White Paper

Designing 5G-Ready Mobile Core Networks

Prepared by

Gabriel Brown
Senior Analyst, Heavy Reading
www.heavyreading.com

on behalf of

www.affirmednetworks.com

and

www.juniper.net

September 2016

5G Core Service & Performance Requirements

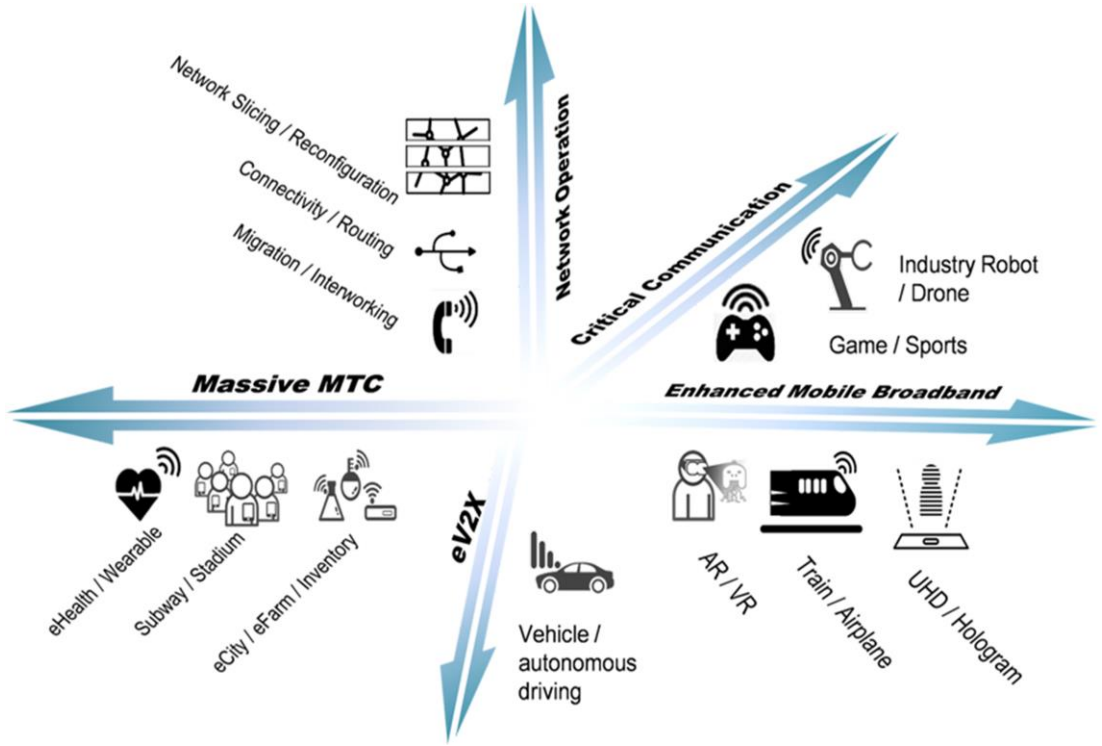
An industry "Vision for 5G" is now established and supported worldwide. Services and performance requirements have been identified, and the industry is moving into the development and implementation phases with a view to commercial launch of 5G mobile service around 2019 and fixed wireless access using mmWave radio from 2017. To meet these schedules, leading operators – especially Tier 1s with high expectations and aggressive deployment timelines – are working to make their networks 5G-ready ahead of commercial deployment.

This white paper argues that with a 5G-ready technology strategy, operators can prepare for rapid 5G service launch in a way that optimizes their investment in next-generation IP and mobile core platforms over the next three years. Specifically, it discusses 1) how 5G services drive a requirement for an IP services fabric to connect the distributed data centers that will host 5G network functions, content and applications and 2) the development of cloud-native, service-orientated core networks for advanced 4G and 5G networks.

5G Services Dimensions

Development of the 5G system architecture, and associated core network, is being led by 5G service requirements. **Figure 1** shows the output from the 3GPP SMARTER study group that has investigated the service dimensions of 5G.

Figure 1: Service Dimensions in Future 5G Networks



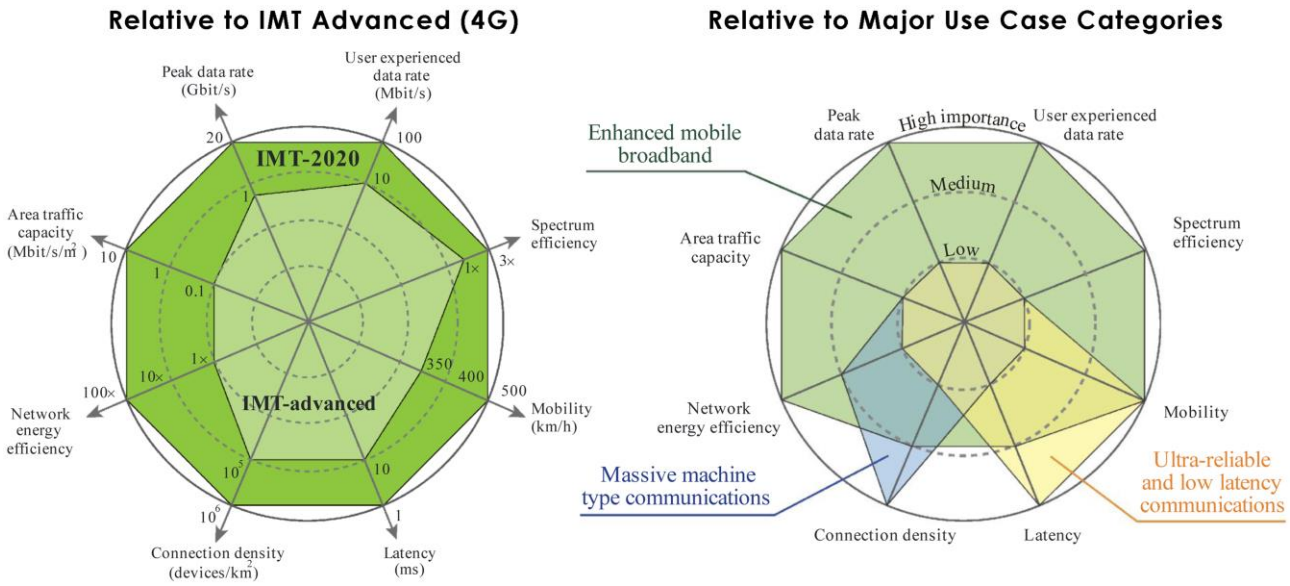
Source: 3GPP SMARTER

The chart maps closely to the now famous triangle representation of 5G services made up of enhanced mobile broadband (eMBB), low latency and mission critical communications and massive machine-type services (mMTC). However, it adds two important axes: a dedicated eV2X axis for vehicular services and, significantly, a network operations axis.

The dedicated operations axis is important because it makes explicit that automation is fundamental to 5G. To address new commercial opportunities, operators need the ability to create service-specific "network slices" that extend end to end across the infrastructure. Automating set up and management of these "network slices," including configuration of the underlying IP network, is critical to the 5G business case.

These service dimensions incorporate many different performance requirements. **Figure 2** shows IMT-2020 performance targets, developed by the ITU, for 5G. To the left, these are shown in comparison to IMT Advanced (4G). To the right, performance requirements are mapped to the three major use-case categories.

Figure 2: IMT-2020 Performance Targets



Source: ITU Recommendation ITU-R M.2083-0, September 2015

Probably the most influential factor on 5G physical network deployment will be the low-latency requirement. To deliver services over a wide-area network (WAN) with a consistent 5-10ms roundtrip time is extremely challenging and will drive a new system architecture deployed on a distributed cloud infrastructure.

A 5G-Ready Network Investment Strategy

Operators now have a good idea of what 5G will look like in scope and a reasonable view into the initial deployments models. Development of normative standards for 5G new radio (NR) and the next-generation core network (NG Core) is underway

with the first specifications – and the first "compatible" pilot deployments – expected in mid 2018. Although there is uncertainty about the details of the specifications under development, the industry has already undertaken significant R&D work on 5G candidate technologies, making it is possible for operators to pursue a 5G-ready network strategy with a reasonable degree of confidence that they will be able to rapidly deploy 5G, at scale, as the technology becomes commercially available.

From an IP and core network perspective, five tiers of a 5G-ready investment strategy are, from the ground up, as follows:

1. **Develop a distributed data center footprint**. The Central Office Re-architected as Data Center (CORD) model provides a good reference. The idea is for operators to develop physical assets located close to the edge and transform them into distributed data centers. The edge location is critical for low-latency 5G services and, as Mobile Edge Computing (MEC) shows, the edge-cloud is valuable in many networking scenarios – it is not solely a 5G investment. In time, the distributed data center will also come to support Cloud RAN (C-RAN) and Virtual RAN (V-RAN) hub sites and will host core network functions, such as distributed user-plane nodes.
2. **Create an "IP Services Fabric" for 5G with software-defined networking (SDN) control**. Edge-cloud locations running 5G network functions and services will require high-performance, secure connectivity. Investment in wide-area SDN is already underway to create an "IP services fabric" that connects centralized data centers, distributed cloud locations and cell sites. This IP services fabric provides routing, security, service chaining, redundancy and orchestration. Again, this investment is, in many aspects, not unique to 5G, but common to high-performance networking at the cloud edge.
3. **Invest in cloud-based Evolved Packet Core (EPC) and NG Core.** With a distributed cloud infrastructure, it is logical to redesign core network functions to take advantage of this asset to meet the strict latency requirements of 5G services and efficiently manage data traffic. Some aspects of NG Core are already reasonably well understood, (for example, control- and user- plane separation or "CUPS"), while others remain works in progress (for example, mobility and session management). In general, expect control-plane separation to apply across the 5G architecture, from radio to core.
4. **Use network slicing for a service-optimized core.** Network slicing is an important bridge from a 4G core to a new 5G core and is a key component of a 5G-ready network investment strategy. In this model, virtual network functions (VNFs) are created per service and data is routed via an optimized processing path across the network. By isolating services into virtual network slices, operators can offer better security, more efficient transport, optimized core network processing and appropriate service quality. A network slice created for an Internet of Things (IoT) customer, for example, may support both 4G and 5G access networks.
5. **Automation and service orchestration.** Automating network operations provides the ability to launch new services rapidly, to reach new market segments, to evaluate success/failure quickly, and then to modify, scale and repeat. Significant progress is being made in applying service orchestration to 4G-Long Term Evolution (LTE) networks, especially where software-based core networks are deployed. Network slicing is a good example of the need for automation because there will be many more logical entities to commission, provision and manage. "Automated service orchestration" is a good, tangible example of 5G-ready functionality.

The 5G Core Network

The new 5G core network is now reasonably well understood: It will be "cloud-native," it will make extensive use of network slicing, and it will operate in concert with a new model-driven service orchestration layer. The industry can therefore start to develop NG Core products that can be adapted as standards emerge and solidify.

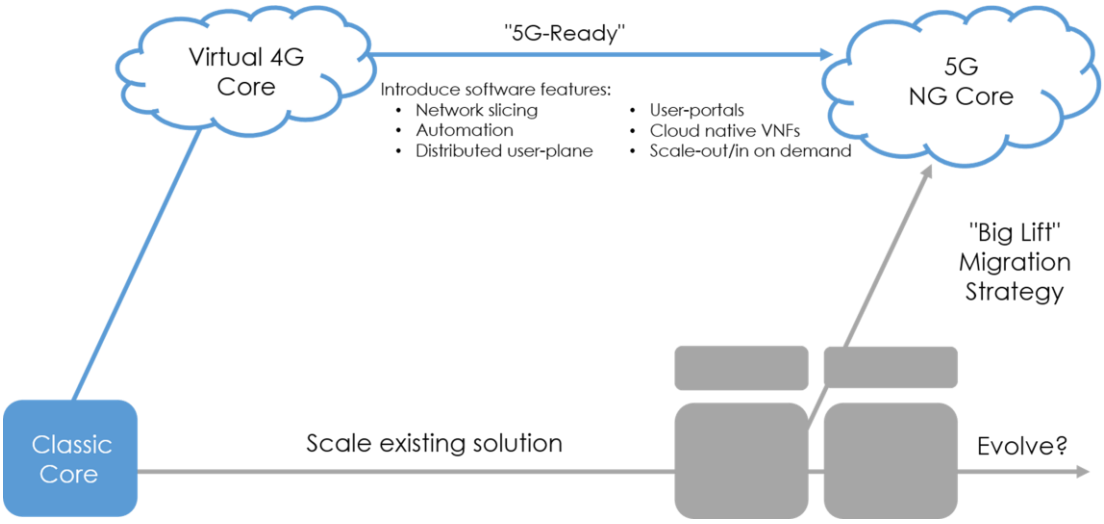
From a practical perspective, for the initial deployment, operators can support 5G radio access on a "5G-ready" EPC and then migrate to a new NG Core over time. In both cases, the "IP services fabric" provides SDN-controlled connectivity and related IP services, across a distributed cloud infrastructure.

Virtualized & Cloud-Native Mobile Core

Operators have made good progress on virtual EPC over the past couple of years. The largest networks now support more than 15 million subscribers (AT&T has discussed this publicly); some progressive operators have started to refresh their main EPC networks using multivendor network functions virtualization (NFV) (e.g., Docomo, Etisalat); and others (e.g., Vodafone) have deployed virtual core elements for IoT services.

Figure 3 shows two migration paths to a "cloud-native" 5G core. The blue line shows a virtualized 4G core as a stepping stone to a "5G-ready" core and then a full 5G core. The grey line shows a more conservative option where the operator scales the classic EPC in the near-term and then makes a bigger leap to 5G later.

Figure 3: A Preliminary Cloud-Based 5G Architecture



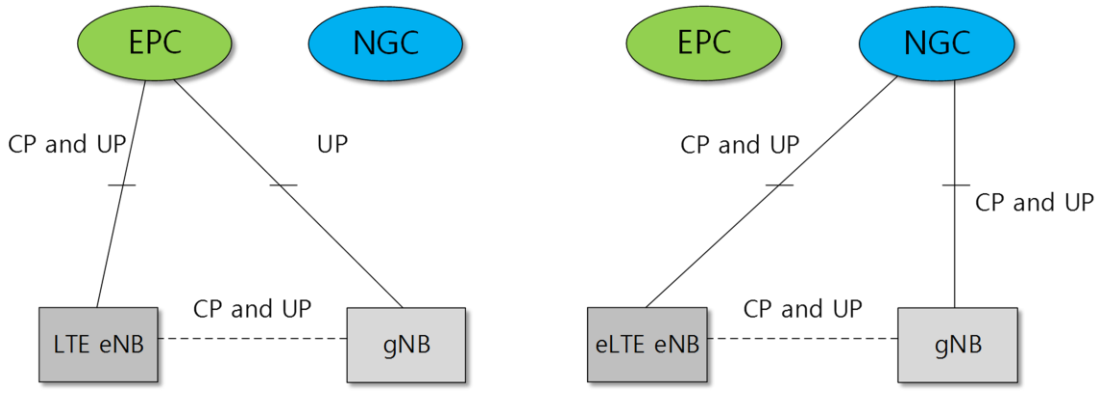
Source: Heavy Reading

There is debate about what exactly "cloud native" means, particularly in the context of stateful telecom network functions. Nevertheless, Heavy Reading's established view is that operators that deploy virtual EPC – even at small scale – gain experience in how these systems work, how to operate them, and how to evolve them, and that in turn will give them a sustainable advantage in the longer-term transition to software-centric 5G mobile networks.

Migrating 5G RAN From EPC to NG Core

A 5G-ready core strategy is determined, in part, by how the operator plans to introduce 5G radio. There are two basic possibilities: operate 5G RAN using an EPC or using a new NG Core. In practice, operators with plans to launch 5G early are likely to start with an EPC and migrate to NG Core over time, as shown in **Figure 4** below.

Figure 4: Migrating 5G RAN from EPC to NG Core



Source: 3GPP TR 38.801 v0.4.0

The diagram is taken from the most recent version of the 3GPP "Study on New Radio Access Technology" (TR 38.801), which will inform development of standards in Release 15. Several permutations of the architecture are under consideration; however, in simple terms, to the left, the new 5G base station (gNB) user-plane interface connects directly to a 4G EPC, while the control-plane functions, such as tracking, paging, etc., are provided by an evolved 4G base station (eNB), which in turn also connects to the EPC. In this scenario, the EPC requires little or no modification, making this a fast and simple way to deploy 5G radio from a core perspective.

Over time, both 4G and 5G base stations can migrate to a new NG Core, which will provide both control- and user-plane functions. At this stage, NG Core becomes the primary core network for 4G and 5G access, as shown to the right in the diagram. This is conceptually similar to how EPC supports 3G and 4G access networks.

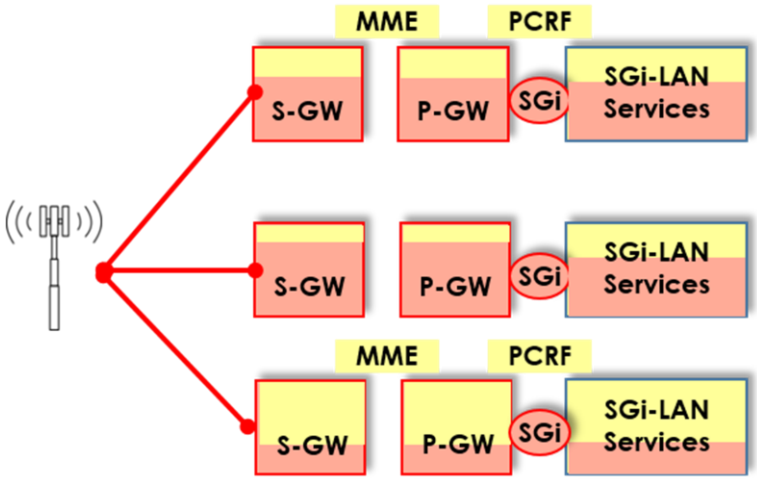
Note that in cases where 5G radio is deployed for fixed wireless access, there is no need for an LTE RAN to provide control-plane functions to the 5G user device; a standard EPC is sufficient, although it would need to provide session management for the 5G access. Some of the first 5G deployments are expected to use this model.

Network Slicing: A Key Bridge to 5G

Network slicing is one of the key bridges between the 4G and 5G core. To support diverse service types, operators will use multiple core networks deployed as "network slices" on a common IP services infrastructure. The idea, shown in **Figure 5**, is to create virtual core network instances (or "slices") dedicated to different services. Each slice can be optimized for the traffic profile and the commercial context of the associated service – for example, IoT, public safety, mobile virtual network operator (MVNO), connected car, voice over WiFi or enterprise services. Network slices can be two dimensional in the sense that they can be both service- and customer-specific.

Figure 5: Dedicated Packet Core Network Slices

- Virtualization enables multi-tenant and single-tenant private core networks
- "Slices" configured according to service type & traffic profile
- Multiple options to steer user traffic into a core network processing path
- Mechanisms in 3G/4G include APN routing, MoCN, DECOR



Source: Heavy Reading

IoT is an example of how operators can use "slicing" to support a 5G-ready core strategy. Since many of the core network design parameters are similar between 4G and 5G, investment in a software-based IoT core network can be made with the expectation that the same core (with software updates) will also support IoT services on 5G in future. Moreover, because devices from both access types will connect to a common IoT core, operators will be able to develop integrated 4G/5G IoT strategies that optimize investment and enable them to go to market with narrowband IoT services before 5G specifications are released and radio equipment is deployed.

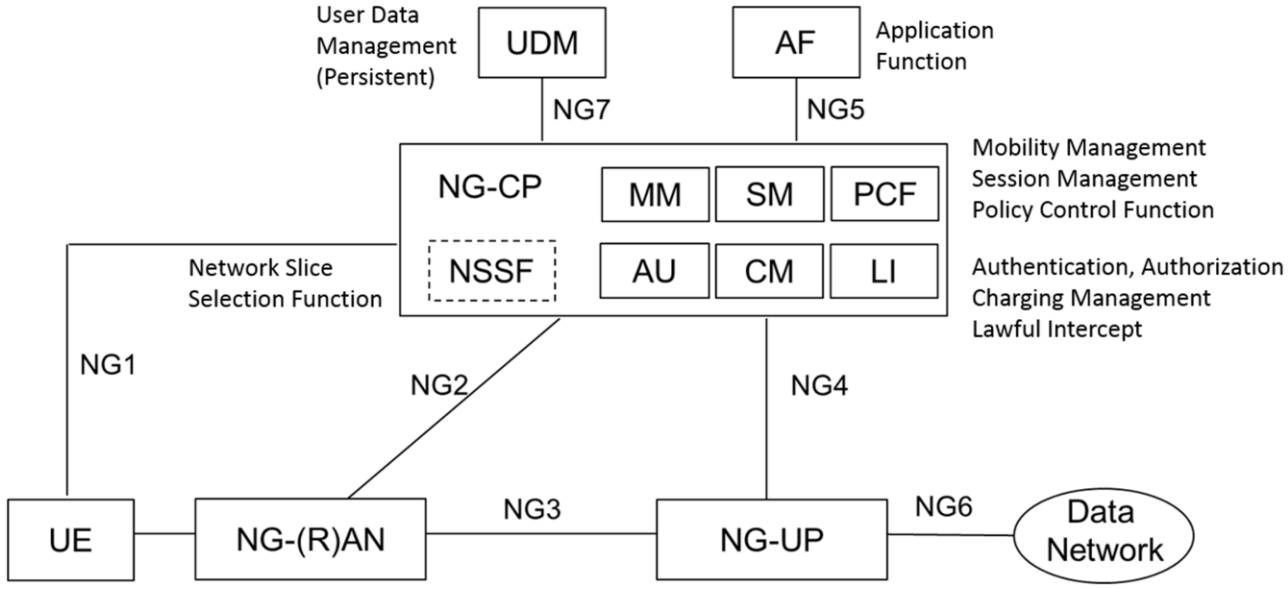
5G will, naturally, bring more capabilities to network slicing. Most importantly, it is expected that the slice will run end to end across the RAN, core and transport network. Radio is typically deployed as a shared resource, which means that 5G slicing will involve advanced self-organizing network (SON) capabilities and hierarchical, slice-aware scheduling on the air interface. On the network side, the operator can use SDN to reserve resources for the network slice on the IP services fabric.

NG Core for 5G

A new core network offers important benefits, particularly for services with demanding performance requirements, and is an important part of the 5G architecture. The high-level view of the NG core network architecture, as it looks in September 2016, according to TR 23.799 (the Technical Report on the NextGen System Architecture) is shown in **Figure 6**.

The chart shows the primary elements of the new 5G system architecture, including the device (UE), the radio access network (NG-RAN) and the core network that comprises the control-plane (NG-CP) and user-plane (NG-UP) functions. The separation of control and user planes is a direct extension of the "CUPS" concept being developed for advanced EPCs in 4G networks and is another example of how operators can invest in 5G-ready core network designs in the near term.

Figure 6: High Level View of NG Core Architecture (Proposed)



Source: 3GPP, Adapted from TR 23.799 v0.8.0, September 2016 (Figure 7.2.4-1)

It is interesting to compare this new functional architecture with the existing EPC. **Figure 7** maps the new NG Core elements to their equivalent in EPC. It shows significant overlap between the EPC and NG Core and provides confidence that today's state-of-the-art EPC is reasonably close to the next-generation core in terms of functionality, albeit that the interfaces and protocols will change/evolve.

Figure 7: Mapping NG Core & EPC Functions

NG Core Network Function	Approximate Equivalent EPC Function
UDM (User data management)	SPR
AF (Application function)	AF
NSSF (Network slice selection function)	MME (DECOR) + HSS
MM (Mobility management)	MME + SGW / MME + SGW-C
SM (Session management)	PGW/PGW-C
PCF (Policy control function)	PCRF
AU (Authentication & Authorization)	HSS/AAA
CM (Charging management)	PGW + PCRF, OCS, OFCS
LI (Lawful intercept)	LI
NG-UP (User-plane function)	SGW + PGW / SGW-U + PGW-U

Source: Heavy Reading, Affirmed, Juniper

Key Capabilities of the Next-Gen System Architecture

In addition to proposed architectures, the Technical Report provides guidance on the features and services NG Core should support. It identifies 19 key issues for the Next-Gen System Architecture, which can be grouped into four categories:

- **Flexible Deployment**: Including network slicing, user-plane network selection, network function granularity (e.g., decomposition of VNFs and service chains), interworking and migration, a policy and charging framework.
- **Flexible Access Support**: Including variable core/access splits, a flexible authentication framework (with varying credentials according to device-type, use case and policy), support for relays and multi-hopping, improved network discovery and selection mechanisms.
- **Connectivity**: Including session management, mobility management, a quality-of-service (QoS) framework and service and session continuity (e.g., across accesses).
- **Adapting Existing Capabilities**: Including network/service capability exposure, multicast and broadcast, IP Multimedia Subsystem (IMS) support, off-network communications (e.g., *ad hoc* networks for public safety).

Implementation of these capabilities is tightly linked to both the design of the NG Core and the underlying IP service infrastructure. The 3GPP specifications do not reference connectivity services, but, in practice, there is a close relationship.

A Distributed 5G Services Fabric

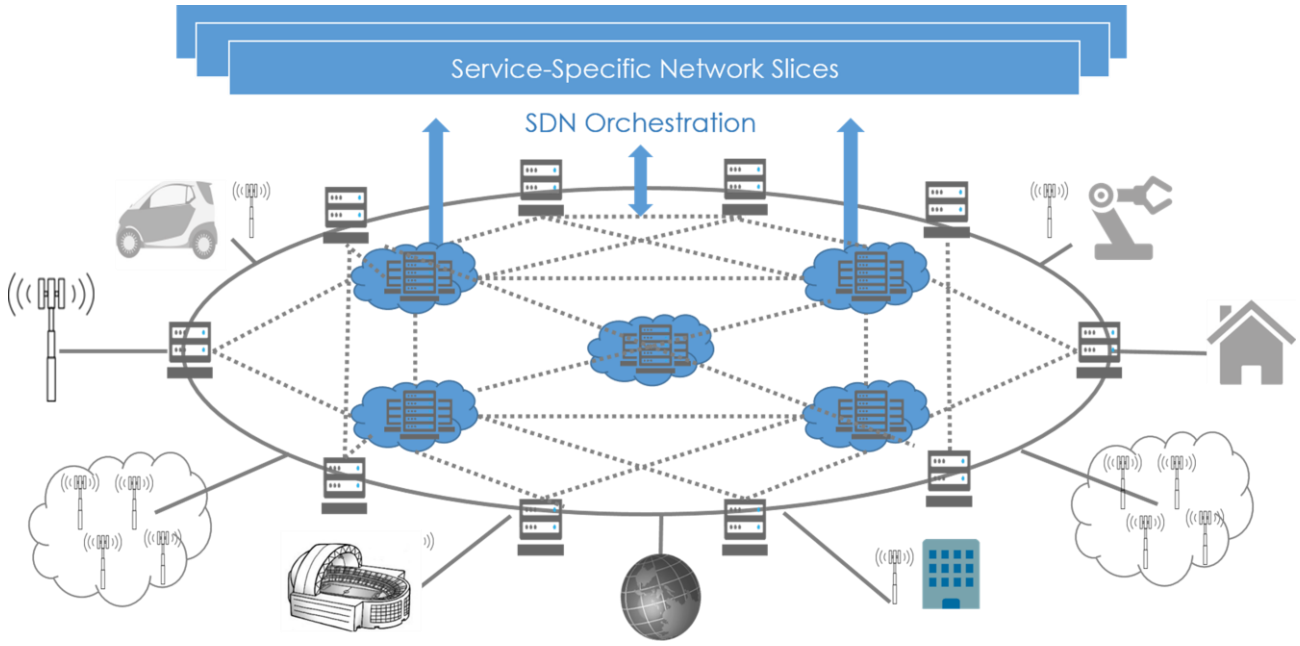
The infrastructure onto which 5G will be deployed should support multiple, demanding use cases. This drives a need for a high-performance wide-area IP services fabric controlled (or "orchestrated") by SDN. This IP services fabric should provide connectivity between many different distributed data centers in a meshed architecture that provides resiliency and scalability, and should be programmable such that it can support dynamic, service-specific network slices. In effect, the IP services fabric makes distributed centers act in a unified manner – i.e., behave as one integrated data center. The concept is shown in **Figure 8**.

The design of the distributed IP services fabric is formed by the variety of use cases, the requirement for low latency and high availability, and the need to scale efficiently. In tandem, the mobile network architecture will evolve to a more distributed model. To meet 5G service requirements, NG Core will be deployed using the "CUPS" model, with user-plane nodes hosted at distributed data center locations, and control-plane nodes at more centralized locations, interconnected by the IP services fabric.

Service orchestration will generally be domain specific, with SDN-controlled IP services, cloud resource management, NFV lifecycle management and end-user service orchestration, all operating quasi-independently in a layered architecture. Coordination between the layers will use "cross domain orchestration" to ensure a network slice contains all the networking components needed to deliver the service.

Depending on how the architecture evolves, there will be a need to support highly accurate, 1588v2-based timing (frequency and phase) within this IP services fabric. Virtual RAN/Cloud RAN is an example on the network side that benefits from accurate network timing. On the customer side, it is expected that some market segments, such as the financial industry, will also require timing to be incorporated into network slices.

Figure 8: IP Services Fabric for 5G Networks



Source: Heavy Reading

Distributed Data Centers

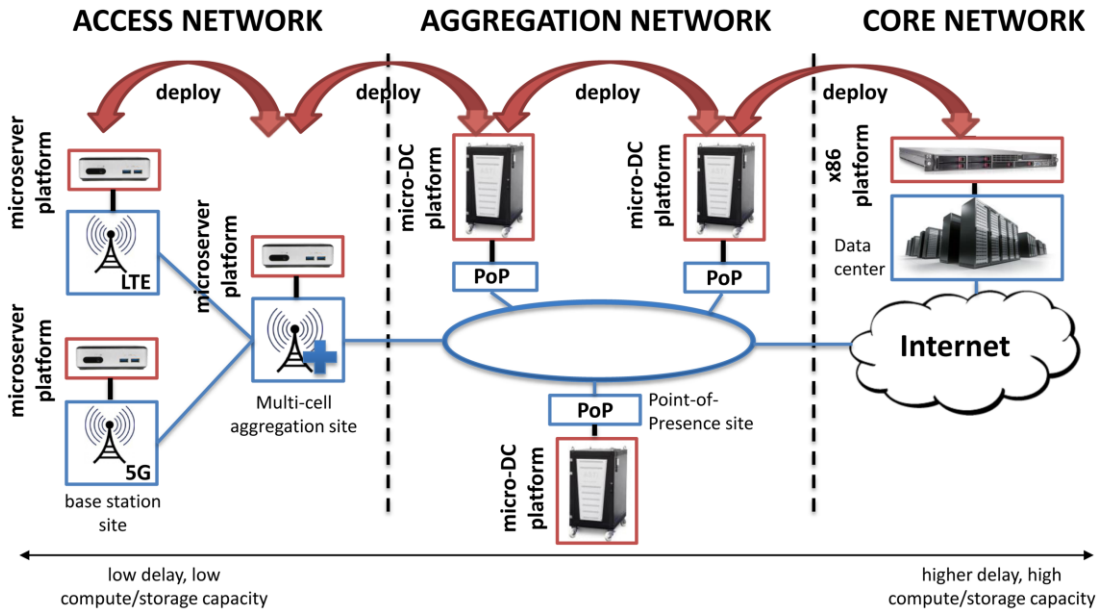
To provide access network services, telecom operators have a physical footprint – in the form of central offices, base station controller sites and transport aggregation sites – that they can convert into distributed or micro data centers.

Depending on the service requirements, these data centers can be used to terminate access connections using virtualized edge functions and become the obvious place to host latency-sensitive content and applications. For centralized cloud providers, this capability is harder to replicate unless they build out an extensive physical footprint or partner with an access provider, and, therefore, this strategy provides operators with a potentially important and sustainable competitive advantage.

The "edge cloud" model will be critical to 5G and is the subject of detailed R&D across the industry. As an example, a preliminary architecture developed by the SuperFluidity project (a European 5G research project) is shown in **Figure 9**. Work on this particular project, which is academic in focus but contains several industrial partners, is scheduled to finish by the end of 2017.

As the name implies, the project aims to achieve superfluidity in the network: the ability to instantiate services on the fly, run them anywhere in the network (core, aggregation, edge) and shift them transparently to different locations. This is the essence of the "IP Services Fabric" for 5G. In this new network, operations will also need to move from a "red light, take action" model to an automated model where the orchestration and associated next-gen operations support system (OSS) detect and fix problems with the IP and NFV layers, and re-optimize the network accordingly.

Figure 9: A Preliminary Converged Cloud-Based 5G Architecture



Source: SuperFluidity

Central Office Re-architected as a Data Center (CORD)

There are several industry and service provider initiatives to define the architecture for SDN-enabled distributed data centers. The Central Office Re-architected as a Data Center (CORD) initiative hosted by the Linux Foundation is one example of how operators aim to make better use of these assets.

According to the CORD initiative, AT&T alone operates 4,000 to 5,000 central offices, each serving 10,000 to 100,000 residential, enterprise and mobile customers. These central offices contain fragmented vendor hardware with multiple physical appliances installed per site (AT&T said it has 300+ unique appliances deployed in its central offices nationwide). The opportunity is to re-architect these central offices to support edge cloud infrastructure and deploy VNFs in place of appliances.

Each CORD location will be connected using an SD-WAN, making it possible to load-balance content and NFV workloads across the distributed cloud using the same SuperFluidity concepts discussed above. In the CORD case, the focus is on the ONOS controller, but in the sense that this is a generic architecture, multiple SDN controller options are viable and attractive. Note that for I/O-intensive workloads, the selection of an edge location should consider the physical transport resources available at the site. It would not make sense, for example, to deploy a 5G user-plane node for fixed wireless access in a central office that does not have a high bandwidth connection to the Internet.

In March 2016, the Linux Foundation announced the M-CORD initiative for mobile operators, backed by AT&T, SK Telecom, Verizon, China Unicom and NTT Communications. It highlighted three key aspects of the architecture, saying that it:

- Will use the same CORD principles of elastic commodity cloud and SDN to bring data center economies and cloud agility to the mobile edge.
- Will demonstrate integration of disaggregated/virtualized RAN, disaggregated/virtualized EPC and mobile edge services.
- Will partner with the SDN controller groups to accelerate adoption of open source SDN and NFV solutions and realize the benefits of the cloud.

Given the timing of the work, we expect M-CORD and similar initiatives to turn their focus toward "5G-ready" core networks as the architecture, interfaces and protocols for NG Core become more clear. A common CORD and M-CORD implementation, with the same architecture and foundational technologies, will create a good foundation for fixed-mobile network convergence, enabling access agnostic services – an important objective of many operators pursuing 5G.

Mapping Mobile Core to Distributed Cloud

The NG Core for 5G and EPC for advanced 4G networks must be mapped to the distributed cloud architecture. One approach would be to simply deploy more packet gateways (and mobility controllers) at the edge of the network to meet capacity and performance demands.

The challenge with this is that today's centralized packet core deployments are characterized by complex integration with surrounding network functions, such as policy, charging, IMS, SGi-LAN and routing services. By moving this model to the edge, the operator would, in effect, have to "distribute complexity," which is costly to deploy and, in particular, to manage. To meet 5G performance, scalability and automation requirements, a new architecture for packet core is needed that will make operation in the edge cloud infrastructure simpler and faster.

There are many aspects to this new architecture. Part of the solution is CUPS, as is currently being developed for 4G-LTE core networks. This involves extracting the control-plane functions from the gateway to leave a simpler, user-plane node. The gateway thus is "split" into S/PGW-U and S/PGW-C components that can that can scale independently, as shown in **Figure 6** above. A key benefit of the architecture is that the control plane, and all the associated complex interactions, can be centralized, while the user plane is distributed across the IP services fabric and scaled as required by the traffic load. This is shown in **Figure 10**.

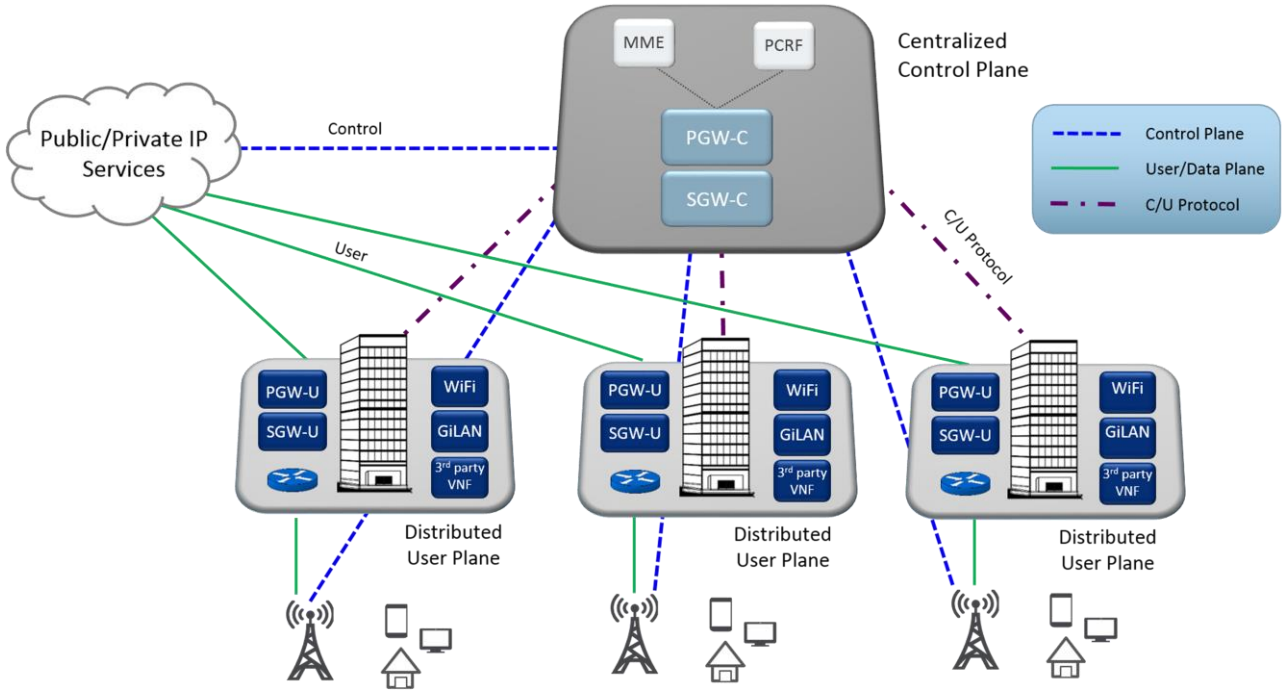
Depending on the scaling needs, the S/PGW-U functions deployed at the edge data center can be implemented in several different ways: on a router, on a white box switch (potentially), on an existing PGW platform, as a virtualized function, or as part of a vRouter. Virtualized user-plane nodes can more easily be placed at the optimal location, as determined by the use case and this flexibility is a strong argument to deploy S/PGW-U nodes as VNFs rather than as hardware functions.

The S/PGW-C components would similarly be deployed as virtualized functions on a cloud platform, typically at a more centralized location. There may be an opportunity to collapse MME, SGW-C, PGW-C functions and perhaps other 3GPP control-plane functions into some form of unified "mobility controller" node. This has the potential to simplify operations and this same model is being pursued for NG Core.

In this architecture, virtualized SGi service LAN components can be placed where appropriate for the traffic coming from the distributed user-plane nodes. For example, some SGi functions could be on router-based compute blades, or on COTS x86

servers deployed at the distributed site, while others could be in the central data center or close to Internet peering points. The selection of the site for the SGi functions is carried out by the central orchestration system, which steers traffic into service chains on a per subscriber, application, bearer, device or combination basis.

Figure 10: Distributed User-Plane at Edge Data Centers



Source: Affirmed Networks

This model of discrete control- and user-plane functions is expected to be fundamental to 5G and the NG Core. In this sense, CUPS can be viewed as an important part of a 5G-ready investment strategy.

A 5G-Ready Core Is Now a Priority

The industry now has a reasonable view of 5G service requirements, and progressive operators with aggressive deployment timelines are now working to prepare their networks for rapid deployment of 5G when equipment is available. This is driving investment in the critical IP services network needed to connect the edge cloud locations that will run 5G network functions, content and services.

On the mobile core side, development of cloud-native, service-orientated core networks for advanced 4G and 5G networks is underway. Network slicing provides a conceptual bridge between 4G and 5G investment and facilitates a faster insertion of new services into the network. Similarly, CUPS provides a reference for NG Core and the new 5G network architecture. With a 5G-ready technology strategy, operators can prepare for 5G service launch in a way that optimizes their investment in next-generation IP and mobile core platforms over the next three years.

About Affirmed Networks

Affirmed Networks' NFV solution has become the standard for the world's top mobile operators, who are embracing new business models and building new revenue streams by making the transition to virtualized architectures. The company's technology portfolio includes the Affirmed Mobile Content Cloud, the Affirmed Wi-Fi Gateway (serving as a TWAG/TWAP and an ePDG), Affirmed Service Automation Platform (ASAP) and Affirmed Virtual Probe and Analytics Solution. These virtualized solutions have come to represent the present and the future of virtualized mobile networks with extreme scalability, remarkable flexibility, comprehensive network orchestration and future-proof solutions for a 5G-ready architecture. Please find more information at www.affirmednetworks.com.

About Juniper Networks

Juniper Networks (NYSE: JNPR) is in the business of network innovation. From devices to data centers, from consumers to cloud providers, Juniper Networks delivers the software, silicon and systems that transform the experience and economics of networking. The company serves customers and partners worldwide. Additional information can be found at www.juniper.net, or connect with Juniper on Twitter and Facebook.