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

HEAVY READING
WHITE PAPER

4G & 5G-Ready Network Slicing

A Heavy Reading white paper produced for Affirmed Networks

AUTHOR: GABRIEL BROWN, PRINCIPAL ANALYST, HEAVY READING

STRATEGIC IMPORTANCE OF NETWORK SLICING

The ability to support diverse services, each with specific performance requirements, on a common network platform is a powerful idea, with compelling commercial opportunities. Network slicing is fundamental to this vision because it enables operators to configure virtual network instances, optimized to the specific functional requirements of a customer or application, more quickly and at lower cost than building traditional dedicated networks.

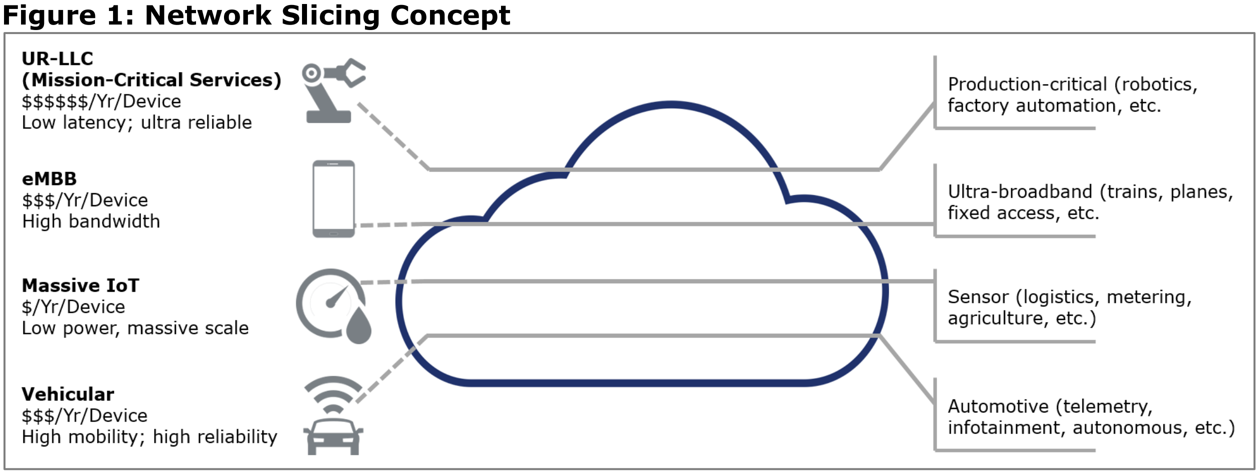
This white paper discusses why network slicing is important, how operators can use virtualization to support different industrial sectors, why this capability should be introduced in today's networks, and how it will contribute to commercial success in 5G. The paper focuses on using Access Point Names (APNs) and virtualized Evolved Packet Core (vEPC) to serve different user groups in 4G networks and looks ahead to the migration to end-to-end network slicing in 5G. It argues that virtual core networks and slicing capabilities are critical for operators seeking to scale business activities across different industry verticals and diversify their revenue base.

Network Slicing & Operator Business Models

A useful way to describe a network slice, an adapted version of the Next Generation Mobile Network Initiative (NGMN) definition, is as follows: "A set of network functions instantiated to form a complete logical network that meet the performance requirements of a service type(s)." A network slice is typically made up of sub-network instances to create an end-to-end service. In a mobile network, this can include radio access network (RAN), core and service platforms.

Slices can be fine-grained, at the individual user or service level (for example, a video streaming slice), or can be more coarse-grained, at a company or industry level (for example, an automotive slice, a utilities slice, etc.). The concept is outlined in **Figure 1**. In practice, we expect to see coarser-grained slices come to market first, with increasing granularity over time. Network slice templates can be used to replicate service slice types across different customers. Devices may connect to one, or more than one, slice at time. Ultimately, 5G network slices can be thought of as the network adapting itself, in software, to the needs of the application.

Figure 1: Network Slicing Concept



Source: Heavy Reading

Network slicing should help operators offer new services and broaden their revenue base. By making network connectivity inherent to different use cases and enterprise processes, operators can diversify their customer base and increase revenues massively and sustainably. This is the fundamental reason to pursue slicing and underlines why it is important to develop these market segments on today's widely deployed 4G infrastructure.

In 4G networks, state-of-the-art virtual EPC, in combination with dynamic resource selection, offer operators the opportunity to introduce commercial "slicing" propositions in the near term. These virtual core network slices have great potential in their own right and provide a commercial and technical bridge to 5G.

Real-World Network Slice Service Examples

Depending on the strictness of the definition, network slicing in the form of services isolated from one another on common infrastructure, has existed for many years – VLANs, VPNs, SDN micro-segmentation, etc. In mobile networks, techniques such as APN routing, MOCN, DECOR and SGi service chaining are also used to varying degrees.

Some example "new-wave" services being delivered over LTE networks today are shown in **Figure 2**. Generally speaking, these services use the public RAN and a dedicated, virtualized core network. The most common form of traffic separation is to use APN routing mapped to a class of service in the underlying IP network.

Figure 2: Examples of "Network Slicing" in Commercial 4G Networks

	Connected Car, AT&T: A platform that allows automakers to develop their own, customized connected car offers. The platform offers service modules such as diagnostics, telematics, software updates, remote services, infotainment, etc. It is a global platform that uses dedicated virtual core networks and partner access networks outside AT&T's own coverage area.
vodafone	**Narrowband IoT, Vodafone:** The operator is making an ambitious push into IoT over the cellular network using NB-IoT, with plans to launch in multiple European markets in 2017. It uses a dedicated, virtual packet core for IoT services; this is one of the earliest, most progressive examples of network slicing in 4G.
verizon	**5G Fixed Wireless, Verizon:** One of the earliest applications of 5G is expected to be for fixed residential access. Using mmWave operators are targeting 1 Gbit/s data rates. This also requires its own dedicated core network that can be scaled to support high data rates and rapid subscriber growth.
Enel	**Electricity Smartgrid, ENEL/3 Italy:** A major Italian power plant is covered with LTE and a dedicated core network to offer integrated communications between personnel, sensors, machines and applications. The network offers live video and voice services, very low latency and service continuity in critical emergency situations.
verizon	**Private LTE, Verizon:** This service securely extends corporate networks to employees' mobile devices or branch offices. It offers wide-area connectivity over the public, shared RAN. A dedicated per-enterprise network gateway is used to allow the enterprise itself to control user polices, data caps, traffic management, QoS, etc.

Source: Heavy Reading

Automation & Operations

Virtualization of the 4G core allows fast and cost-effective service turn-up for customers with specific security, management or performance requirements and, as such, is a critical enabler of mobile network slicing. Prior to virtualization, deployment of dedicated physical packet gateways, and associated SGi appliances, was limited to a small number of "high-touch" customers, due to the up-front expense and operational overhead.

Network slicing, by definition, increases the number of functions active in a network. If operated in the same way as traditional networks, the management overhead would be prohibitive. Operators therefore need tools to automate service slice management. Network slice lifecycle management (creation, monitoring, scaling) involves multiple elements that need to be dynamically configurable using automated processes. With the development of resource and service orchestration tools for NFV, the industry as a whole is advancing at a decent pace; however, there is scope for much greater, and more rapid, progress.

Network analytics – the ability to monitor infrastructure and services – is important to closed-loop automation because it allows the network to adapt to prevailing conditions: for example, to predictively re-assign services or resources in the network/cloud infrastructure. Note also that analytics is not only useful for network operations; the underlying intent is to use network intelligence to drive business decisions.

More efficient, and more automated, operations are among the major near-term opportunities in virtual 4G core networks. The mechanisms used to direct traffic into different core network slice instances are overly manual and absorb too much planning and testing time. The cost of change effectively restricts the operator's addressable market and leads to commercial stagnation. To be more commercially and operationally agile, and to scale network slices to a large number of use cases and customers, it is critical that operators address these rigid engineering processes.

NETWORK SLICE SELECTION MECHANISMS

Traffic separation, and the ability to provide specific resources to users or groups of users, is fundamental to network slicing. Ideally, this should apply end-to-end across the network. The 4G RAN is inherently shared, but has a QoS model that allows operators to prioritize services over the air. (Note: A new flow-based QoS model is being developed for 5G.) The task is to map radio bearers, mapped to APNs, into the appropriate transport path and core network instance. There are a number of mechanisms for this.

4G Core Network Selection

Virtualization makes it easier to create dedicated core network instances. There are several mechanisms, standardized in 3GPP, to direct traffic from the RAN into the correct core network. The three main ones in use today are:

- **APNs & Bearer Routing:** APNs are used to logically segment traffic across the network from RAN to core. They can be routed, according to policy, to the correct gateway for the service. A VoLTE APN, for example, would use a priority radio bearer and route traffic to a dedicated P-GW. APNs are widely used, useful, and play a role in the evolution to 5G slicing. There is, however, room to improve the usability and scalability of APNs in association with dedicated core networks.
- **Multi-Operator Core Network (MOCN):** This technique was developed for mobile network sharing and MVNOs, where operators share the same RAN, but route traffic to their own core networks. This is well standardized, but is rather "heavy" and coarse-grained due to the overhead associated with allocation of dedicated PLMNs. This creates challenges to automating and scaling (down) this solution to scenarios such as private enterprise networks or IoT services.

- **Dedicated Core Networks (DECOR):** A relatively new option that enables a device to be directed to a specific EPC on association with the network. It is well suited to virtualized packet core deployments because no additional PLMN IDs need to be allocated. DECOR was standardized in 3GPP Release 13 and enhanced in Release 14 as eDECOR (TS 23.711). The Release 13 version is deployable today, whereas eDECOR requires updates to devices, which will take time to penetrate the market.

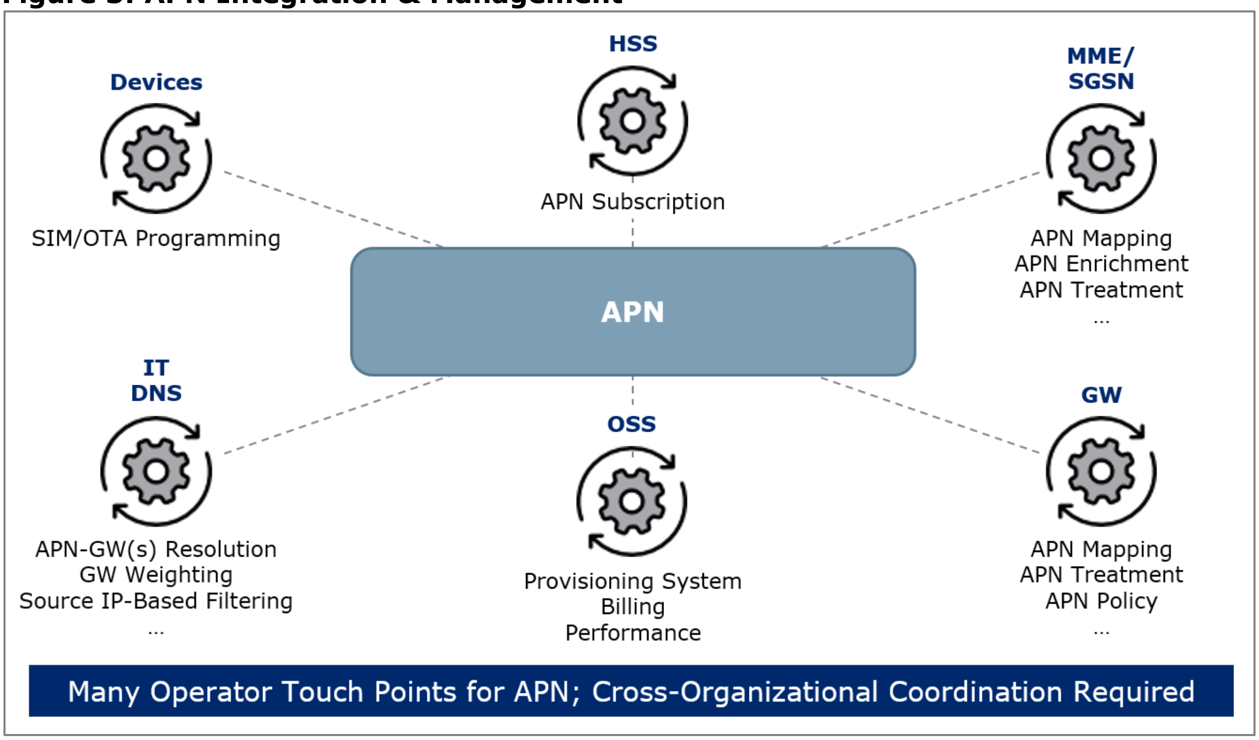
One of the challenges is that these mechanisms require coordination across different network elements – MME, HSS, DNS, PCRF, devices, etc. – which increases the integration and management overhead, making them less suitable for fine-grained and dynamic network slicing. With current generation technology, the more elements a new slice "touches," the harder it is to make the business case. In a dynamic slicing environment, this overhead is multiplied.

Optimizing "High Touch" APNs

APNs are a good example of how multiple touch points with the rest of the network can increase complexity. An APN is a network identifier that determines the Packet Data Network (PDN) to which the user device connects – it can be used, for example, to connect a device to a P-GW associated with a specific service, such as VoLTE, or a dedicated enterprise P-GW. In general, operators today are reluctant to add lots of new APNs; however, as the rollout of virtualized P-GWs continues, and with more diversity in the services and user groups supported on these virtual gateways, the number of APNs will grow.

APNs apply a logical separation from device to the end service or external network (somewhat like a network slice). Thus, each APN impacts many different network elements, which need to be configured accordingly. **Figure 3** shows the range of elements that an APN "touches."

Figure 3: APN Integration & Management



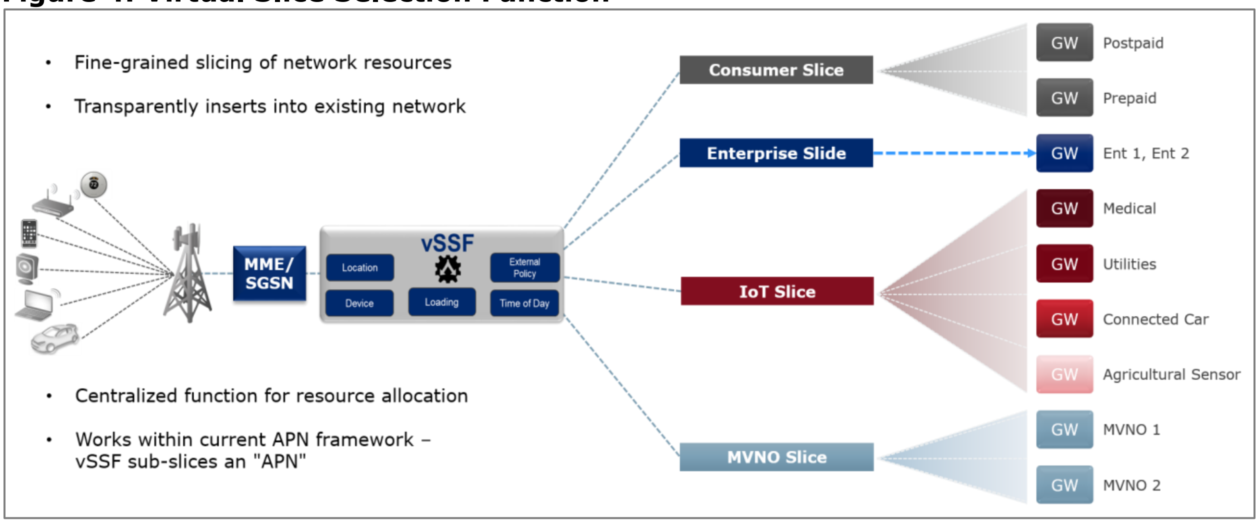
Source: Affirmed Networks

Any change, particularly a new service, therefore involves multiple functional teams. Consider when an operator wants to perform a maintenance-related activity for an APN – for example, to expand capacity (deploy a new gateway), to load balance traffic, or to re-home specific enterprise users to a new gateway. This is a complex process involving multiple updates to the HSS, the MME, the DNS configuration, etc. Today's manual processes are not scalable and, as more APNs are added, the complexity of network operations increases.

Virtual Slice Selection Function

A new option available to operators to simplify operations and streamline gateway (re-)selection across multiple dedicated virtual core network instances (and legacy gateways where needed) is to utilize a control-plane application internal to the operator core network. This new function is called a virtual Slice Selection Function (vSSF), a lightweight/stateless control-plane VNF that can perform dynamic gateway selection. The aim is to unwind the existing APN-centric operator complexity typical of the current 4G architecture, and enable operators to more fully realize the benefits of their investment in virtualization, as shown in **Figure 4**.

Figure 4: Virtual Slice Selection Function



Source: Affirmed Networks

The advantage of this model is that it limits the impact on other parts of the core network. The vSSF is a centralized function where operators can steer traffic to specific gateways without the need to make configuration changes to adjacent network elements. It is transparent to network operation in the sense that other parts of the EPC are unaware of the vSSF. Similarly, DNS configurations – so often a source of operational overhead and/or service outage – can remain unchanged or even be simplified.

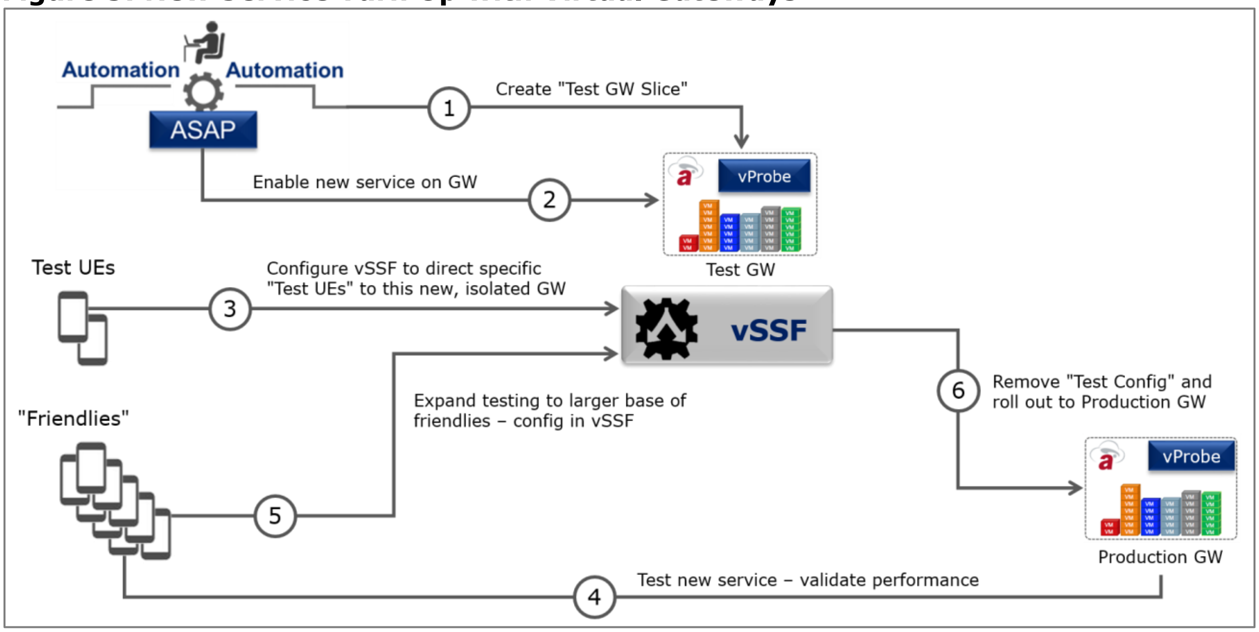
Classification can be multidimensional. The vSSF can use information within the control plane itself, such as IMSI, IMEI, location, APN, RAT type, etc., to direct user equipment (UEs) to the appropriate virtual core, and can supplement this with other policy information (plan type, time-of-day, quota, etc.) to optimize traffic routing. This offers the potential for fine-grained, dynamic service control at the per-user level.

It is possible in principle to achieve these capabilities by extending the existing gateway selection techniques supported by the existing MMEs. This reduces the need to add an additional vSSF function to the network, but has widespread implications for a large number of sensitive nodes already deployed, and will be dependent on vendor support, and typically multiple vendor support, given the number of impacted network functions.

Example: New Service Turn Up

An example of how a vSSF type function can minimize overhead and increase agility is new service turn-up is illustrated in **Figure 5**. The figure shows how an operator can deploy, test and roll out a new service into a production network using the vSSF to steer user traffic into the correct dedicated virtual gateway. This is "low touch" compared to traditional methods.

Figure 5: New Service Turn Up with Virtual Gateways



Source: Affirmed Networks

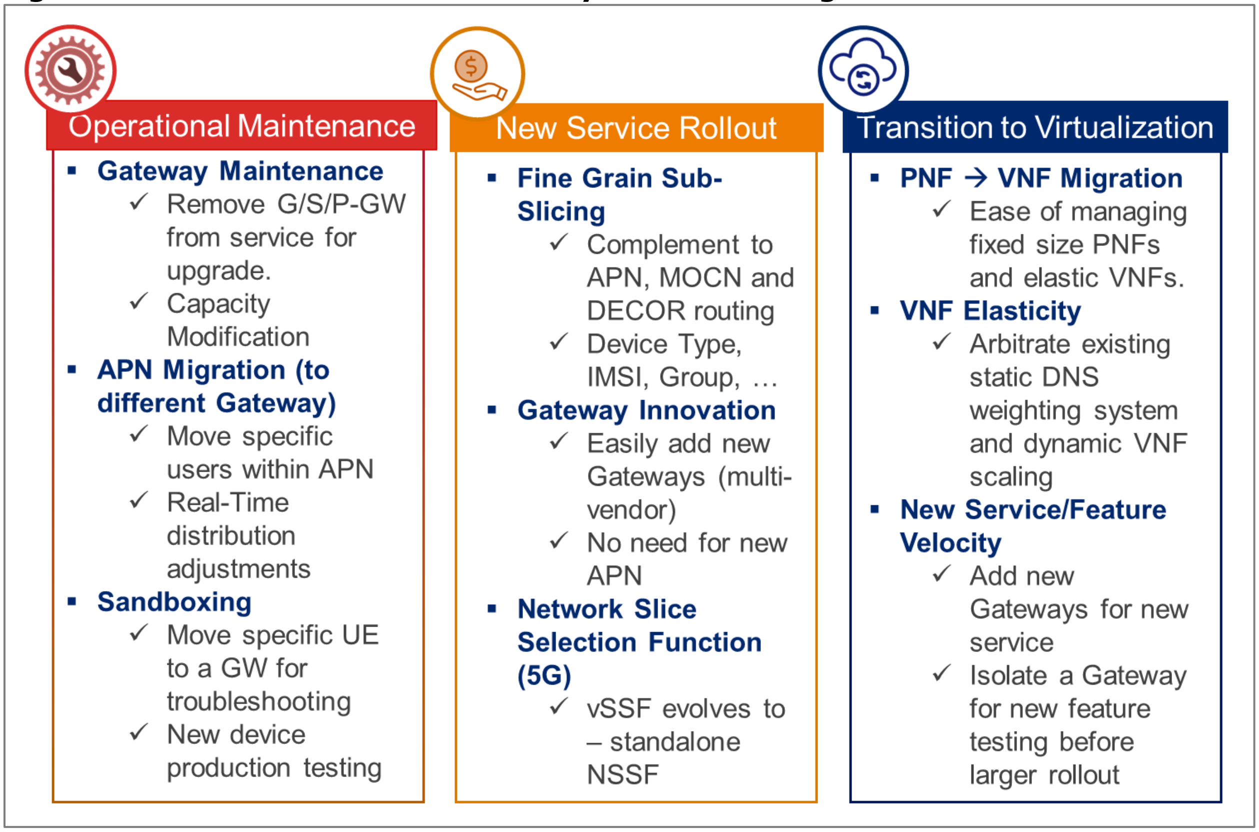
From the point at which the operator's business team identifies the new feature or service, the process outlined **Figure 5** is as follows:

1. Spin up new test virtual Gateway (in a production environment)
2. Enable the new feature (e.g. via new software) on the Gateway
3. Configure the vSSF to direct specific test UEs to this new isolated Gateway
4. Test the new service and adapt as necessary
5. Expand testing to a larger base of customers (friendlies)
6. Make decision to roll out new feature across network
7. Remove SSF steering configuration associated with this new service testing
8. Roll out feature to all production Gateways

Use Case Summary

Figure 6 summarizes the wide range of use cases for a vSSF in the 4G core network, shown in three categories. The SSF is transparent to the rest of the core, and complements the standard MME/SGSN gateway selection methods with additional capabilities, such as visibility of gateway loading and dynamic gateway allocation based on operator-defined policy.

Figure 6: Use Cases for Virtual Gateway Selection Using a vSSF



Source: Affirmed Networks

MIGRATION TO 5G NETWORK SLICING

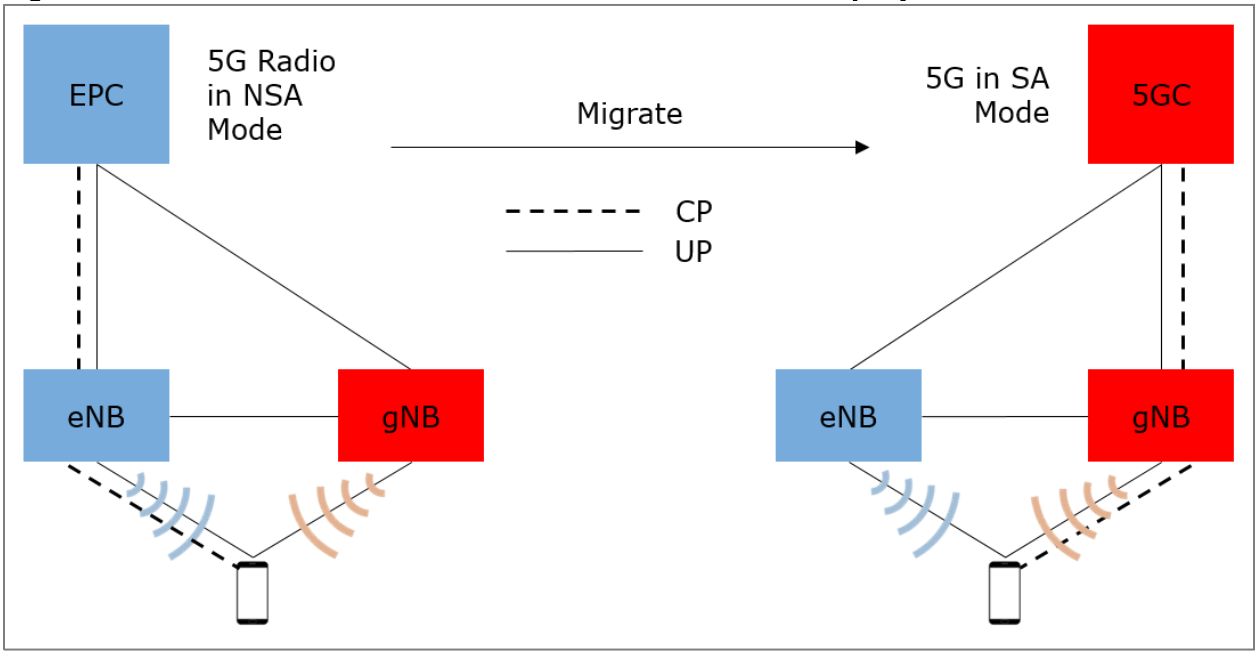
The 5G core network (5GC) is being standardized in 3GPP, with the first specifications due as part of Release 15 in June 2018. It will be further developed in Releases 16, 17 and beyond. The 5G system architecture will contain the functionality needed for dynamic, end-to-end network slicing, although full capability may not be available commercially until 2022 or later – for example, for massive machine-type communications (mMTC).

5G Radio on a 4G Core

Before the 5G core is standardized, 3GPP will develop support for 5G radio to be deployed in non-standalone (NSA) mode, attached to a 4G network, with dual-connectivity to the device. NSA mode, shown to the left of **Figure 7**, will allow fast deployment of 5G and requires relatively little work on the EPC from a standards perspective. NSA 5G radio will, however, have a practical impact on the EPC and, if deployed widely, is likely to accelerate deployment of virtualized, cloud native core networks, the use of control and user plane separation (CUPS), and distributed user-plane processing, to meet the cost, flexibility and scalability demands of 5G radio.

Some of the first mobile 5G services will be delivered using NSA mode on a 4G network – large operators such as AT&T and Vodafone were among the chief proponents of this 5G acceleration initiative – and it is likely that this will be the predominant mode of 5G operation worldwide in the early years. In other words, many operators will need an upgraded 4G core to introduce 5G services, and many services (for example, connected cars) will operate across both networks. This underlines the need to support network slicing in 4G.

Figure 7: Non-Standalone & Standalone Modes for 5G Deployment



Source: Heavy Reading

Interworking 4G Dedicated Core & 5G Slicing

In the technical study on the 5G system architecture (TR 23.799) carried out to inform 5G standards development, it was agreed that it should be "possible to handover a UE from a slice in the 5G Core to a dedicated core network in EPC." This is important for a UE moving between 5G and 4G – for example, due to mobility and radio coverage. Connected car services are an obvious use case.

5G, however, needs to address use cases that are not possible or economically viable on 4G networks. This means the 5GC will evolve to support new and different services than EPC supports – otherwise, what is the point of a new core network? It is logical to expect 5GC to become progressively more capable than, and differentiated from, EPC over time. Therefore, it will not necessarily be effective to develop one-to-one mapping between a 5G network slice and a 4G core network slice. To allow for this decoupling, interworking between 5GC and the legacy 4G core should be relatively simple.

Current proposals in the 3GPP favor restricting the initial set of "Slice Service Types" that should interwork between 5G with 4G to correspond to the three major use cases being developed for 5G release. As shown in **Figure 8**, these are: enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (UR-LLC) and massive IoT (MIoT).

Figure 8: Standardized SST Values

Slice/Service Type	SST Value	Characteristics
Enhanced Mobile Broadband (eMBB)	1	Slice suitable for the handling of 5G enhanced mobile broadband, useful, but not limited to the general consumer space mobile broadband applications including streaming of high-quality video, fast, large file transfers, etc. It is expected this SST to aim at supporting high data rates and high traffic densities.
Ultra-Reliable Low-Latency Communica-tions (UR-LLC)	2	Supporting ultra-reliable low latency communications for applica-tions including, industrial automation, (remote) control systems.
Massive IoT (MIoT)	3	Allowing the support of a large number and high density of IoT devices efficiently and cost-effectively.

Source: 3GPP, TS23.501, V1.0.0 (2017-06), Table 5.15.2.2-1

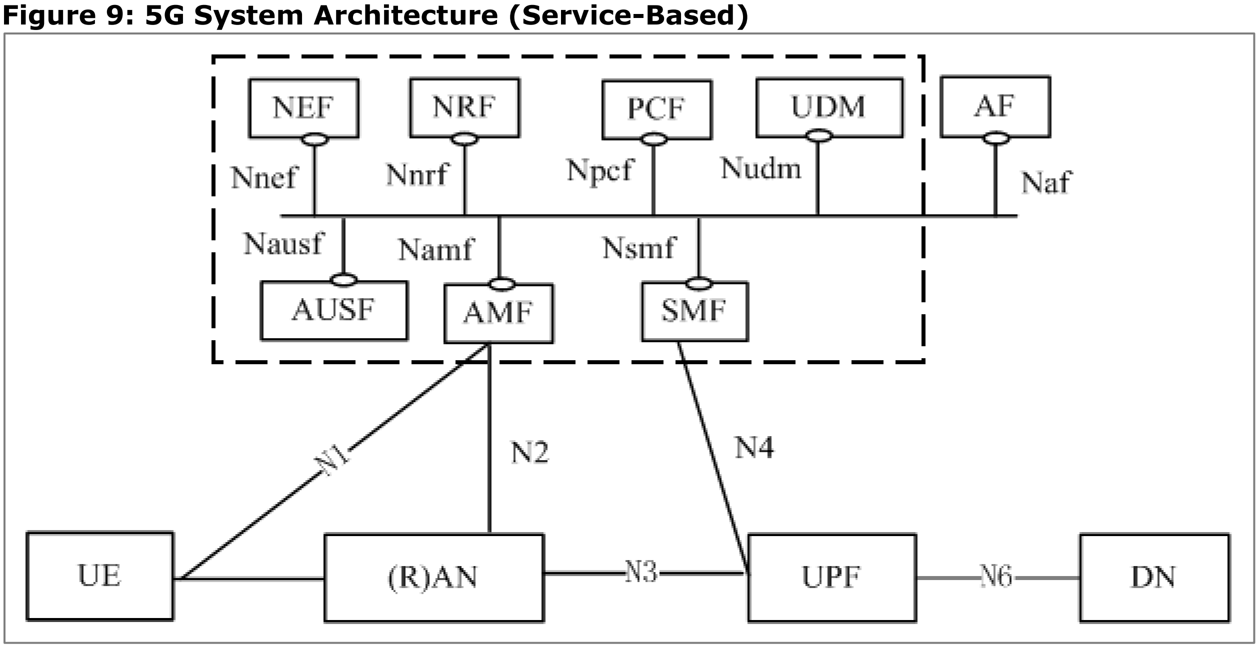
Ultimately LTE access will migrate to the 5G core, shown on the right of **Figure 7**, and this will become the common core for many access types across fixed and mobile. This will enable operators to offer converged, "follow-the-user" services.

5G Core Network & Architecture

The 5G network architecture and 5G core are being developed by the Technical Specification Group 23.501. The work is part of Release 15, due to freeze in June 2018, and will enable 5G to operate in standalone mode, without dependencies on LTE. This is an ambitious time-table, and the work will be phased such that some more advanced, or less-pressing, capa-bilities will come in later releases.

Figure 9 shows the service-based version of the 5G system architecture.

Figure 9: 5G System Architecture (Service-Based)



Source: 3GPP, TS23.501, V1.0.0 (2017-06), Figure 4.2.3-1

The major components of NG Core are listed below:

- **Access and Mobility Management Function (AMF):** Manages access control and mobility. The AMF also includes the Network Slice Selection Function (NSSF).
- **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. These are equivalent of GWs in 4G.
- **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.
- **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 APIs.

Network slicing is fundamental to 5G, and a number of new terms have been introduced during the standards development process. These include Network Slice Instance (NSI), Slice/Service Type (SST), Network Slice Template (NST) and Network Slice Selection Assistance Information (NSSAI).

SUMMARY/CONCLUSION

The network slicing concept is fundamental to operators seeking to scale their business activities across different industrial sectors. By making network connectivity part of the fabric of an enterprise process, operators can diversify their customer base and increase revenues massively and sustainably. In our view, it is important for operators to develop these market segments on today's widely deployed 4G infrastructure, ahead of a more advanced implementation in 5G.

In 4G networks, virtual EPC, in combination with dynamic resource selection, offers operators the opportunity to introduce commercial "slicing" propositions in the near term. These virtual core network slices have great potential in their own right and provide a commercial and technical bridge to 5G. Virtualization makes it easier to create dedicated core network instances and allows for fast and cost-effective service turn-up for customers.

One of the challenges of using dedicated core networks in combination with APN routing is that these mechanisms require coordination across different network elements – MME, HSS, DNS, PCRF, devices, etc. – which increases the integration and management overhead and makes them less suitable for fine-grained and dynamic network slicing. To fully take advantage of this capability, operators need to automate engineering processes.

One potential new way to do this is to use a "Slice Selection Function" that is transparent, operationally, to other core elements. This complements resource selection methods performed by the MME/SGSN in 4G today and is similar to the NSSF/AMF in the 5G core.

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.