell sites will increase significantly.

4: Ethernet and IP to the edge

NGMN's "RAN Evolution Project Backhaul and Fronthaul Evolution" has a good discussion of the physical layer options. From fiber to microwave, there are a range of physical layer options, and through various standards groups (eCPRI, IEEE 1914, 802.1cm, TSN), the industry is increasingly moving toward a converged transport option for fronthaul/midhaul/backhaul scenarios supported by different standards:

- eCPRI is significantly more efficient than CPRI and can be transported over Ethernet. The eCPRI frames are carried over IP and an Ethernet frame with different sections carried over different layers of Ethernet frames.
- IEEE 1904.3 Radio over Ethernet (RoE) is an open standard that specifies a transport protocol and an encapsulation format for transporting time-sensitive RAN-related application streams over Ethernet-based transport networks.
- IEEE 1914.1 Packet-based Fronthaul Transport Networks also uses Ethernet framing. IP packet technologies offer cost-efficient speed and capacity growth, driven by the enterprise, access, and data-center markets.

  IEEE Standard 1914.1 specifies details that allow packet-based fronthaul transport networks to be a flexible and efficient solution for the transport of 5G cellular services.

IEEE 1914.1 discusses many different models with the key being the ability of the network to service all different requirements. Key resource elements, like the DU and CU, must be physically located to minimize overhead while still ensuring minimum latency. The network is being split into two key components. NGFI-I connects the RU and DU, while NGFI-II connects the DU and CU.

Implication: The transport network will use Ethernet to carry IP.

5: New services with stringent requirements

Some of the new 5G services will take advantage of the higher bandwidths available in 5G such as Enhanced Mobile Broadband (eMBB). High-speed connections in the gigabit range get a lot of attention because the existing transport networks were not designed to handle the cumulative impact of this much bandwidth from the user equipment. At the other extreme is Massive Machine to Machine Communications (mMTC) which, as the name implies, will provide connections for a massive number of low-bandwidth IoT devices such as sensors.

The third class of services will be supported by Ultra-reliable Low Latency Communications (URLLC) which will provide new communications services for industrial automation, Smart City intelligent transportation systems, connected and autonomous vehicles, and tele-healthcare.

URLLC is different from the other services. For many existing industrial, medical, drone and transportation, reliability and latency requirements surpass bandwidth needs. Thus, the network must deliver very low latency with high reliability. Moreover, most of these services will rely on some form of cloud computing. Conventional cloud data centers are concentrated in a few areas and are often quite distant from the user. Since signals cannot travel faster than the speed of light, distance translates into latency. The resulting latency will not meet the requirements for these emerging applications. These services will only work if the computing resources are much closer to the end user or device.

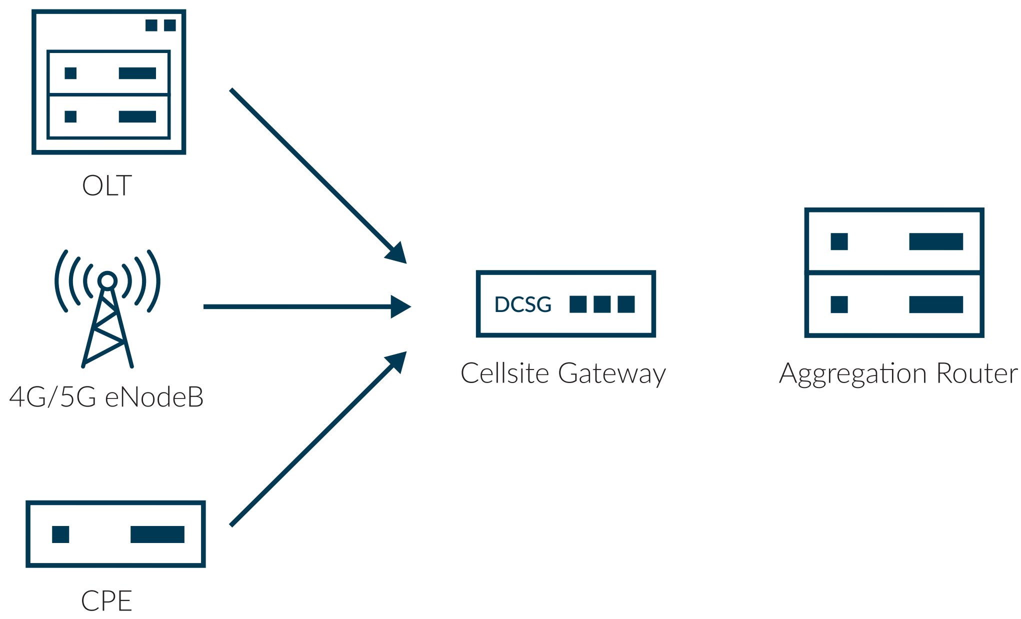
Implication: New services will have stringent and varying requirements in terms of latency and bandwidth.

6: Routing

Since the networks will be IP based and there are stringent requirements including latency, service providers expect to use routing extensively. A well-documented example is the Telecom Infra Project's (TIP) Disaggregated Cell Site Gateway (DCSG) project. TIP specifies the hardware and software requirements to connect a small number of cell sites per router. Moreover, they envision using the DCSG to connect and cover business and broadband services. The DCSG is expected to support MPLS, Segment Routing, QoS, and timing and synchronization.

The ability to deliver different service levels is critical to services using network slicing. Fundamentally, using routing for the service intelligence in this type of platform will be critical to ensure the performance of the various applications running on the 5G transport network.

Figure 4: Cell site Gateways will need to connect multiple cell sites and enable convergence of enterprise and broadband services.



Given this complexity, routing must be flexible and very cost effective, which is driving a high interest in virtual routing. Legacy routing vendors have attempted to implement router virtualization with approaches like logical routers, virtual routing and forwarding (VRF) and partitions, such as Cisco's Secure Domain Routing and Juniper's Node Slicing. However, none of these meet the full requirements, especially on cost.

We have seen how virtualizing network functions is critical to cost effective and scalable 5G deployments. Further, there are new resources in the network that can enable virtualized functions while also being used for new services. Can routing be virtualized?

Let's start with a definition of a virtual router. Like server virtualization, a virtual router must be one of **multiple, separate workloads** within a given piece of hardware. This means that there should be multiple routing elements on a single piece of hardware. A routing element must be a **separate management domain** with its **own control plane and dedicated forwarding resources (logical or physical ports).**

We see two basic approaches that meet our definition of a virtual router:

- A router OS like JUNOS or IOS can be implemented as a Virtual Machine (VM) running on a hypervisor, which runs on a general-purpose processor like an Intel x86 server or equivalent. The vendor will specify which hypervisors are compatible with their software. It is noteworthy that this is also a VNF. The limitation of this approach is that these x86 CPU-based data planes do not provide the throughput necessary for much of the routing in 5G environments. These VNFs are limited to 10 to 40 Gb/s. Moreover, the necessary server hardware has a very high price per port for high bandwidth applications like connecting cell sites.
- Alternatively, the control and management planes can be fully disaggregated from the underlying hardware, such as a white box switch. The control plane can run on the cloud and manage the switch. Each control plane can be thought of as equivalent to a Routing Engine or Route Switch Processor. It runs its own set of routing protocols as a separate management domain. This allows for multiple routing engines to run in the cloud. These routing engines are then associated with switch resources such as physical or virtual ports (versus with a whole line card as in SDR or Node Slicing). Thus, each set of the virtual routing engine and the switch ports functions as a virtual router. The cloud provides the most cost-effective way to scale processing and since the control plane portion is not dependent on the processor on the switch. Thus, the number of virtual routers can scale significantly on a single white box switch.

By definition, a virtual router is disaggregated, meaning the MNO can choose the hardware and software separately and are not locked into a single vendor. However, not all disaggregated implementations support multiple virtual routers. A Network Operating System (NOS) is disaggregated but is a single instance of the routing software running on an open network device. Moreover, the CPU and memory on a white box switch is limited to keep cost reasonable. Their strength is the switching ASIC which provides the real power of these devices. There simply is not enough processing to run multiple virtual machines or VNFs on these switches.

A cloud-based approach to router virtualization provides the only means of implementing multiple virtual routers on an ASIC-based platform. It is noteworthy that MNOs are specifying the ASICs to be used in cell site gateway routers. This will be critical in order to support applications like RAN sharing, network slicing and MEC. VM or VNFs of router software running on servers may have a role within the compute infrastructure but are unlikely to meet the needs of the 5G transport infrastructure.

RAN sharing is a way for multiple mobile network operators to share radio access network infrastructure. Two mobile operators share cell sites and networking infrastructure to share capital costs and better serve customers. It is popular among carriers because it is a cost-effective way to increase their coverage. Virtual routing allows each service provider to have their own router that can operate and be managed with complete separation of both the control plane and the management plane.

This will be best accomplished by having multiple virtual routers on the cell site gateways.

Implication: Disaggregated cloud control plane routing to leverage ASIC based, white box switches.

7: Network slicing

Network slicing allows operators to deploy different "slices" of the network, where these virtual networks run on a common infrastructure. Network slices are defined as a set of virtual routers under the same administrative domain. Virtualization of network functions is a key enabler of network slicing. The dynamic provisioning and management of network slices must go all the way to the cell site (and its router) where having separate virtual routers each with its own administrative domain is essential to make this service practical.

Each network slice is isolated and tailored to the specific requirements required by very different applications like machine-type communication, ultra-reliable low latency communication, and enhanced mobile broadband content delivery.

Implication: Network slices will require virtual resources in common network infrastructure including sets of virtual routers to support different applications.

Conclusion

We started by noting how 5G will use both software and the cloud. Clearly the use of software-based network functions coupled with virtualization is critical to delivering the next generation of services cost effectively. The applications supported by these services will have to integrate computing much closer to the end user. This means the cloud must come to the user. The same infrastructure that delivers cloud services can be used to deliver network functions which can now be extended to routing.

5G transport networks must support the following:

- A disaggregated approach using virtualization for software-based network functions
- Much more bandwidth per cell site
- Many more cell sites, especially small cells.
- All IP over Ethernet down to the cell site
- New services with stringent delay and bandwidth requirements
- Disaggregated virtual routers with scalable control plane
- Network slicing to support different applications in common network infrastructure

Given that the initial investments in 5G edge buildouts will include much more routing, MNOs will need to ensure that their choices can meet this set of requirements in order to futureproof their choices. Virtualization and disaggregation will help keep costs low. They will also be critical to responsiveness and service agility which will allow MNOs to maximize their time to revenue.

Volta Networks delivers the first cloud-native routing engine which reduces cost by an order of magnitude by using the cloud to optimize routing for low cost, white box switches. Unlike legacy routers or routing appliances, the Volta Elastic Virtual Routing Engine enables much lower cost by virtualizing and scaling the routing control plane while using automation to accelerate new network services.

Find out more at www.voltanet.io.

Produced by the mobile industry for the mobile industry, Mobile World Live is the leading multimedia resource that keeps mobile professionals on top of the news and issues shaping the market. It offers daily breaking news from around the globe. Exclusive video interviews with business leaders and event reports provide comprehensive insight into the latest developments and key issues. All enhanced by incisive analysis from our team of expert commentators. Our responsive website design ensures the best reading experience on any device so readers can keep up-to-date wherever they are.

We also publish five regular eNewsletters to keep the mobile industry up-to-speed: The Mobile World Live Daily, plus weekly newsletters on Mobile Apps, Asia, Mobile Devices and Mobile Money.

What's more, Mobile World Live produces webinars, the Show Daily publications for all GSMA events and Mobile World Live TV – the award-winning broadcast service of Mobile World Congress and exclusive home to all GSMA event keynote presentations.

Find out more at www.mobileworldlive.com

[1] "C-RAN: The Road Towards Green RAN", China Mobile Research Institute
[2] "https://www.zdnet.com/article/what-is-5g-the-business-guide-to-next-generation-wireless-technology/
[3] "ACG Research, The Business Case for a Common NFV Platform, October 2015.
[4] "U. Dotsch et al, Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE, Bell Labs Technical Journal 18(1), 2013
[5] Larsen et al, A Survey of the Functional Splits Proposed for 5G Mobile Crosshaul
[6] Networks, IEEE, 2018
[7] https://www.forbes.com/sites/bobodonnell/2019/11/19/how-fast-will-5g-really-be/#55eb88625cf3

Disclaimer: The views and opinions expressed in this whitepaper are those of the authors and do not necessarily reflect the official policy or position of the GSMA or its subsidiaries.