NOKIA

Dynamic end-to-end network slicing for 5G

Addressing 5G requirements for diverse services, use cases, and business models

White Paper

Next-generation 5G networks will enable a fully mobile and connected society. The proliferation of wearable devices and wireless connected objects will pave the way to a wide variety of new use cases and business models. But a fundamental rethinking of the mobile network is needed to realize the full potential 5G offers to support a variety of very diverse and extreme requirements for latency, throughput, capacity, and availability. The network architecture must shift from the current network of entities architecture to a network of capabilities architecture. And network models must shift from the current network for connectivity model to a network for services model. Network slicing offers an effective way to meet the requirements of all use cases using a common network infrastructure. This paper outlines a dynamic network slicing concept that enables design, deployment, customization, and optimization of different network slices on a common infrastructure. The concept leverages Nokia products and solutions and Nokia Bell Labs innovations in cloud mobile access and core, software-defined networking (SDN), network function virtualization (NFV), containerization, end-to-end orchestration, network applications, and analytics.

NOKIA

Contents

The need for 5G network slicing

The next generation wireless network will need to support new demands from a wide variety of users, machines, industries, governments, and other organizations. It will be a door opener for new possibilities and use cases, many of which are as yet unknown.

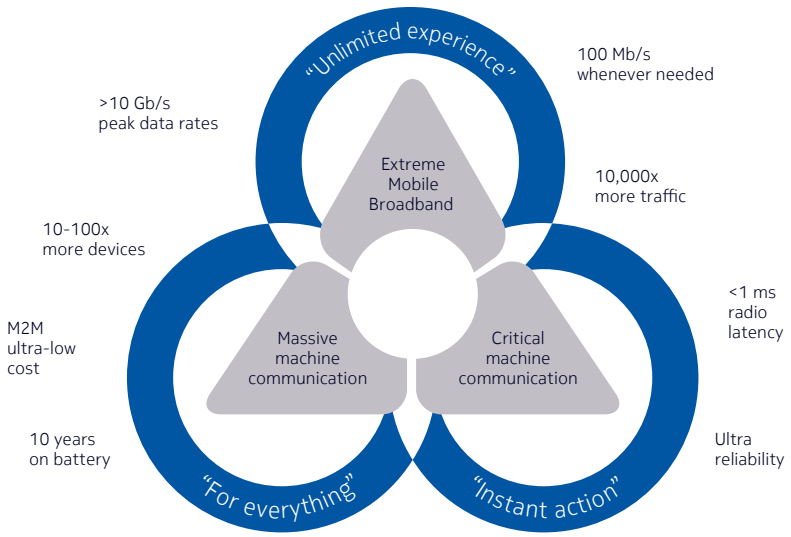
For example, 5G networks will connect the factory of the future and help create a fully automated and flexible production system.

In the healthcare industry, hospitals will be able to arrange remote robotic surgeries as if the surgeon were physically present next to the patient. At the same time, 5G connected healthcare chips will constantly monitor vital signs, prevent conditions from becoming acute, and adapt medication to meet changing conditions.

While on the roads, self-driving cars and smart infrastructures enabled by 5G networks will reduce accidents and save millions of lives every year.

Obviously, the "one-size-fits-all" network solution we have today for all use cases and services to every device everywhere is no longer viable. To enable new possibilities, 5G networks will have to support a variety of very diverse and extreme requirements for latency, throughput, capacity, and availability (Figure 1).

Figure 1. 5G must support a variety of very diverse and extreme requirements



NOKIA

A fundamental rethinking of the mobile network is needed to support all use cases and all requirements:

- The network architecture must shift from the current network of entities to a network of capabilities architecture.
- Network models must shift from the current network for connectivity to a network for services model.

Network slicing offers an effective way to meet all of the diverse use case requirements and exploit the benefits of a common network infrastructure. It enables operators to establish different capabilities, deployments, and architectural flavors for each use case or service group and run multiple network instances in parallel.

Software-defined networking (SDN) and network function virtualization (NFV) will play an important role in the shift to network slicing.[1] Virtualization will enable separation of the software from the hardware and offer the possibility to instantiate many functions on a common infrastructure. With this approach, the infrastructure can be shared by different tenants and provide different services.

This paper outlines a dynamic network slicing concept that enables design, deployment, customization, and optimization of different network slices and services on a common network infrastructure. And it explains how dynamic network slicing leverages Nokia products and solutions and Nokia Bell Labs innovations in cloud mobile access and core, SDN, NFV, end-to-end orchestration, network applications, and analytics.

5G network slicing requirements and approach

Yesterday's and today's mobile networks were designed for the delivery of personal communication services and content, such as voice, video, and web browsing. Next-generation 5G networks must enable a fully mobile and always connected society. The proliferation of wearable devices and wireless connected objects will pave the way to a wide variety of new applications and services. In addition to more pervasive human-centric applications, such as virtual reality augmentation and UHD video, 5G networks will support machine-to-machine (M2M) and human-machine type applications that will make our lives safer and smarter. These applications will automate and mobilize a variety of industries, including energy, health, public safety, smart city, manufacturing, logistics, media, and automotive.[2]

Beyond connectivity, 5G will offer operators unique opportunities to create new business models and enable new use cases for consumers, enterprises, and industry specific services, as well as content and application providers.

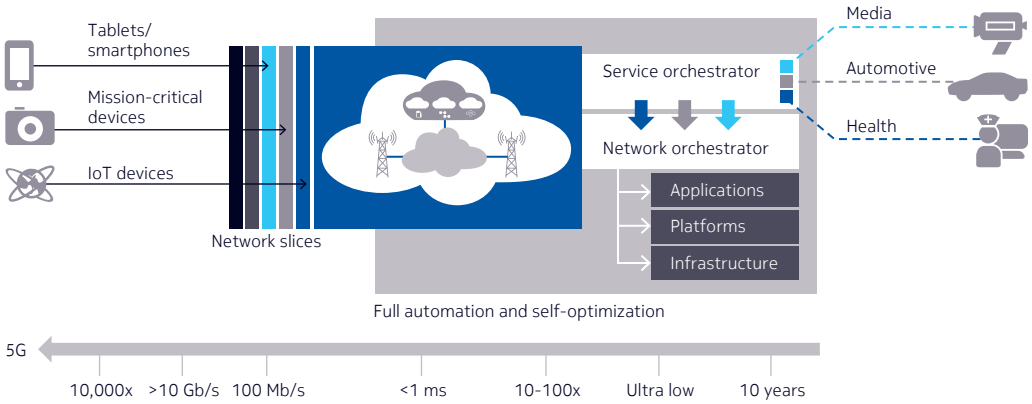
1 "Relationship of SDN and NFV", issue 1, Open Networking Foundation, Technical Report 518, October 2015. https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical-reports/onf2015.310_Architectural_comparison.08-2.pdf

2 "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond", ITU-R Recommendation M.2083.

Meeting the diverse requirements of this large set of stakeholders with current network architectures and model is difficult due to the variety of requirements that must be addressed at any point in time. But, as noted by the Next Generation Mobile Networks (NGMN), the success of 5G technology will come from its ability to provide multiple value solutions for all requirements and at the same time provide all stakeholders with a unique solution tailored to their needs.[3]

5G network slicing can be used to ensure that end-to-end performance meets customer expectations, as well as service and application requirements (Figure 2). To leverage network slicing properly, the individual segments (radio access network (RAN), transport, metro, core, edge cloud, central cloud), which were formerly treated separately, must be examined as a whole. And, performance optimization must be adapted and coordinated, across the entire network.

Figure 2. 5G network slicing can address a variety of industry and multi-tenancy use cases with multiple independent instances on one physical network



An effective 5G network slicing process must:

- Support a variety of business models, including industry-specific models and multi-tenancy
- Significantly reduce new service creation and activation times
- Provide extreme agility in the network to meet diverse service needs
- Provide massive elasticity in the network to meet very dynamic traffic demands
- Exploit analytics and context to adapt services and networks predictively and in real time
- Enable an open services ecosystem where different parties can cooperate to introduce innovative services tailored to specific user or industry demands
- Expose actionable network insights to application and content providers, enterprises, and industry verticals

3 "5G White Paper", Next Generation Mobile Networks (NGMN), March 2015, http://ngmn.org/5g-white-paper.html

- Provide full programmability to enable easy integration of new network capabilities, extension of existing capabilities and easy creation of new services and business models
- Intelligently manage and orchestrate resources and capabilities for dynamic (re)configuration of the network to meet end-to-end performance
- Support a high level of automation powered by advances in analytics, and machine learning

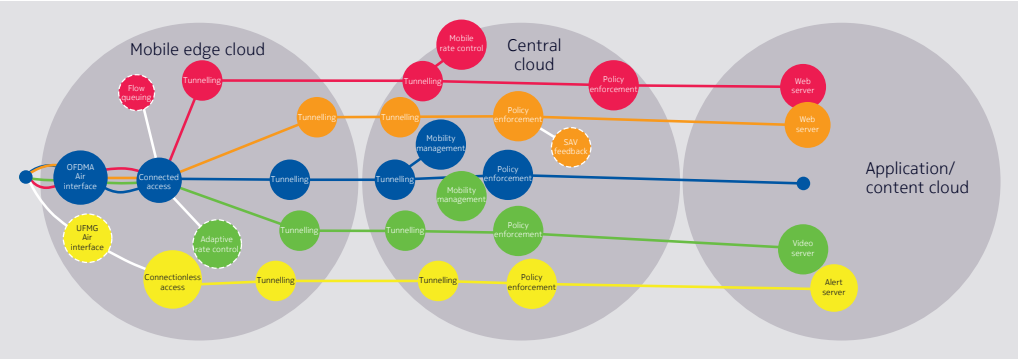
Addressing requirements with dynamic network slicing

The dynamic network slicing concept extends the slicing vision of the NGMN and addresses all requirements. It partitions a common network infrastructure into multiple, logical, end-to-end, virtual network instances or slices with several key characteristics:

- The slices support a group of services, use-cases, and business models with similar requirements. For example, an operator can run enhanced broadband slices to offer a variety of broadband services to its customers, which include web browsing, audio and video streaming, and chat.
- The slices are built with only relevant network capabilities that match the needs of the supported service, use case or business case. For example, an ultra-low latency capability can be created for a slice supporting ultra-low latency use cases. The capabilities in the slice are not restricted to the user plane. The slices can also control and manage plane-relevant capabilities, such as a dynamic video stream controller or a specific type of billing application relevant to the business case.
- The slices are dynamic in runtime. They include an automation framework that uses real-time analytics and monitoring for efficient use of network and cloud resources, and optimization for the dynamic needs of services or dynamic traffic demands.

Figure 3 shows how multiple network slices would appear in a network based on the above characteristics. The slices are created and deployed over a distributed cloud infrastructure, each with unique capabilities dedicated to a supported service group, use case, or business model.

Figure 3. Dynamic network slices running over a single infrastructure



To enable this approach, dynamic network slicing leverages key technology advancements in:

- Distributed cloud infrastructure and cloud native applications
- NFV
- End-to-end orchestration
- SDN and programmable networking
- Network big data, analytics, and machine learning
- Services oriented architectures
- Intent based network programming

In addition, Nokia has pioneered a number of new technologies and innovations that address specific network slicing challenges (Figure 4).

Figure 4. Network slicing challenges and innovations

The challenges

- Diversity of next-generation services creates extreme divergence in network requirements
- Monolithic network functions creates a lack of elasticity and scalability
- Rigid infrastructure makes it difficult to evolve and extend the network with new capabilities
- Statistically placed network functions require forklift upgrades to relocate
- Complexity of new network makes it difficult to manage with traditional management systems
- Common network capabilities for diverse service types
- Extremely large number of service paramaters, due to diversity of use cases

The innovations

- Dynamic network slices, each tailored for corresponding services
- Modular and granular network capability units (NCU) that are easy to scale up or down
- Highly agile cloud infrastructure that easily evolves and extends network capabilities
- Dynamically orchestrated lightweight NCUs that can easily be placed or relocated for optimum performance
- Automation powered by distributed native analytics (DNA)
- Differentiated network capabilities uniquely tailored for diverse service types
- A managed group of Labeled Service Parameter Sets (LSPs)

Key principles

This dynamic network slicing approach is guided by several key principles, as outlined below.

A distributed cloud and virtualization infrastructure

The infrastructure to support 5G network and network slices is built using proven virtualization technologies, and cloud native applications. Wherever possible, the 5G network functions and capabilities are developed upon a distributed cloud and virtualization infrastructure. And dedicated and purpose-built network entities are only used when necessary.

A network of capabilities instead of network of entities

The 5G network slices are composed of modular network capabilities working in unison. Network Capability Units (NCUs) are the abstractions of these network capabilities. The NCUs can have varying degrees of granularity, but they will all be modular for easy plug-and-play deployment and operation. The NCUs can be implemented as containerized applications, as native micro services, or as complete virtualized network functions.

NOKIA

A network for services instead of a network for connectivity

The 5G network slices are designed from the ground up to offer and support classes of services. This is a paradigm shift from past networks, which were primarily built to offer connectivity. With this approach, the 5G network slices will have unique capabilities that are required for the supported group of services. Moreover, the capabilities of each slice can be dynamically optimized to meet the specific needs of individual services.

Easy design and dynamic creation of end-to-end slices

Dynamic 5G network slices can be easily designed by packaging necessary NCUs into a forwarding graph. Network slices can be created dynamically by orchestrating these forwarding graphs on a distributed cloud infrastructure using proven orchestration and management technologies. A template-based approach can be used to ease the creation of slices and to eliminate routine errors. The templates can be used to create various slices and then each slice can be further customized based on the needs of supported services, use cases, or business models.

Dynamic programmability and control

The 5G network slices support dynamic programmability and control by leveraging SDN principles. And the dynamic programming of network slices can be accomplished either by custom programs or within an automation framework driven by analytics and machine learning.

Automation of the network operation and optimization

Creating more network slices inherently adds complexity to network operations and ongoing optimization efforts. Therefore, the dynamic network slices are designed from the ground up to enable maximum possible automation of operations and optimizations. Automation is enabled by analytics, machine learning, network big data, and network programmability

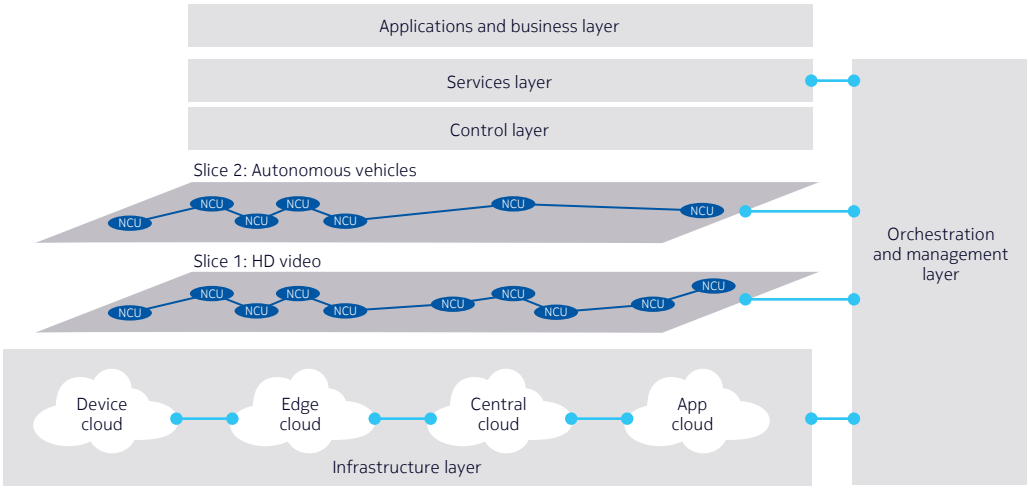
End-to-end perspective and approach

The 5G network slices are composed of capabilities from multiple network segments that span the network from access to core, as well as network applications. This end-to-end perspective is needed to meet the needs of diverse services, use cases, and business models. It eliminates the need to make assumptions about what specific sub-segments in the network will or will not support.

Conceptual architecture

Figure 5 presents a conceptual architecture for a 5G network that uses the dynamic network slicing concept.

Figure 5. Conceptual architecture for dynamic network slicing in 5G



In this architecture, the services layer acts as the logical interface between network and business applications. It provides abstraction of the network towards applications and interfaces for easy service creation and optimization. It also helps to reduce the large number of individual service parameters into a manageable group called a Labeled Service Parameter Set (LSPS).

The control layer hosts the logical end-to-end control for the network. It provides abstraction of the network towards the services layer. In deployments, the control layer may consist of multiple controllers coordinated or federated appropriately to provide end-to-end control and programmability.

The orchestration and management layer supports design, creation, and activation of individual slices on the common infrastructure.[4] Orchestration and management is enhanced using analytics and machine learning. In addition, analytics are exploited to optimize the infrastructure resources within a slice, as well as across different slices sharing the same infrastructure.

Finally, the infrastructure layer hosts the physical and virtual resources needed to create end-to-end network slices. These include both virtualization software and hardware comprised of memory, compute, storage, and networking resources.

4 ETSI specification, Network Function Virtualization (NFV): Management and Orchestration, ETSI GS NFV-MAN 001, V1.1.1, December 2014.

NOKIA

Conclusion

The Nokia dynamic network slicing concept enables dynamic end-to-end slicing of 5G networks. It supports the very diverse and extreme requirements for latency, throughput, and availability operators will have to meet to deliver 5G services to a wide variety of users, machines, industries, and other organizations. It is extremely important to consider that all parts of the network — from radio access, to transport and core — must be included to achieve the desired levels of service. With this concept, operators could create different deployments and architectural flavors for each business model, use case or service group. They could run all network instances in parallel on a common network infrastructure. And they could open the door to new human possibilities of technology, enabled by new services, use cases, and business models.

Acronyms

5G	5th Generation
LSPS	Labeled Service Parameter Set
NCU	Network Capability Unit
NFV	Network Function Virtualization
RAN	Radio Access Network
SDN	Software-defined Networking

Nokia Oyj
Karaportti 3
FI-02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Product code: PR1604019341EN (June)