analysys mason

White paper

5G network slicing: cross-domain orchestration and management will drive commercialization

September 2020

Anil Rao

RESEARCH

Contents

List of figures

1. Executive summary

Enterprises rely on communications service providers (CSPs) to supply critical network connectivity in order to conduct their business. Service providers have historically delivered this connectivity with limited flexibility and control, meaning that it is not fit for purpose for new-generation digital enterprises. Enterprises in a range of industries are rapidly accelerating their digital transformation initiatives; these require highly flexible network connectivity services that they can provision on-demand and according to their unique performance requirements. 5G network slicing enables this.

The 5G network (with a programmable transport network, the cloud-native core, new radio and edge clouds) is expected to become the 'services creation platform' for next-generation communications. The platform will be used to create separate network partitions (or 'slices') with unique network performance and latency characteristics to serve a particular use case or enterprise. A combination of networking technology innovations and enablers such as segment routing and software-defined networking (SDN) in the transport network, and network function virtualisation (NFV) and cloud-native computing (CNC) in the 5G core makes domain-level slicing and end-to-end network slicing possible.

Network slicing allows service providers to offer differentiated slice-based services on a per-use-case basis and provides enterprises with slice-level control and the necessary mechanisms to self-provision connectivity services. To deliver network slicing, service providers must do the following.

- **Implement end-to-end network slicing orchestration.** The network slicing solution must be cross-domain by design in order to stitch together an end-to-end slice spanning the RAN, edge, transport and the core. It must also support a model-driven multi-layer architecture with open APIs in order to provide necessary abstraction both at the domain and cross-domain level. Multi-vendor support will be essential to give service providers the choice to deploy their preferred vendors' solutions at the network resource layer, domain slice orchestration layer and cross-domain slice orchestration layer.

- **Automate network slicing orchestration.** It will become critical for service providers to control the incremental cost of creating new slices as more enterprises consume network connectivity in the form of network slices to support new use cases. This can be achieved by maintaining a catalogue of slice templates for the most popular use cases so that they can be very quickly instantiated and provisioned, both at the domain and cross-domain level.

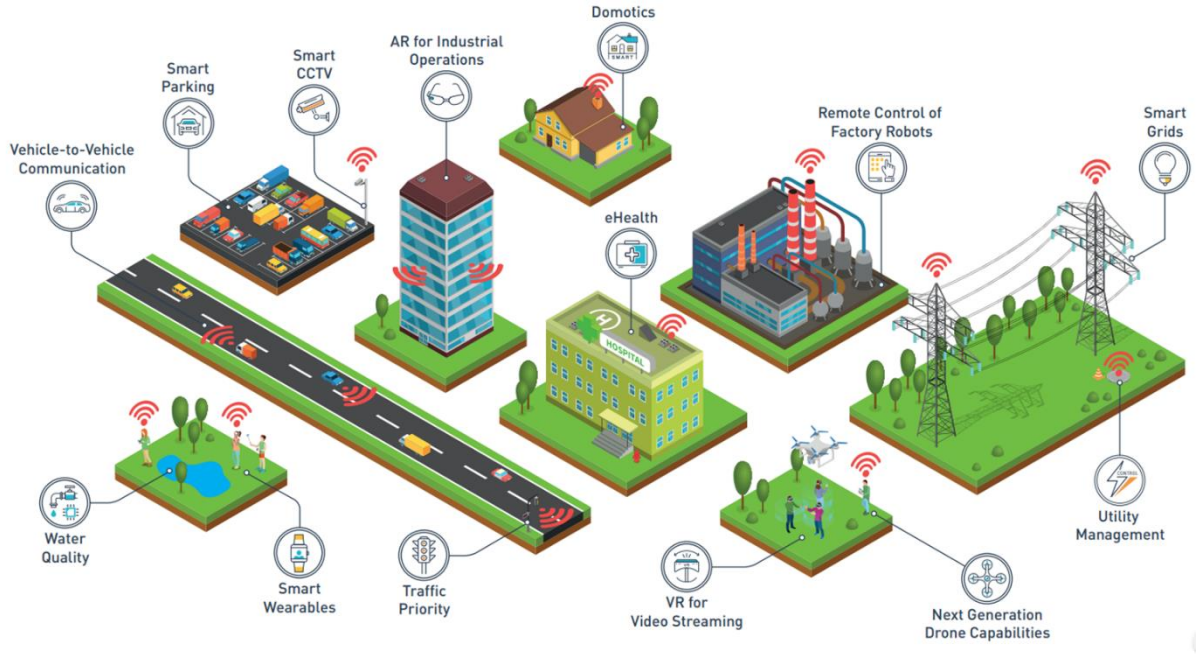
Domain-level network slicing technology will vary greatly depending on the network domain. Slicing in the transport domain can be enabled by a combination of IP routing innovations such as segment routing and SDN controller technology and applying the slice forwarding configurations as needed. Higher scale and automation can be achieved by applying predefined slice forwarding configurations when an enterprise creates a slice.

Conversely, core network slicing will depend on the ability to configure the virtual network functions (VNFs) or containerised network functions (CNFs) with the slice requirements. Core network slices will be created as either shared or dedicated network functions according to the slice policy, which is in turn driven by the service offered to the end customer. The core domain orchestrator is likely to be supplemented with slice orchestration capabilities and must interwork closely with the Kubernetes container orchestrator for CNF-level resource configuration for network slices.

2. Enterprises in a range of industry verticals are accelerating their digital transformations

Digital transformation means different things in different industries. At a high level, terms such as Industry 4.0 and industrial Internet of Things (IIoT) have been used to explain digital transformation as a broad concept for 'connected' enterprises in industries such as manufacturing, energy and utilities, transport and logistics. These terms have also been used to explain how enterprises can deliver superior experiences to their end customers and improve business agility, operational efficiency, innovation and competitiveness. Technologies such as robotics, big data, ML/AI, blockchain, IoT, cloud, edge and next-generation connectivity are the key enablers for digital transformations. Figure 1 illustrates some examples of digital transformation in various industries.

Figure 1: Digital transformation in a range of industries

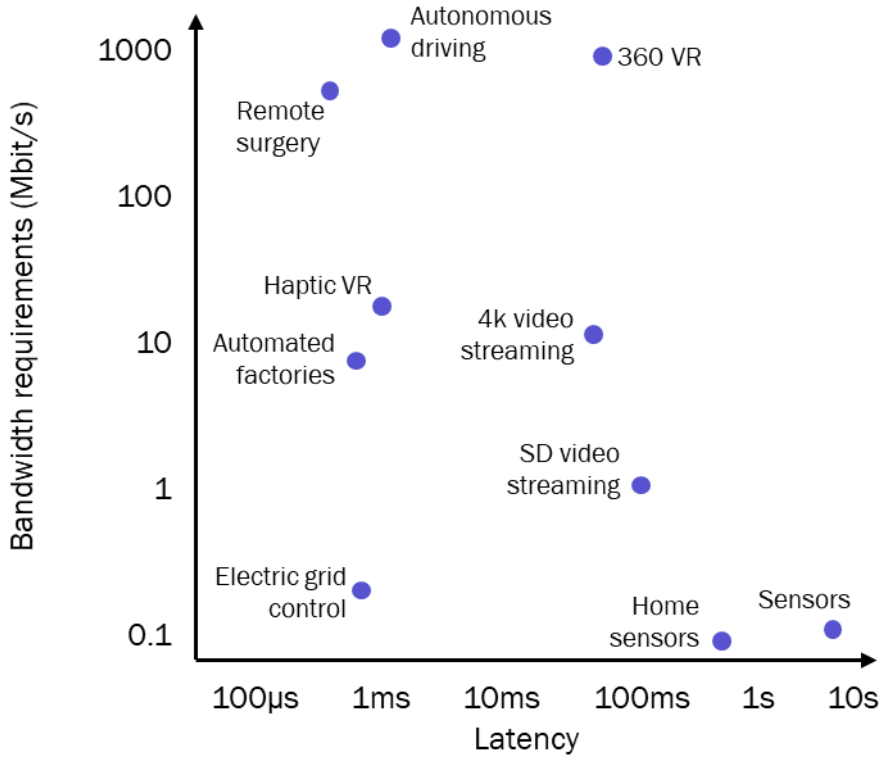


Source: Analysys Mason, 2020

However, it is now becoming clear that broad-ranging, 'one size fits all' approach is not sufficient and that each industry and individual enterprise must consider what digital transformation looks like to them. Enterprises are carefully assessing each of the technology enablers and partner ecosystems that can help them to achieve the transformation objectives of their businesses. Next-generation connectivity is right at the heart of these considerations.

Service providers traditionally offered connectivity via inflexible, static connections with little or no scope for customisation. Such services also required overprovision for peak demand, which meant that enterprises always paid for unused capacity. This 'one size fits all' approach to offering network connectivity services is the antithesis of meeting the unique demands of a digitally transformed enterprise. Consequently, enterprises are looking for service provider partners that can supply fit-for-purpose next-generation connectivity solutions that give them the flexibility and control to request, configure and modify the network resources on demand and according to their unique performance requirements. Figure 2 depicts the wide range of network performance requirements for a selection of use cases.

Figure 2: Network latency and bandwidth requirements for a range of use cases



Source: Analysys Mason, 2020

3. 5G will support a range of enterprise use cases and network performance requirements

The service provider community and industry stakeholders are unanimous in the view that 5G is not just another mobile technology generation. The term 5G is typically used to describe a set of mobile network standards that have been sanctioned by the 3GPP, but 5G offers more than just mobility. It has the potential to be a game changer for service providers, industries and societies and will support:

- a diverse set of use cases
- the running of applications that require very low latency, very high availability or extremely high device density on the same network
- far higher data rates and coverage than previous generations, with lower costs of delivery, thereby putting high-speed broadband in the reach of every business and citizen
- a wide variety of spectrum bands to improve the price/performance of mobile broadband for all stakeholders
- advanced multi-tenancy and network slicing, which will enable many more industries and their service providers to address specialised performance requirements at an affordable price.

This whitepaper will:

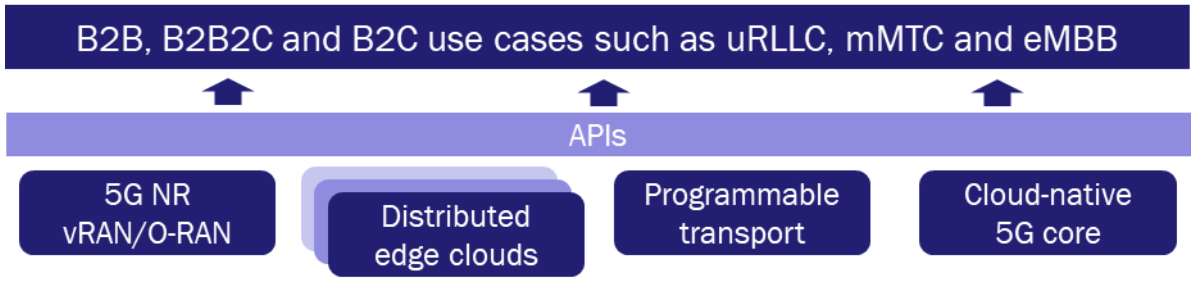
- outline how service providers and enterprises can implement end-to-end network slicing and operationalize network slices at scale for enterprise services on a 5G macro network
- provide an in-depth discussion on how slicing can be achieved in the transport network and cloud-native 5G core.

3.1 True end-to-end network slicing will be possible with a programmable, 5G network platform

The concept of network slicing is not new. Service providers have previously achieved slicing by using access point name (APN) and virtual private networking (VPN) technology in 4G LTE and fixed networks, but the inherent limitations of the physical network meant that the configurations had to be hardcoded, which made the process highly manual and cumbersome. Furthermore, these slices did not provide enterprises with flexibility and differentiation, so the service providers could not charge a premium.

5G, on the other hand, is being developed as a 'services creation platform' (Figure 3). It combines new radio (physical and open RAN), edge clouds, programmable transport and the cloud-native core to create and deliver differentiated, on-demand IoT and ultra-low latency use cases to enterprises; end-to-end network slicing makes this possible.

Figure 3: The 5G network services creation platform



Source: Analysys Mason, 2020

NFV, SDN and cloud-native computing are the foundational capabilities upon which network slicing can be realised. These base capabilities provide the fundamental building blocks to transition from the traditional rigid physical network into a flexible, dynamic and a programmable 5G network platform. NFV uses the ability to create and modify network resources to reflect the changing service requirements to enable service providers and enterprises to allocate network resources precisely when they are needed. They can therefore achieve higher network efficiency through optimum resource utilization.

SDN introduces dynamicity to the transport network by enabling the programmatic control of the traffic management processes. This means that service delivery can be optimized and network costs can be reduced. Furthermore, designing network functions using cloud-native technologies such as containers and microservices prepares the network to be deployed in the cloud, thereby allowing service providers to use DevOps and CI/CD methodologies to independently manage the lifecycle of the microservices without affecting the overall service. This also enables the automatic scale-up of only those network function service modules that require more capacity, and allows for rapid service innovation through automated network and service orchestration by interlinking select network function modules into end-to-end services, thereby significantly reducing service

creation timescales to minutes. 5G services that require network functions to be instantiated on demand will depend on cloud-native virtualization and SDN technologies.

The control plane and user plane separation (CUPS) concept proposed in 3GPP release 14 enables the decentralization of the data forwarding component, the formation of the user plane function (UPF) and the placement of this function closer to the network edge. This allows packet processing and traffic aggregation to be performed in the distributed edge clouds or on-premises in enterprises.

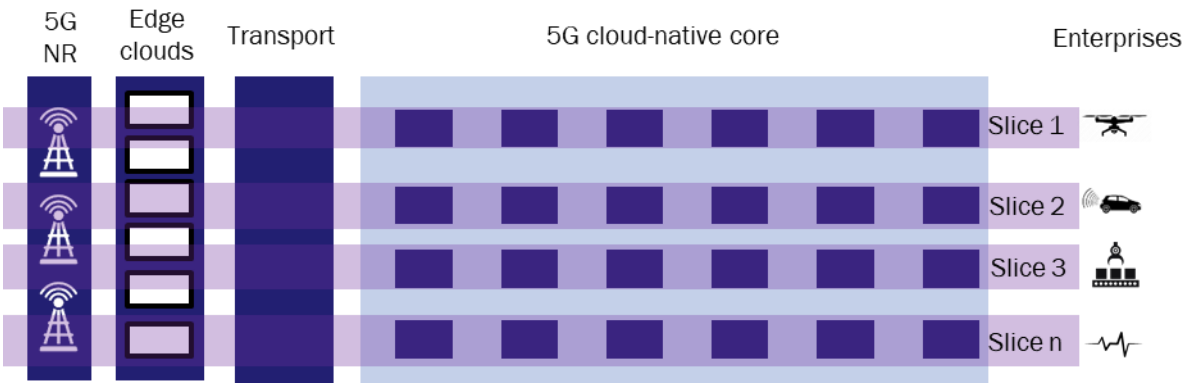
Service providers can use these new capabilities to create multiple, separate 'end-to-end slices' of the network, which in turn allows for use-case-based slice instantiation, SLA-driven instantiation (dedicated or shared) and network function placement (at the edge or core) based on the specific latency, performance, reliability and availability requirements of each use case/service. The network slices run in isolation from each other; each slice is used to solve a specific business problem. Network slicing paves the way for service providers to offer differentiated slice-based services for whole industries or granular, differentiated slices for each use case, subscriber type, application or enterprise with slice-level control, management and quality of service. Network slicing also enables enterprises to create multiple services within their own network domains without the help of service providers.

Some use cases (such as factory automation) are likely to be fulfilled through the deployment of private 5G networks, where the enterprise deploys the 5G radio on-premises with either an on-premises private 5G core or a public-cloud-hosted/centralised instance provided by the service provider. A combination of factors, such as economic considerations (capex/opex) and technology maturity, will dictate whether the enterprise will opt for private networks or a network slicing strategy, or both.[1] Key stakeholders in the telecoms ecosystem and specific industry verticals (including service providers, enterprises, standards bodies and vendors) are evaluating all options.

3.2 The automation of cross-domain slice management and operations is critical for economies of scale

To take advantage of the monetization opportunities made possible by network slicing, the 5G network must be treated as one whole entity and must be supplemented by suitable slice management and operations technology in order to partition, assign and lifecycle manage the network slices for consumption by enterprises (Figure 4).

Figure 4: End-to-end network slicing



Source: Analysys Mason, 2020

[1] Network slicing can also be implemented in private networks.

A network slice is therefore inherently end-to-end and cross-domain in nature. This presents the following two 'how to' challenges for the service providers.

- **How to implement end-to-end, cross-domain network slice management and operations.** A slice management and operations solution (or network slice orchestration solution) must inherently be cross-domain in nature. As an example, a network slice using a CUPS-based topology will require the instantiation of specific control plane network functions (such as SMF and PCF) in the 5G core domain and remote user plane functions (UPF) in the edge cloud data centre domain. The appropriate transport network resources must also be configured in the backhaul domain.

- **How to deliver and manage network slices at scale.** Both enterprises (slice consumers) and service providers (slice providers) will need a highly efficient and cost-effective way of ordering, instantiating, provisioning, monitoring and managing the lifecycle of network slices as their adoption increases. Service providers must consider the economics of deploying a slice and to what extent they can control the incremental cost of deploying tens or hundreds more slices. Enterprises must contemplate how flexibly they can control their own slices and introduce new service capabilities more quickly. Extreme automation of network slice management and operations is therefore critical to achieve the requisite economies of scale.

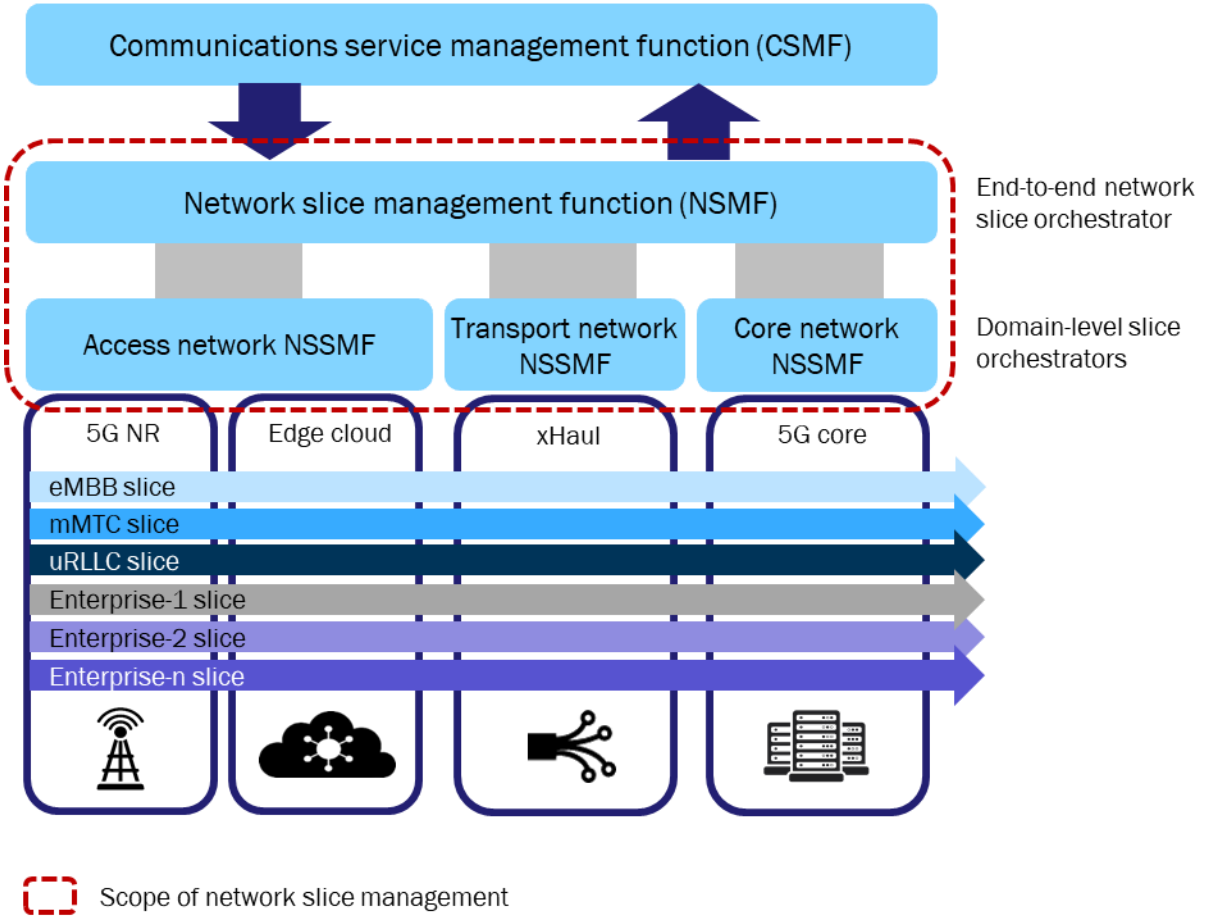
4. Cross-domain network slice orchestration enables end-to-end slicing

The high-level network slice management framework published by 3GPP outlines four key management functions for network slicing: the communications service management function (CSMF), the network slice management function (NSMF), the network slice subnet management function (NSSMF) and the network function management function (NFMF). The CSMF is a higher-layer OSS/BSS capability aimed at performing customer order management, and the NFMF is for the application-level management of VNFs, PNFs and CNFs.

The NSMF will perform cross-domain network slice orchestration using the domain-level slice management functions, with each network domain having its own NSSMF. This architecture allows for the instantiation and configuration of network slice resources for each of the use case types (eMBB, mMTC or uRLLC) in each subnet or domain, dictated by the end-to-end network slice intent and governed by the end-to-end slice orchestrator.

It is expected that additional new innovative use cases will be identified to fit the needs of enterprises that require specific levels of service. In some cases, the combination of slice parameters will be new and unique, while in other cases, multiple services will share common characteristics, albeit with differences in the target infrastructure. It is therefore expected that the NSMF will maintain a library of slice templates for popular use cases to enable the sharing and/or fast provisioning of slices due to their similarities.

Figure 5: 3GPP reference architecture for network slicing



Source: 3GPP and Analysys Mason, 2020

The NSSMFs are likely to be built as extensions of the existing domain orchestrators (for example, the NFV orchestrator for the 5G core) and the domain controllers (such as the SDN controller for the transport network). The cross-domain nature of the NSMF has been discussed at a high level in various standards bodies (most notably by ETSI, where it is referred to as the Service Orchestrator). However, the concept of slice-based service orchestration and lifecycle management is still new, and the standards are still emerging.

4.1 The cross-domain network slice orchestration solution must exhibit certain foundational traits

In addition to having an end-to-end cross-domain capability, the network slice orchestration solution must be developed based on the following design principles.

Multi-layer architecture. One of the foundational capabilities of the solution is to provide multi-layer control at the domain and cross-domain level. A multi-layer orchestration solution will ensure abstraction and operational demarcation in what is going to be a highly distributed mobile network architecture with multiple networking, data centre and cloud domains including the RAN, WAN, edge and core. The lower-layer domain orchestrators will perform the individual domain-level orchestration and software-defined control of each of these domains, while the cross-domain orchestrator will be the 'orchestrator of orchestrators', and will provide the higher-layer abstraction and end-to-end cross-domain service orchestration.

Model (or intent)-driven. There will be a mix of network infrastructure in the mobile network environment, ranging from the traditional physical network functions to VM-based network functions and the new container network functions for 5G standalone (SA). Each of the components will demonstrate a different level of maturity. The orchestration solution should therefore support a model-based approach that provides a high level of customisation to cater for various network function permutations and requirements in order to enable service providers to develop new services and use cases without depending significantly on the individual components of the underlying network.

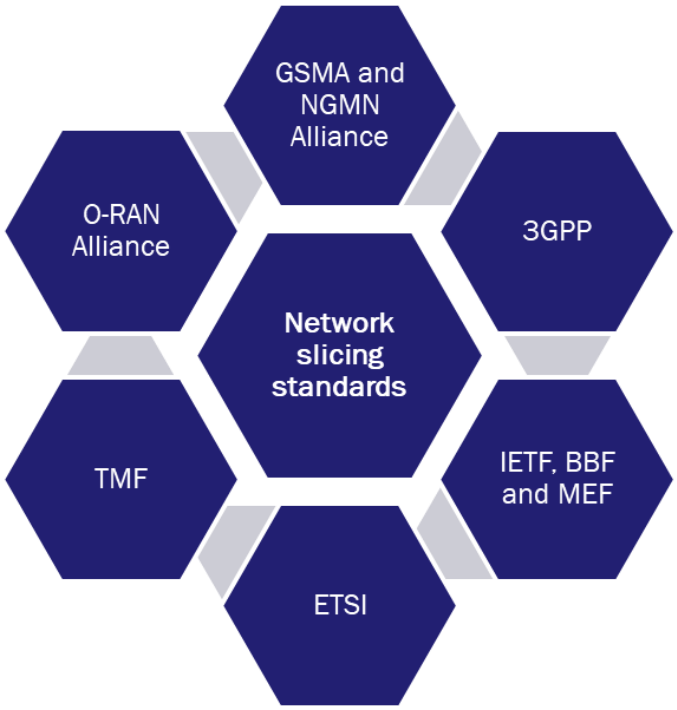
Platform-oriented and open API-driven. Adopting a platform approach will enable the solution to be built in a modular form and be microservices-based. This will make it easy to introduce new capabilities, make the solution composable and allow service enhancements with minimal disruption to the run-time environment. The solution must comply with open, industry-standard APIs for frictionless solution composition and easier integration with third-party software components to enable best of breed orchestration. Additionally, the platform should be developed using a highly curated set of tools for managing the DevOps and CI/CD pipelines to automate the build, test and deployment processes. This will prevent 'tool sprawl' and increase the automation efficiency of solution development and deployment.

Multi-vendor support. The cross-domain network slice orchestration (NSMF) solution must support multi-vendor capabilities, that is, it must easily integrate with any third-party domain-level slice orchestrators (NSSMF). The domain-level orchestrators must be vendor-agnostic and have the ability to orchestrate a network consisting of multi-vendor network equipment, VNFs and CNFs. As the networks become more virtualized, it is likely that some of the network functions will be hosted in a hybrid-cloud environment (private and/or public). This will require multi-cloud slice orchestration capabilities across both the private cloud environment within service providers' data centres and public clouds such as those from AWS, Azure and Google.

4.2 Several standards bodies are working to define the specifications for network slice management

A range of standards development organisations (SDOs) are working on initiatives to create standards and specifications for network slice lifecycle management and operations. Figure 6 gives examples of some of the active network slice management SDOs. In general, the SDOs are focusing their efforts on the most relevant network domains based on their specialism.

Figure 6: Examples of the standards bodies that are developing specifications for network slice management



Source: Analysys Mason, 2020

The GSMA and the NGMN Alliance are working as umbrella liaison SDOs to ensure industry alignment and convergence of technology standards around network slicing. 3GPP is leading the creation of technical specifications for slicing in the RAN and core domains. IETF is looking at IP router protocol enhancements (for example, segment routing and L3VPN) and is working with 3GPP to define the interfaces between RAN and core network slicing management. BBF is working on the slicing specifications for the 5G bearer networks and the related transport network slicing management architecture. MEF is looking at slicing in the shared fronthaul and backhaul networks, among other things. The ETSI ISG NFV is responsible for providing technical specifications for the management and orchestration of NFV-based and cloud-native 5G core. The ETSI ISG ZSM is focusing on cross-domain network slicing management and how the end-to-end orchestrator should interface with the individual domain-level slice orchestrators. The TMF is incorporating network slicing requirements into their ZOOM initiative, while the O-RAN Alliance is developing specifications for the OpenRAN and is working to include the RAN slicing requirements. In addition, MEC is looking at slicing in MEC-based edge clouds and the ONAP project under the Linux Networking Foundation is implementing the end-to-end 5G network slicing use case as part of the Frankfurt release.

A broad range of organisations from a variety of vertical industries are collaborating with many of the SDOs listed above. These include the 5G Automotive Association (5GAA), the Industrial Internet Consortium, Industrie 4.0 and ZVEI.

The very large number of SDOs and other organisations involved in developing the wide-ranging specifications for network slicing lifecycle management demonstrates the complexity of the task facing the industry. Deep collaboration between SDOs, service providers, the vendor ecosystem and industry consortia will be critical to achieve consensus and standardisation.

5. Networking innovations powered by domain orchestration enables transport network slicing

The transport network consists of the fronthaul, mid-haul and backhaul transit legs. As such, transport slicing technology is based on packet technology from the antenna site to the mobile core. Cell site routers are located at antenna sites and go through a series of packet switching layers to provide aggregation towards the core of the transport network and access to the data centre assets that are distributed around the network.

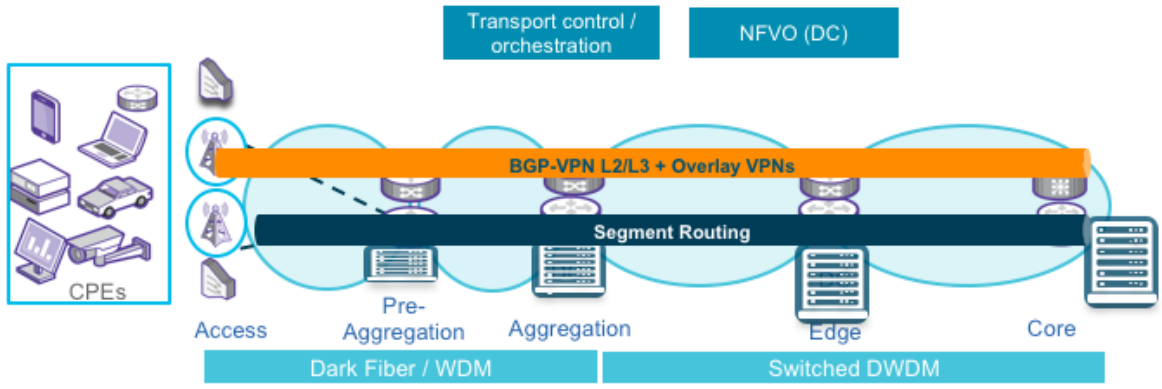
5.1 Transport network slicing architecture

The transport architecture includes common underlay infrastructure consisting of a segment routing (SR) control and data plane, which can be used for 5G services or can be segmented using BGP-based L2/L3 VPNs or overlay VPN technology (Figure 7).

The key characteristics/parameters of a transport slice include the following.

- **QoS.** IP packet marking/classification is required, along with per-hop-behavior (PHB) scheduling and/or shaping to meet SLAs for latency, throughput, loss and jitter.
- **Path forwarding policy.** This provides the links for specific traffic to traverse. Latency and bandwidth requirements may be the obvious inputs to the policy, but other options are possible with SR technologies, such as the use of disjoint paths, only traversing encrypted links or only traversing highly reliable circuits.
- **Connectivity.** This defines which sites can and cannot talk to various other sites. The logical separation of traffic across slices is a critical requirement for increasing security and is enabled using BGP-based VPNs.
- **Service assurance.** Slice performance must be monitored to ensure that it is meeting the SLAs.

Figure 7: Transport slicing architecture



Source: Cisco

The transport network will support a set of transport slice forwarding behaviors that are suitable for the generic 5G use cases (URLLC, eMBB and mMTC). Additional forwarding behaviors can be added to these, for example a secure forwarding plane that only utilizes encrypted links or a transport forwarding plane that is dedicated to a very large enterprise. At this level, the emphasis is primarily at supporting a small set of discrete and different

forwarding behaviors. It is anticipated that the forwarding behaviors will be set up in advance for the generic 5G use cases.

For enterprise-specific network slices, the transport domain-level forwarding planes can be configured in a more dynamic manner using the transport domain controller and attaching it to the end-to-end slice. For example, when an enterprise requests a slice, VPNs associated with that enterprise can be provisioned and configured in the appropriate data centres and in the RAN resources for the new slice. Connectivity between these network resources will be built for the specific slice using a set of transport extranet VPNs that ensure communication within the slice, but not with components in other slices. In this way, enterprises can have their own individual slices using transport network-level VPNs. The traffic is treated in a fashion that is appropriate to the slice type by associating these VPNs with a specific set of transport forwarding behaviors.

The key to achieving this at scale is to build a catalog of predefined 'slice forwarding behavior' templates and instantiating the most suitable template when an enterprise requests a slice. The slice characteristics are dynamically configured as the slice is instantiated. With this approach, potentially hundreds or thousands of 'slice instances' can be automatically instantiated using a limited set of 'slice types'. The slice types can be used to identify the requirements for the 'underlay' QoS and the forwarding behavior. The slice types are defined as network slice subnet templates (NSSTs) according to the 3GPPP terminology, and the slice instances are the network slice subnet instances (NSSIs).

5.2 There are both hard and soft approaches to transport network slicing

The terms hard and soft slicing were coined by the IETF and have taken hold in the wider industry. However, they are still evolving. They support the same set of slicing functionalities, but the way in which the slice is built differs for each approach. For hard slicing, the transport resources are dedicated to a specific NSI and could take the form of dedicated links, dedicated forwarding planes and even dedicated routers. For soft slicing, the resources used to build the slice are shared and can be re-used by other slices.

However, hard and soft slicing are loose definitions; there is, in fact, a spectrum of slice hardness and softness. At one extreme, each slice has its own totally dedicated network (links, routers and management), while at the other extreme, all slices use a common internet-type environment. There are many intermediate points between these two extremes; how the transport slice characteristics are configured determines whether the slice is hard or soft. Figure 8 presents a high-level view of the ways in which the slice parameters can be configured to create different transport slice architecture. These options can be mixed and matched using the same infrastructure to create, for example, a hard slice architecture for URLLC services but also a soft slice architecture for eMBB and mMTC services.

Figure 8: How to configure soft and hard slicing for each slice parameter

	SOFT SLICING ⟷ HARD SLICING		
Edge QoS	H-QoS ingress policing / H-QoS egress scheduling	H-QoS ingress policing / H-QoS egress scheduling	H-QoS ingress policing / H-QoS egress scheduling
Core QoS	Shared queues on core links	Dedicated queues / mix of shared / dedicated queues on core links	H-QoS on core links
Forwarding plane	Single core forwarding plane based on IGP and IGP metrics	Multiple forwarding planes based on IGP and non-IGP metrics using Flex-algo or SR-TE.	Multiple forwarding planes tied to core sub-interfaces using flex-algo or SR-TE
VPNs	BGP based routes with no route coloring	Mix of BGP based routes with and without route coloring	BGP based routes with route coloring
Forwarding plane selection	Default forwarding plane only	Mix of default / destination / flow based forwarding plane selection	Destination and/or flow based forwarding plane selection based on ODN/AS

Source: Cisco

5.3 The transport slicing toolset

Service providers can apply a set of tools at each layer within the transport network to create slices with different characteristics and levels of hardness. Figure 9 illustrates how this toolset enables the network to be designed to concurrently support a mix of hard and soft slices on the same network at the same time.

Figure 9: Transport slicing toolset

Tools	Description
Underlay infrastructure	The underlay network spans the WAN portion of the packet network and uses a combination of the data and control plane, core QoS and performance management tools to create one or more forwarding planes, also called slice forwarding planes. These slice forwarding planes are designed to meet a set of optimization and constraint objectives that are associated with 5G services. Example objectives and constraints include latency and jitter, bandwidth, availability, security and economic constraints.
QoS	End-to-end packet-level QoS is one of the most important components in building viable transport slice architecture, regardless of whether the aim is to support soft or hard slices, or a mix of both. The proposed approach is based on Diffserv architecture; this is packet-based architecture that relies on edge equipment classifying, metering and marking into a small number of core transport classes on ingress.
Edge QoS	Edge devices (PE) need to support hierarchical ingress and egress QoS.
Core QoS	This refers to how slice traffic is treated on core links. It is assumed that the edge PE device has conditioned and marked the traffic on ingress to the network. Ingress marking will set the MPLS traffic class and core scheduling will use this field. Three core QoS models are examined to support 5G services and the level of resource sharing varies for each, hence some slice solutions are harder in nature, but are also more complex and less scalable.
Core QoS: shared core queuing	This is how most service provider networks are built today. It relies on a small number of shared queues in the core of the network. These are shared between all slices running on the network and are configured to achieve a PHB that is appropriate to the service they are carrying. It is a soft QoS approach because different slices share common queues in the core.
Core QoS: dedicated core queuing	In this model, traffic for some slices or slice types is given a dedicated core queue and PHB. For example, the URLLC traffic, regardless of ingress traffic marking, uses a dedicated queue and associated PHB in the core (hard). The eMBB and mMTC traffic could continue to use a shared core queuing solution, such as that outlined above (soft).
Core QoS: dedicated core	In this model, each slice has dedicated logical links in the core network. Each logical link has an underlying class-based queuing system, so there is not only link isolation, but also queue isolation

Tools	Description
bandwidth and class queuing per slice	between slices. This is achieved by running VLANs on core Ethernet interfaces. The core links run H-QoS so that each VLAN has an associated bandwidth implemented through a shaper and an underlying class-based queueing system.
Underlay forwarding plane	The forwarding plane determines how traffic is sent over the underlay core infrastructure. The following three approaches are used to construct the underlay forwarding plane/planes for a 5G transport infrastructure. • Single forwarding plane for all slices. The underlay network relies on the IGP (ISIS or OSPF) to calculate the forwarding tables. The routing protocol sees all links and there is a single forwarding table that is calculated based on IGP metrics. All traffic takes the shortest path between two points within the network and utilizes ECMP. • Forwarding plane per slice service. Within the underlay network, a forwarding plane per 5G slice service is built to meet the forwarding behavior required by the slice (such as having a dedicated slice forwarding plane per 5G service type (URLLC, eMBB and mMTC)). • Hard forwarding plane per slice customer. This is a variation on the previous scheme. In this case, a forwarding plane for individual customers is defined, rather than having a forwarding plane per 5G service type. This will be suitable for large customers only (such as MVNOs and MOCN customers).

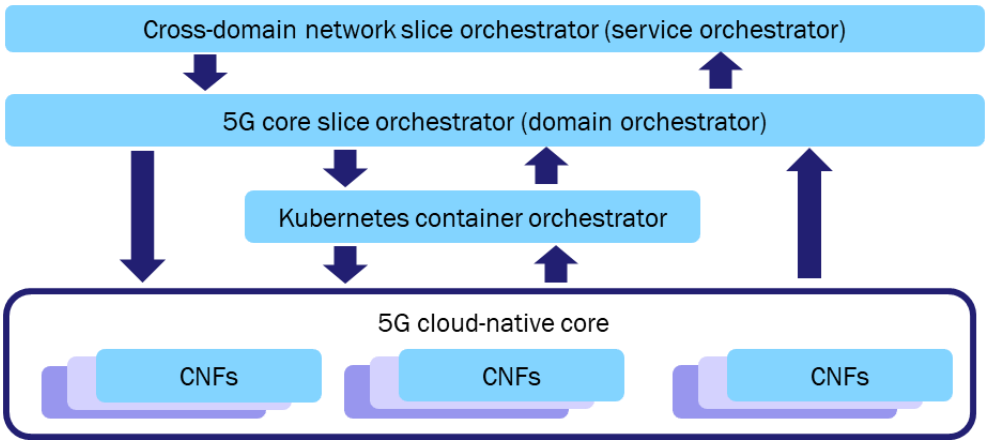
Source: Cisco and Analysys Mason, 2020

6. The 5G core domain orchestrator enables slicing in the cloud-native 5G core

A key goal of the containerization of 5G network functions is to be able to deploy such functions more granularly than with monolithic or virtual machine-based designs, thereby enabling rapid instantiation and the ability to scale only those functions that need it. To achieve this goal, the core network orchestration capability must evolve beyond its traditional role of instantiating and activating VNFs and incorporate additional functionality to manage CNFs. To do this, the orchestrator must work closely with the Kubernetes (K8s) container orchestrator to manage the lifecycle of 5G control plane CNFs.

Additionally, the 5G core orchestrator, in tandem with the cross-domain slice orchestrator, must facilitate the lifecycle operations of a core network slice instance, including slice instantiation, provisioning, service chaining, modification and termination. (Figure 10). The network slice use case and the slice intent will dictate the slice parameters for each of these operations (such as whether the network functions should be dedicated or shared).

Figure 10: 5G core and slice orchestration



Source: Analysys Mason, 2020

6.1 The 5G core domain orchestrator must interwork with the Kubernetes orchestrator

The mobile core network functions for 5G SA will be developed as CNFs using cloud-native computing technologies such as K8s.

Cloud-native architecture is a requirement not only for 5G networks, but also for the applications deployed on top of these networks. Service providers need to fully understand this architecture, which is characterized by containerized applications composed of microservices. Microservices-based applications are deployed and managed by the K8s container orchestrator, the abstraction layer that permits the portability of CNFs and other containerized applications across different infrastructure.

K8s facilitates unprecedented levels of service scalability and automation, but it also presents a challenge for service providers because they must decide what level of control should be exposed for management by the K8s container orchestrator versus the 5G core orchestrator. Therefore, service providers need an orchestration solution that can manage the VNFs/CNFs as well as manage the K8s clusters and the associated microservices.

6.2 The dynamic core slice orchestration function will be delivered as part of the 5G core domain orchestrator

Service providers are still at a very early stage in their journey to deploying a full 5G network with a cloud-native core. As explained in Section 2, service providers are expected to support a broad variety of use cases across a range of industry verticals. The associated network slicing requirements are therefore expected to be quite diverse and subject to change.

As an example, service providers may have to support use cases where the session management function (SMF) and user plane function (UPF) support multiple network slices, as well as those where there are dedicated SMFs and UPFs for each network slice. Furthermore, the evolving network and slice performance may lead to a change from shared to dedicated instances to meet the SLA. As such, decisions on the network function deployment and the associated configuration may happen in a dynamic environment based on evolving performance requirements. A model-based approach will therefore be required to adapt to the changing needs of network slices. With a model-based approach, service providers can customize the deployment scenarios based on the evolving use case requirements, rather than specific static configuration options.

These capabilities will be developed as part of the 5G core orchestrator, which will act as a domain-specific orchestrator for the cloud-native core. It is expected to contain the slice orchestration functions that will map onto the 5G core NSSMF as defined by the 3GPP. The intent of the overall slice will be consumed by the end-to-end cross-domain network slice orchestrator (the NSMF) and the core domain-specific intent will be expressed towards the 5G core domain orchestrator, which will translate it in order to configure and customize the VNFs/CNFs in the cloud-native core.

7. Conclusion

5G network slicing enables service providers to move away from the rigid 'one size fits all' business model to offer differentiated enterprise connectivity services such as ultra-low latency and massive IoT services with varying network performance characteristics. To deliver on this promise, service providers need an end-to-end network slice orchestration and operations solution that can create and provision network slices spanning the RAN, edge, transport and core network domains.

The slice orchestration solution must be cross-domain by design, and must support multi-layer architecture in order to abstract domain-level complexities and delegate slice resource allocation responsibilities to the relevant domain-level slice orchestrators. The solution must also demonstrate multi-vendor support and use model-driven architecture in order to encapsulate the end-to-end slice characteristics with a reasonable level of abstraction, while allowing the domain-level slice orchestrators to take care of the specifics of configuring the network domain resources, irrespective of the form factor (physical, VM-based or containers) and the supplier.

Transport domain network slicing will be enabled by new IP routing technologies, such as segment routing, that can be configured to address key slice characteristics such as the QoS and path forwarding policy. A transport domain controller can be used for on-demand configuration, thereby providing a higher level of dynamic control for transport network slices. Service providers can achieve scale by maintaining a catalog of predefined slice forwarding templates that can be rapidly instantiated upon request.

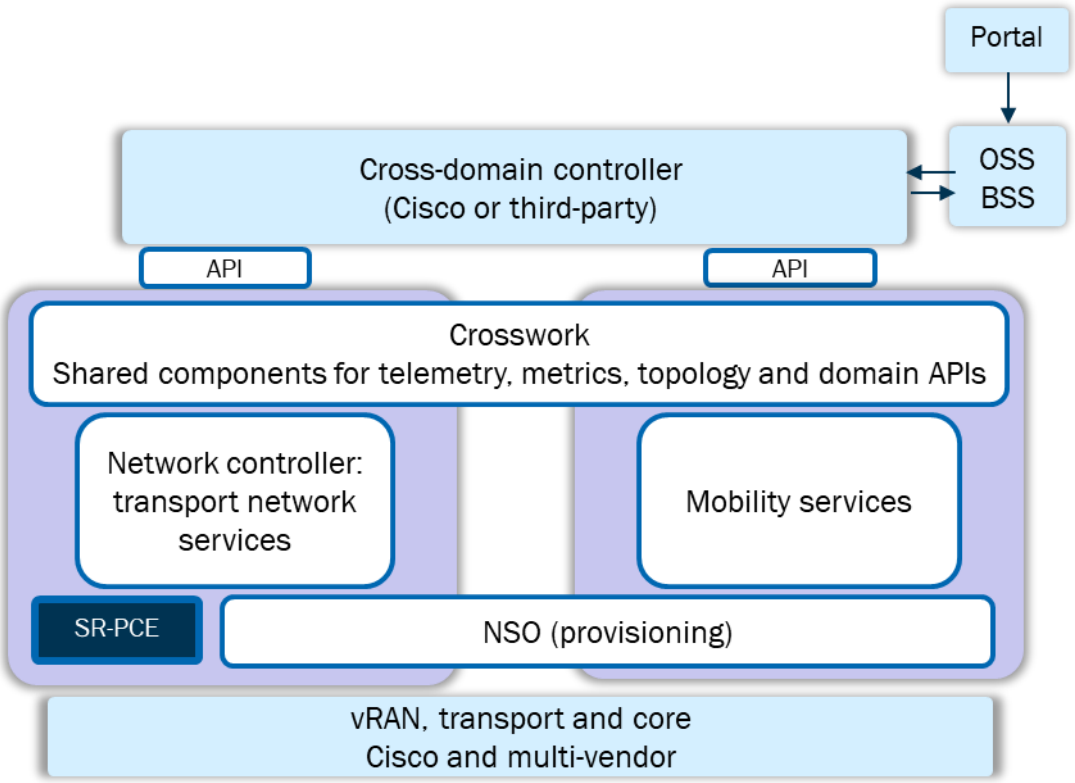
Core domain network slicing will be enabled by the 5G core domain orchestrator and will be made possible by the development of the containerized cloud-native core as part of the 5G SA implementation. The core domain orchestrator will contain the core network slicing functions and will closely interwork with the K8s orchestrator to instantiate and configure the CNF resources for the core slice.

8. Cisco's network slicing solution

Cisco's network slice orchestration solution is built using slicing-enabled components with the ability to span the entire service chain. Cisco believes that it is important to build a solution that is flexible and can evolve as standards, technologies and design patterns change. Slicing exists as part of a larger business process and must easily integrate into that business workflow. These perspectives have shaped the current portfolio and are driving future investment decisions.

Figure 11 depicts Cisco's network slicing solution, which is built on the Crosswork Network Automation suite combined with the Cisco Network Services Orchestrator (NSO) product.

Figure 11: Cisco's network slice orchestration solution



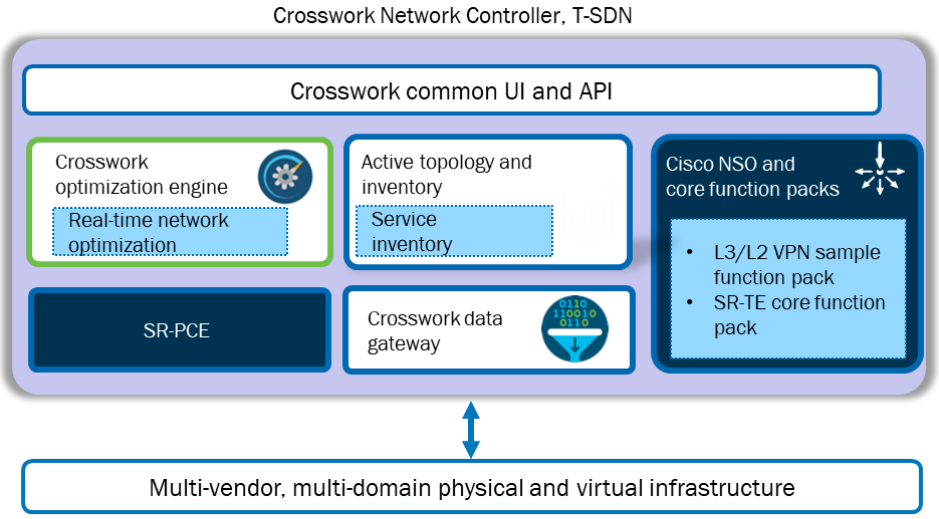
Source: Cisco, 2020

Cisco Crosswork and Cisco NSO are aligned to the 3GPP structure and bring the capabilities for managing the 5G environment to each of the domains listed in the following sections.

8.1 Transport network slicing

Cisco's Crosswork Network Controller is the company's transport domain controller and is depicted in Figure 12.

Figure 12: Cisco's Crosswork Network Controller, with associated use cases



Use case	Description
Service-oriented transport provisioning	Provision segment routing traffic-engineering policies for services with SLAs.
Service provisioning	Provision L2VPN and L3VPN services (sample)
Bandwidth optimization	Tactically optimize the network during times of congestion
Real-time network optimization	Collect real-time performance information and optimize the network as needed to maintain the SLA
Topology and inventory	Collect and expose information on network and services

Source: Cisco, 2020

The Crosswork Network Controller uses intent-based network automation techniques to deliver capabilities for service orchestration and fulfilment, network optimization, service path computation, device deployment and management and fault remediation.

It offers service providers a turnkey network automation solution that delivers increased service agility, cost efficiency and optimization in order to increase customer value and lower operating costs.

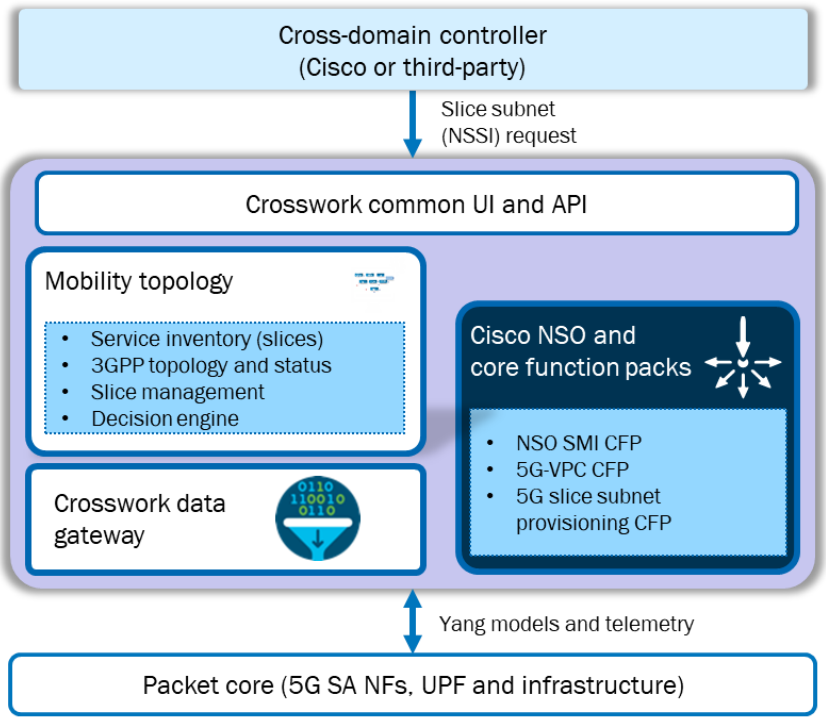
The fully integrated solution combines core capabilities from multiple products, including the Cisco NSO, Cisco Segment Routing Path Computation Element (SR-PCE) and the Cisco Crosswork suite of applications, with common APIs and integrations with a common user interface, providing control via a single pane of glass

8.2 Core network slicing

Cisco provides 3GPP-compliant, slicing-capable core network functions such as the SMF, AMF and PCF in addition to the transport UPF functions. Cisco uses the same tooling with the Cisco NSO to provide resource-facing automation to control the lifecycle of the network functions and the lifecycle of the service. Cisco's NSO has been extended to control the infrastructure required to support the lifecycle management of VNFs and CNFs through a platform designed from the ground up for the next-generation services, thereby enabling 5G deployment for both consumer and enterprise use cases.

The efficient management of the infrastructure platform, data plane, control plane and network functions required for the 5G core that is provided as part of the solution ensures a strong foundation for closed-loop automation (Figure 13).

Figure 13: Cisco's Crosswork Core Controller, with associated use cases



Use case	Description
Core provisioning	Provisioning workflow: 1. Customer requests new NFs based on policy 2. Mobility controller applies capacity information and decision logic 3. Drives NSO for deployment and provisioning of 5G SA core NFs
Slice manager	Provisioning workflow: 1. Receive new slice subnet request based on policy 2. Mobility controller applies capacity information and decision logic 3. Drives NSO for deployment and provisioning of slice parameters in mobility core, as needed
Topology, inventory and telemetry	Collect and visualize mobility core data, integrate for use in slice planning and decisions • Packet core topology and slice topology • Slice subnet inventory • Slice subnet status and metrics

Source: Cisco, 2020

The ability to provide core network function slicing allows service providers to efficiently operate the core of the network and provide differing levels of service on a per-slice basis. The complexity with core network slices is that the number of slices could be greater than the number of customers connected to the network. The ability to provision and assure those slices increases with the number of network functions and use cases supported by the core network.

8.3 Unifying slice control with NSO

Cisco's NSO also forms a strong foundation for building the 5G cloud-native core and 5G network slicing use cases. Historically, Cisco has been able to use NSO for many use cases that require provisioning for both single vendor and multi-vendor environments. Now, with the development of function packs on NSO that are tuned for the easy consumption of product capabilities, Cisco is adding provisioning capabilities for 5G.

Furthermore, NSO can be deployed in a tiered architecture, where the same baseline software is used for provisioning engines for an NSMF and an NSSMF. Combining Cisco NSO with the assurance capabilities of Crosswork enables assurance-driven slice orchestration and lifecycle automation. This combination forms the basis for a 5G core controller that allows for the provisioning of the slice within that domain such that the service level and intent for the slice can be properly maintained and modified as needed.

9. About the author

Anil Rao (Principal Analyst) is the lead analyst on network and service automation research that includes the Network Automation and Orchestration, Automated Assurance and Service Design and Orchestration research programmes, covering a broad range of topics on the existing and new-age operational systems that will power operators' digital transformations. His main areas of focus include service creation, provisioning and service operations in NFV/SDN-based networks, 5G, IoT and edge clouds; the use of analytics, ML and AI to increase operations efficiency and agility; and the broader imperatives around operations automation and zero touch networks. Anil also works with clients on a range of consulting engagements such as strategy assessment and advisory, market sizing, competitive analysis and market positioning, and marketing support through thought leadership collateral.

This whitepaper was commissioned by Cisco. Analysys Mason does not endorse any of the vendor's products or services.

Published by Analysys Mason Limited • Bush House • North West Wing • Aldwych • London • WC2B 4PJ • UK
Tel: +44 (0)20 7395 9000 • Email: research@analysysmason.com • www.analysysmason.com/research

Registered in England and Wales No. 5177472