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

HEAVY READING

WHITE PAPER

Service-Oriented 5G Core Networks

A Heavy Reading white paper produced for Huawei Technologies Co. Ltd.

HUAWEI

AUTHOR: GABRIEL BROWN, PRINCIPAL ANALYST, HEAVY READING

NEW SERVICES & NEW VALUE CHAINS

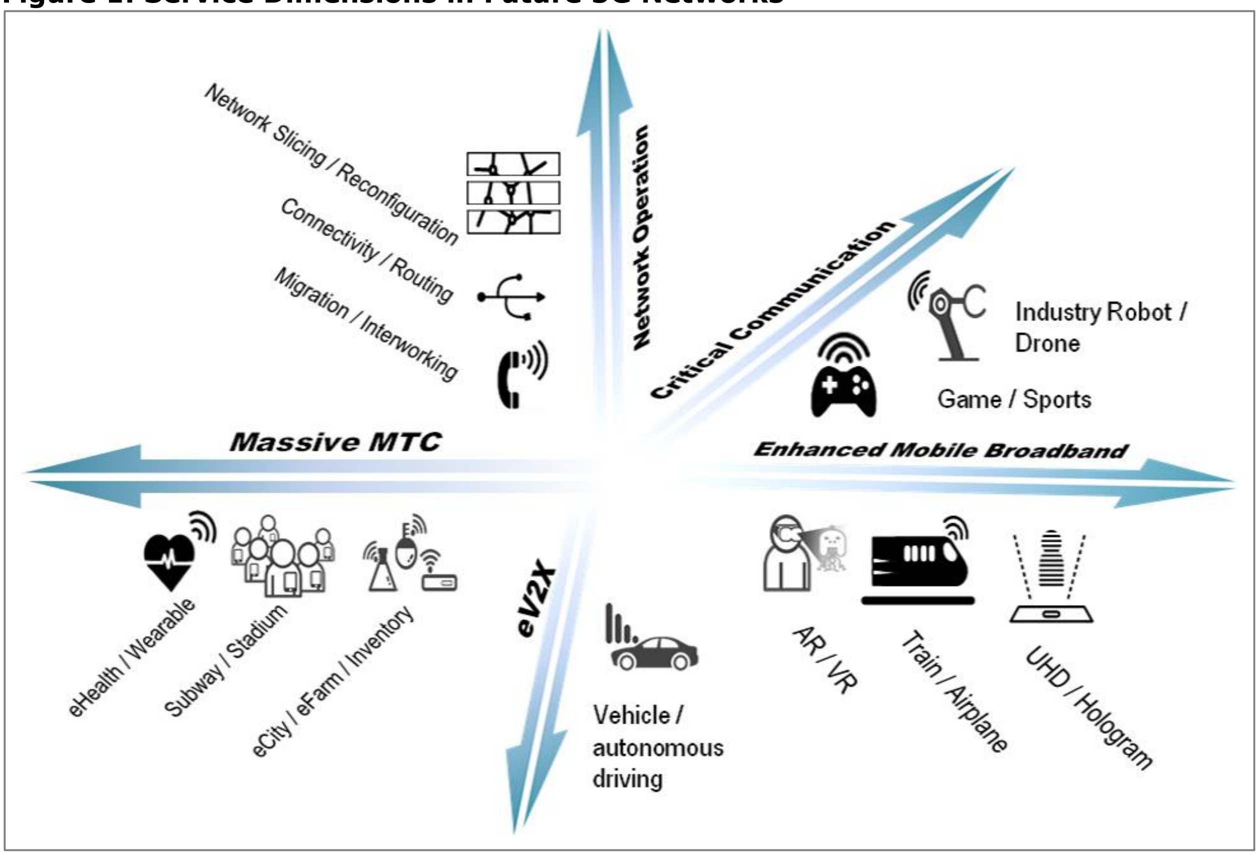
Operators worldwide are evaluating the transition to 5G networks, with a view to commercial launch from 2019 onward. To support the wide range of services envisioned for 5G requires a new core network, known as Next-Generation Core or NG Core. This white paper discusses the development of the NG Core ahead of specification in Release 15 and Release 16. It covers what it means to design and operate a service-oriented 5G core network, as well as the status of NG Core development at the start of 2017.

One Network; Many Services

The service-oriented 5G core network is based on the premise that 5G will support a very diverse range of services with very different performance requirements. In preparatory work ahead of standardization, the industry identified three different service categories for 5G: enhanced mobile broadband (eMBB), massive machine type communications (mMTC, a.k.a. massive-scale IoT) and ultra-reliable low-latency communications (UR-LLC).

This incorporates use cases such as industrial control, augmented reality (AR)/virtual reality (VR) and connected cars, as shown in **Figure 1**. The objective is to use end-to-end network slicing to support these diverse services and industries on one physical network infrastructure. In this way, operators can enable new services in adjacent sectors and insert their networks into new industrial value chains.

Figure 1: Service Dimensions in Future 5G Networks



Source: 3GPP SMARTER

5G Phase 1 & 5G Phase 2

The vision for 5G is expansive and sets out service scenarios that will play out over a decade – or more – of operation. To deploy a new generation of wireless technology is complex and technically challenging. To help the introduction of 5G, the industry has adopted an approach that will see 5G put into commercial operation over two phases:

- **Phase 1** will focus primarily on the eMBB use case, with some aspects of mission-critical communications also supported. Broadband access is the mainstay of mobile operators' business with proven customer demand. Prioritizing eMBB will expedite early deployment by operators with ambitious launch schedules. Phase 1 corresponds to 3GPP Release 15, which is scheduled to freeze in mid 2018. The first commercial networks could be operational from 2019.
- **Phase 2** corresponds to Release 16, currently scheduled to freeze at the end of 2019. It will introduce new features related to mMTC and UR-LCC services, as well as updates and enhancements to Release 15 capabilities. Release 16 is likely to be the version of 5G submitted to the International Telecommunication Union (ITU) for approval as an IMT 2020 technology. Phase 2 work can begin while Phase 1 is being specified.

The next-generation system architecture and associated core network (NG Core) straddles both releases. Release 15 should be sufficiently featured to enable operators to launch commercial 5G radio access for eMBB – either using a 4G core, known as non-standalone (NSA) mode, or with the new 5G core, known as standalone (SA) mode.

Naturally, the 5G radio access network (RAN) and core will continue to evolve over later releases. One of the defining features of 5G, relative to earlier generations, is forward compatibility. This is the idea that network elements will be designed with the expectation that new, as yet unknown capabilities will be introduced in future, so that 5G can meet the requirements of application developers not just in 2020, but in 2025 and beyond.

Operator Trials & Deployment Plans

With standardization work now underway, many operators around the world are trialing prototype and pre-standard 5G network technology. Some of the leading examples, based on public statements, are as follows:

- **Deutsche Telekom:** Has demonstrated end-to-end network slicing across 5G RAN and core. It is also investing in significant outreach beyond eMBB into vertical industries – for example, in the automotive and power-grid sectors.
- **AT&T:** Has an extensive trial program for 28 GHz 5G with test networks deployed in several U.S. markets, using in-house R&D and prototype equipment, as well as vendor partners. More than any other operator, AT&T is paying close attention to the overall architecture and is already deploying the software-defined networking (SDN) fabric and cloud infrastructure needed to host and connect NG Core functions.
- **Verizon:** Already operates 28 GHz trials in several U.S. markets. The initial focus is on fixed access, but mobile is also part of the development work. The pre-standard Verizon 5G Technology Forum has been influential on the industry timeline and on some key technical decisions. The first fixed wireless services could launch at small scale as soon as 2018.

- **KT:** Has a very progressive 5G deployment plan and is an integral part of South Korea's national plan for 5G development. The operator runs the "KT 5G-SIG," which focuses on 28 GHz millimeter-wave technologies for mobile 5G. It intends to provide live 5G service at the Winter Olympics in Pyeongchang 2018, with a view to commercial launch in 2019.
- **NTT Docomo:** Probably the most extensive 5G development and trial program of any operator in the world. It is engaged with all major vendor product development and has an in-house R&D division. It plans to launch commercial service at the 2020 Summer Olympics in Tokyo.

There are, of course, lots of other operators that have publicly discussed their 5G development plans, including China Mobile, Vodafone, Orange, Telefónica, TeliaSonera and more.

SERVICE-CENTRIC CORE NETWORK

Two modes are under development that will enable operators to deploy 5G radio access: non-standalone (NSA) mode, using a 4G Evolved Packet Core (EPC), and standalone (SA) mode, using a new NG Core. There is no clear consensus regarding which will prove most popular for early deployments – both have their merits – but there is agreement that, over time, 5G will require a new core network. NG Core is being designed to serve the range of services envisioned for 5G and therefore too fulfil the service-oriented core network remit. Network slices with specific characteristics to support mission-critical communications or massive-scale IoT, for example, will require a highly automated, highly configurable, cloud-based infrastructure. And ultra-high throughput from 5G and fixed access will necessitate a new architecture that distributes user-plane processing closer to the edge.

Cloud-Native NG Core

The new NG Core should be "cloud native" for agility and programmability reasons. A new core network offers an opportunity to design virtual network functions (VNFs) that operate under a cloud management and service orchestration system. Some of the main differences between classic VNFs derived from virtualized appliances and "cloud native" networks are shown in **Figure 2**.

Figure 2: Cloud-Native VNF Design Principles

Virtualized	Cloud-Native
Limited, workflow/script-driven automation	Extreme model-driven automation
COTS, Linux, hypervisors, OVS/VPP, OpenStack	PaaS, hyper-converged servers, distributed DCs
VNF software derived from virtualized appliances	VNF designed for cloud (e.g., microservices)

Source: Heavy Reading

This is a major transformation for operators, because it changes how they operate services. In particular, it moves some of the key resiliency and failover mechanisms from the network equipment (typically deployed in N+N hardware configuration) to the cloud. In service provider networks, which are mission-critical infrastructure, this is far from trivial; however, success will open the door for far greater use of cloud technologies across the network.

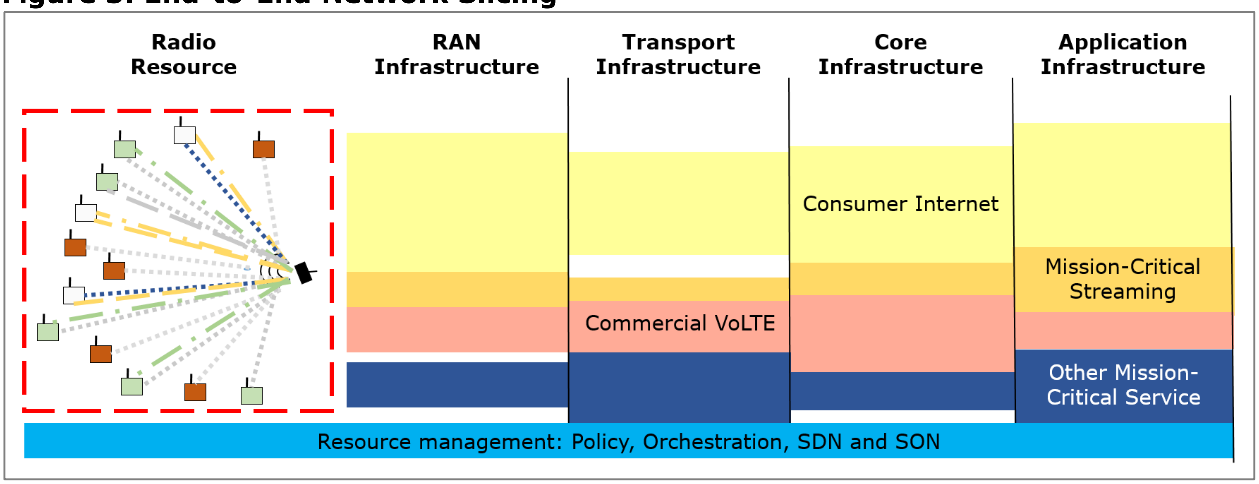
There are already several vendors and operators deploying cloud packet core for 4G. In combination with new architectures, such as control and user plane separation (CUPS), experience with a cloud-based 4G core can give operators an advantage in the deployment of NG Core. Specifically, factors such as service automation, resource orchestration and resiliency will transfer from cloud EPC to NG Core.

Network Slicing & Automated Lifecycle Management

One of the important commercial objectives of 5G is to enable operators to support multiple service types on a common physical network infrastructure. In this way, 5G will act as an enabler to growth markets, such as Industry 4.0, connected car, VR, and so on. As noted above, each of these services has specific performance requirements.

Network slicing is proposed to support these services on a common infrastructure – although in practice, it is likely that discrete equipment optimized for the use case may also be used in some cases. This requires the performance parameters associated with the slice to be supported across the network, from the user equipment (UE) and radio to the cloud-hosted application, as shown in **Figure 3**. In some cases, this will involve bringing the application closer to the users (for example, using edge computing); in other cases, the RAN or transport network will require a particular configuration. NG Core, which manages sessions, quality of service (QoS), security, policy, etc., sits at the heart of the process.

Figure 3: End-to-End Network Slicing



Source: Heavy Reading

There is some precedent for "slicing" in mobile networks with APNs and DECOR, but there is a need for greater granularity and flexibility in 5G, and specifically for network slice lifecycle management. Cloud and network functions virtualization (NFV) provide a reference with platform automation and service orchestration. And on the network side, SDN-controlled micro-segmentation and virtual network overlay techniques used in the data center will evolve to the wide-area network (WAN).

The details of how slicing will work in NG Core are being determined during the normative phase now underway; however, the broad outline is that slice selection is network-controlled, but with the UE able to provide network slice selection assistance based on policy. The UE may belong to more than one slice on the same RAN and core network.

Multiple Access (e.g., for Converged Operators)

The mobile packet core has supported non-3GPP access (typically WiFi) for some years, and recently several operators have deployed successful WiFi calling services using VoLTE core networks. The challenge is that WiFi access is not native to the 4G architecture; it is an add-on function that requires deployment of additional ePDG and IPsec equipment and specialist integration with the EPC. This limits its usefulness and scalability.

The demand in NG Core is to develop an access-independent core network with a common user-plane access and, perhaps most importantly, a control plane that can support common AAA and policy control. This is important for converged operators seeking to offer "follow-the-user" services across fixed and mobile access – for example, policy control and enforcement may need to be applied to a user moving between home WiFi and wide-area cellular.

Multi-access support is an ambitious target for a 3GPP-defined architecture, because it requires participation from other standards groups in the fixed access world. There is liaison between these groups and the 3GPP, of course, and there is some precedent in that 3GPP-specified IMS and policy control, for example, have been adopted in fixed networks. Nevertheless, real-world multi-access deployments are harder to deploy, and more strategic in nature, than making provision for this in standards. Activity on the ground will determine how successful this will be.

In NG Core, the formal separation of control and user plane is important to an access-agnostic architecture. Specifically, the specification of an Access and Mobility Management Function (AMF), discussed in the following section, is a user-plane function that may or may not include mobility management capabilities, depending on the use case.

Performance & Scalability (e.g., for UR-LLC or mMTC)

Performance and scalability are critical to 5G and NG Core. In the same way that the 5G radio interface should be scalable across many service types, the NG Core should also support massive-scale IoT, mission-critical services, low latency and high throughput.

With high-performance radio access – potentially multiple gigabits per single base station – throughput performance is critical. This is far beyond what has been common in EPC and will require a new approach to user-plane forwarding, particularly where multiple radios are aggregated in the transport network. It is complicated by the desire to use virtualized, software-based forwarding, which generally performs worse than hardware-based packet processing. Technologies such as Vector Packet Processing (VPP) will help.

Greater distribution of "forwarders" (equivalent to gateways in EPC) to specific locations – for example, to enterprise campuses or stadiums – will help operators meet processing volume demand. Equally importantly, distribution will enable new, low-latency services by processing data within a few milliseconds of the user, which in turn will support a range of mission-critical services that rely on short response times. A further level of performance challenge is low latency with high-speed mobility – for example, automated vehicles would need between 5 ms and 20 ms latency at 80 km/h.

Connection density, often expressed as number of devices per km^2, is a critical performance metric. Of particular note, the 5G core network will need to support connectionless operation to scale to the high number of connections needed for massive-scale IoT. To enable different parameters to be configured in the core, according to the type of IoT service (low power, high mobility, etc.), network slicing will be used. Connectionless operation is a major evolution from 4G and will have significant impact on the core, making it part of Phase 2. Similarly, support for mesh and relays will be needed in the core for services such as public safety, industrial control and *ad hoc* networks. However, because the radio interface aspects related to relaying will not be specified until Phase 2, this is also a Release 16 capability in the core.

NG CORE ARCHITECTURE

The new 5G network architecture is being specified in 3GPP Release 15. In Release 14, study groups undertook significant preparatory work and produced several technical reports (TRs) that will inform the normative phase. This process has identified an overall architecture and set of interfaces that will be developed further in Release 15 and Release 16, starting in January 2017.

Defining the Next-Generation Architecture

The definition of the next-generation architecture is the responsibility of the System Architecture (SA) Technical Specification Group on Service and System Aspects. The most influential sub-groups are SA1 (Services) and SA2 (Architecture). However, the overall architecture is heavily influenced by decisions made by the Radio Access Network Technical Specification Group, which is responsible for the definition of the functions, requirements and interfaces of the radio and, therefore, the interfaces these radios present to the core. The two study areas that are influential to the new architecture are:

- The study on "RAN Architecture and Interfaces" (TR 38.801) is focused on the RAN, but decisions made here have significant impact on the core. For example, NSA and SA modes are decided by RAN. Even more fundamentally, specification of the N1 interface to the core depends on the radio architecture.
- The study on "Architecture for Next Generation System" (TR 23.799) sets out the requirements for NG Core and has identified an architecture to take forward to the next phase, in which work in the "System Architecture for the 5G System" (TS 23.501) group will define the actual architecture in Release 15.

NG Core Functional Elements

The functional elements of the next-generation architecture and NG Core have been identified in TR 23.799. In some cases, they have an equivalent in 4G EPC, while in other cases there are significant differences. Perhaps most notably, access control and session management are combined in EPC, but separated in NG Core for greater flexibility and scalability, and to better support fixed access. The major components of NG Core are listed below:

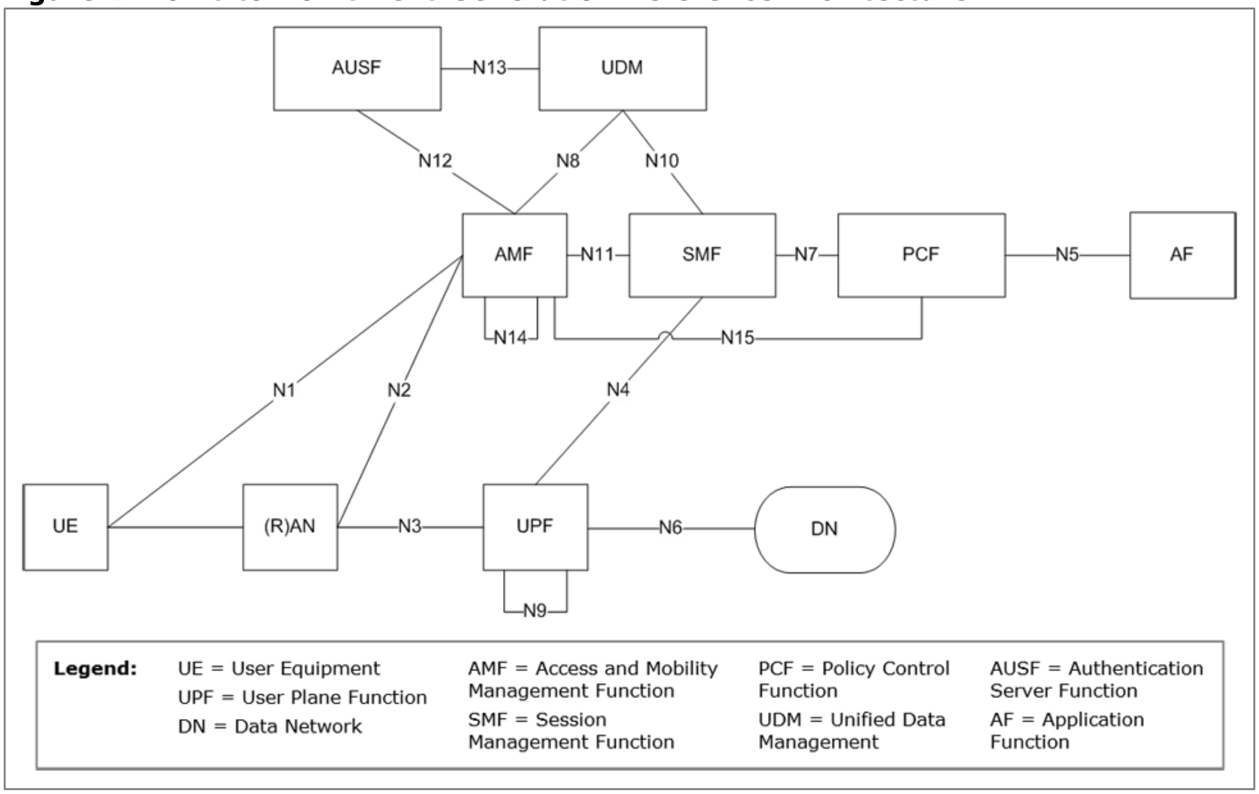
- **Access and Mobility Management Function (AMF):** Manages access control and mobility. The AMF will likely also include network slice selection functionality. Where fixed access is concerned, mobility management will not be needed in the AMF.
- **Session Management Function (SMF):** This sets up and manages sessions according to network policy.
- **User Plane Function (UPF):** UPFs can be deployed in various configurations and locations, according to the service type.

- **Policy Control Function (PCF):** This provides a policy framework incorporating network slicing, roaming and mobility management. Equivalent to a PCRF in 4G.
- **Unified Data Management (UDM):** Stores subscriber data and profiles. Similar to an HSS in 4G, but will be used for both fixed and mobile access in NG Core.
- **NF Repository Function (NRF):** This is a new functionality that provides registration and discovery functionality so that network functions (NFs) can discover each other and communicate via application programming interfaces (APIs).

Point-to-Point Architecture

There are two representations of the next-generation architecture in TR 23.799: a point-to-point architecture, shown in **Figure 4**, and a service-based architecture, shown in **Figure 5**.

Figure 4: Point-to-Point Next-Generation Reference Architecture



Source: 3GPP TR 23.501, January 2017, Figure 4.2.3-2

Both models comprise, more or less, the same functional elements; both offer a formal split of control and user plane; and both support many of the same interfaces. The critical N1, N2, N3, N4 and N6 interfaces, for example, are the same in each case.

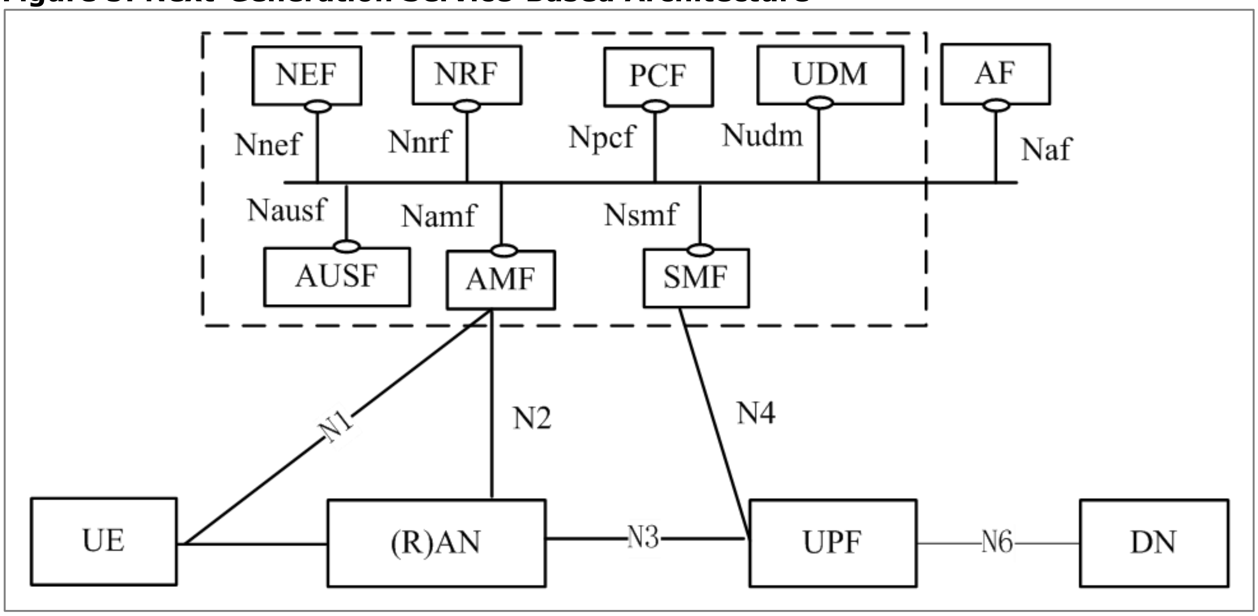
The point-to-point architecture is closer to a traditional 3GPP architecture, in that it defines functions and interfaces between them. It is relatively familiar and is likely to be the architecture used in the first NG Core deployments. The drawback of this model is that it can be complex to add new network elements/instances, because this often requires the operator to reconfigure multiple adjacent interfaces. Over time, it could evolve to the service-based model using reusable APIs between control- and user-plane functions.

Service-Based Architecture

The service-based architecture incorporates the same functional elements and the same user-plane processing path between the UE and external data networks. The major difference is in the control plane, where, instead of predefined interfaces between elements, a services model is used in which components query an NF Repository Function (NRF) to discover and communicate with each other (the NRF is not used in the point-to-point model).

This is closer to the cloud-native networking concept, in which libraries of functions can be requested from a VNF catalog and composed into end-to-end service chains on demand. This software-oriented architecture (SOA) model enables short time to market for new services and greater flexibility for system updates. The challenge for early 5G deployments is that the network cloud platform, and associated operating models, are not mature enough for production networks. Therefore, this may be a Phase 2 deployment.

Figure 5: Next-Generation Service-Based Architecture



Source: 3GPP TR 23.501, January 2017, Figure 4.2.3-1

Release 15 & Release 16

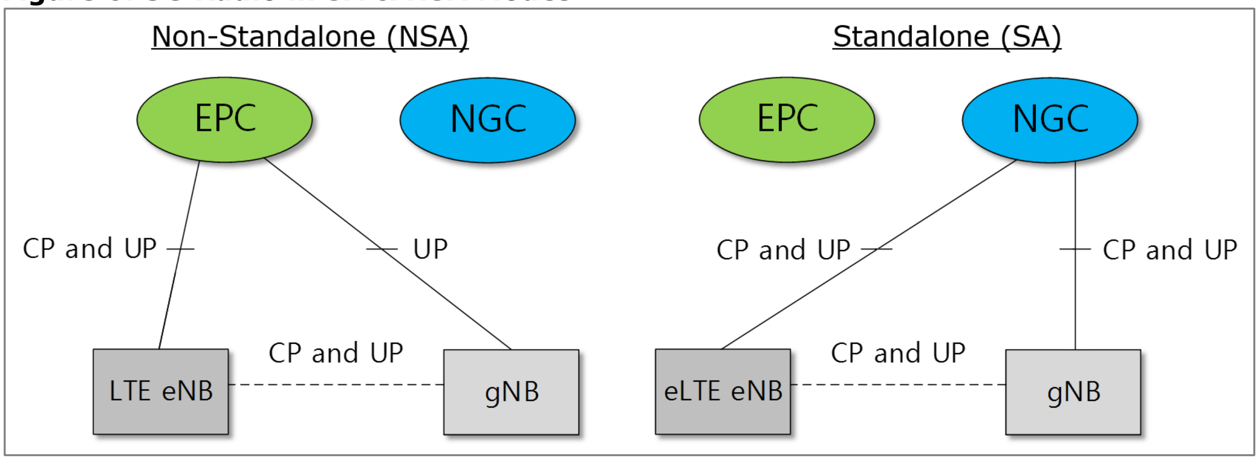
Not all the required NG Core capabilities can be included in Phase 1 (Release 15), which is in some ways a "bare bones" release intended to make early deployments possible. A good number of features have been identified, but specification has been deferred until Release 16 (Phase 2). Some examples are:

- Phase 1 will support 3GPP and untrusted non-3GPP access, but NG Core will not be natively multi-access until Phase 2.
- The first release will support a classic hub-and-spoke cellular topology; more advanced mesh and relay capabilities will come on Phase 2.
- Access network selection mechanisms to determine the best available connectivity according to need or policy are also Phase 2 items; Phase 1 will use EPC-like access selection.

DEPLOYING 5G NR USING EPC & NG CORE

Despite the industry demand for multi-access NG Core, and despite standard interfaces to decouple RAN and core, the deployment of the 5G core network is tightly linked to the introduction of services over 5G radio. As disused above, some operators have very aggressive deployment schedules and are looking for ways to enable commercial launch before specifications are completed. There are two major ways being developed to deploy 5G radio: non-standalone (NSA) mode and standalone (SA) mode. These are shown in **Figure 6**.

Figure 6: 5G Radio in SA & NSA Modes



Source: 3GPP

In NSA mode, the 5G radio (known as gNB) is colocated with an LTE base station (eNB) and connects to the 4G core (EPC) over the standard S1-U interface for user-plane traffic. Control-plane communications (e.g., for session and mobility management) between the network and device remains on the LTE radio and, therefore, the 4G core. In this model, the 5G radio acts as a secondary serving cell to boost throughput and capacity. It is in some ways similar to the use of LTE-Licensed Assisted Access small cells in combination with the licensed LTE anchor. This does not require the NG Core and is attractive to some operators.

To deploy 5G in SA mode requires a new core – the NG Core. This option had been thought of as a Phase 2 (Release 16) item because of the challenges in specifying a new core in a short timeframe. However, over the past year, there has been a push to accelerate NG Core development, and several operators now look set to pursue that model for initial 5G deployment.

In the short term, this is most attractive where the operator will introduce a 5G-only service in a limited geographic area, without 4G interworking. In the medium term, there will need to be integration with the EPC to enable mobility between 4G and 5G access and integration with the evolved LTE RAN, which will connect directly to NG Core.

In the longer term, it is expected that the NG Core will be the common core network for all access types. With a cloud-based EPC and NG Core, operators are able to switch investment from EPC to NG Core in line with the migration rate of the subscriber base. Moreover, to offer advanced new services, such as those that require high mobility and low latency, or access-agnostic, follow-the-user services, the operator will need to deploy NG Core.