NGMN 5G P1
Requirements & Architecture
Work Stream End-to-End Architecture

Description of Network Slicing Concept

by NGMN Alliance

Version:	**1.0.8 (draft)**
Date:	**14th September 2016**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P - Public**
Authorised Recipients: (for CR documents only)	

Project:	**5G**
Editor / Submitter:	**Sebastian Thalanany (US Cellular)** **Peter Hedman (Ericsson)**
Contributors:	**NGMN P1 WS1 E2E Architecture team**
Approved by / Date:	

Abstract: Short introduction and purpose of document

This document describes further details of Network Slicing including the Network slicing concept and definitions.

Document History

Date	Version	Author	Changes
31.08.2015	0.0.2	Peter Hedman **NGMN P1 WS1 E2E Architecture team**	First Version (Draft)
11.09.2015	0.0.3		Scope section removed. Problem statement updated according to mail discussions. Definition section updated with draft proposal.
02.12.2015	0.0.4		Figure and related text updated Resource definition updated
03.12.2015	0.0.5		Changes as agreed at the conference call 2/12 Abbreviations added
11.12.2015	0.0.6		Clean-up as per mail review.
16.12.2015	0.1.0		Minor changes to concept and definition. Version discussed and agreed in the NGMN P1 WS1 E2E Architecture team.
13.01.2016	1.0	Klaus Moschner, NGMN Office	Formatting updates and updates following Board Approval
05.05.2016	1.0.1		Section 5 added
13.07.2016	1.0.2		Updates to Section 5
09.08.2016	1.0.3	Sebastian Thalanany	Updates for clarity, including consistency across figure formats and names, in Section 4, and Section 5. Post August 3, 2016 call.
12.08.2016	1.0.4	Sebastian Thalanany	Added a figure in Section 4.
16.08.2016	1.0.5	Sebastian Thalanany	Revised figure addition in Section 4,
17.08.2016	1.0.6	Sebastian Thalanany	Revisions. Post August 17, 2016 call
07.09.2016	1.0.7	Sebastian Thalanany	Editorial updates. Post September 07, 2016
14.09.2016	1.0.8	Steve Tsang Kwong U	Editorial and Formatting updates

Table of Content Contents

1 INTRODUCTION

The NGMN 5G White Paper-Version 1.0-published in February 2015 has positioned the demands and business requirements beyond 2020 and introduced the network service deployment concept of Network Slicing with some high level examples.

In order to enable SDOs to develop viable standards for Network Slicing there is a need to refine the concept and establish a common terminology for Network Slicing. This would also help actors across the industry to have a common understanding of the expectations and the characteristics of such an important concept.

This document describes further details of the concept of Network Slicing and provides common terminology. It is assumed that the relevant use cases are detailed elsewhere e.g. in the 3GPP TR 22.891 [1]; some relevant network operation requirements will also be listed e.g. in 3GPP TR 22.864 [2].

2 REFERENCES

[1] 3GPP TR 22.891: "Feasibility Study on New Services and Markets Technology Enablers"

[2] 3GPP TR 22.864: "Feasibility Study on New Services and Markets Technology Enablers – Network Operation"

3 PROBLEM STATEMENT

The 5G use cases demand very diverse and sometimes extreme requirements. The current architecture utilizes a relatively monolithic network and transport framework to accommodate a variety of services such as mobile traffic from smart phones, OTT content, feature phones, data cards, and embedded M2M devices.
It is anticipated that the current architecture is not flexible and scalable enough to efficiently support a wider range of business needed when each has its own specific set of performance, scalability and availability requirements. Furthermore, introduction of new network services should be made more efficient. Nevertheless, several use cases are anticipated to be active concurrently in the same operator network, thus requiring a high degree of flexibility and scalability of the 5G network.

4 THE CONCEPT OF NETWORK SLICING

4.1 Network slicing concept

As depicted in Figure 1, the network slicing concept consists of 3 layers: 1) Service Instance Layer, 2) Network Slice Instance Layer, and 3) Resource layer.

The notion of an "instance" of any entity is a run-time construct, and should be construed as being derived from a design time or configuration time "template" or a "blueprint"

The Service Instance Layer represents the services (end-user service or business services) which are to be supported. Each service is represented by a Service Instance. Typically services can be provided by the network operator or by 3rd parties. In line with this, a Service Instance can either represent an operator service or a 3rd party provided service.

A network operator uses a Network Slice Blueprint to create a Network Slice Instance. A Network Slice Instance provides the network characteristics which are required by a Service Instance. A Network Slice Instance may also be shared across multiple Service Instances provided by the network operator.

NOTE: Whether there is a need to support sharing of Network Slice Instances across Service Instances provided by different 3rd parties is up for discussion in SDOs.

The Network Slice Instance may be composed by zero, one or more Sub-network Instances, which may be shared by another Network Slice Instance. Similarly, the Sub-network Blueprint is used to create a Sub-network Instance to form a set of Network Functions, which run on the physical/logical resources.

A conceptual outline of a network slice is shown in Figure 1, depicted as a run-time example.

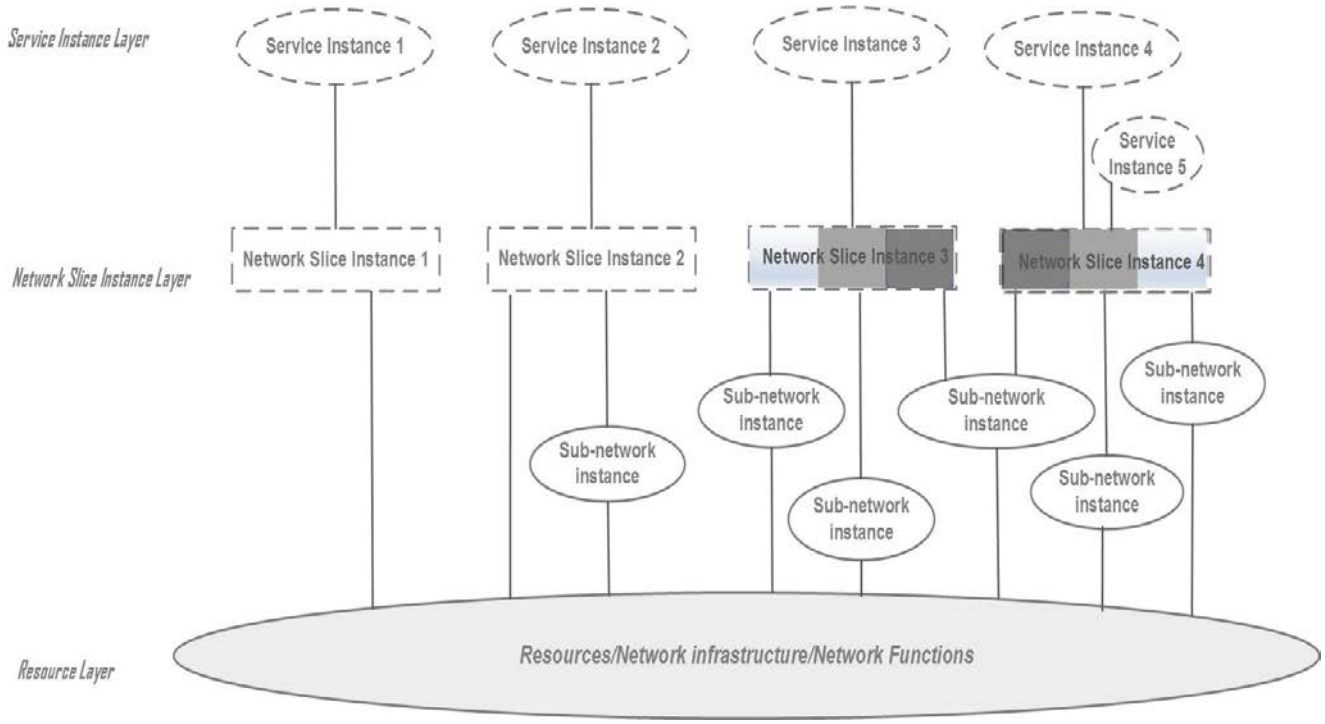

Figure 1: Network slicing conceptual outline

The concept of inheritance, across a “network slice blueprint” and a “network slice instance” is illustrated in Figure 2. For example a “Network Slice Blueprint-R” may be used to instantiate a single “Network Slice Instance-Y”. This example also illustrates that a single “Network Slice Instance-X” may be derived from a composite “Network Slice Blueprint-PQ” that has constituent “Sub Network Blueprint-P” and a “Sub Network Blueprint-Q”. This also depicts the notion of inheritance, where the constituents ““Sub Network Blueprint-P” and “Sub Network Blueprint-Q”, are inherited from a parent “Network Slice Blueprint-PQ”. The “Network Slice Blueprint” may also be a simple composition of “Sub Network Blueprints”, where there is no inheritance.

Some examples of a network slice instance: Enhanced MBB, M2M, Enterprise and Industry etc.
Example of a sub network instance: IMS (IP Multimedia Subsystem) etc.
The concept is extensible to any scenario envisioned for an application of the network slice framework.

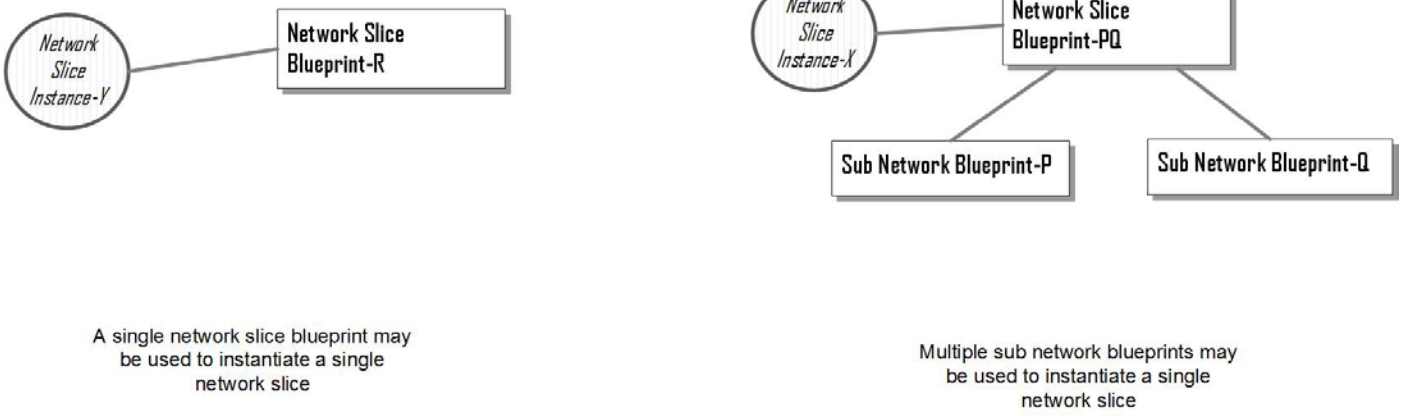

Figure 2: Conceptual model for relationships between "network slice blueprint" and "network slice instance

4.2 Definition

Service Instance: An instance is a run-time construct of an end-user service or a business service that is realized within or by a Network Slice

Network Slice Instance: a set of run-time network functions, and resources to run these network functions, forming a complete instantiated logical network to meet certain network characteristics required by the Service Instance(s).

- A network slice instance may be fully or partly, logically and/or physically, isolated from another network slice instance.
- The resources comprise physical and logical resources.
- A Network Slice Instance may be composed of Sub-network Instances, which as a special case may be shared by multiple network slice instances. The Network Slice Instance is defined by a Network Slice Blueprint.
- Instance-specific policies and configurations are required when creating a Network Slice Instance.
- Network characteristics examples are ultra-low-latency, ultra-reliability etc.

Network Slice Blueprint: A complete description of the structure, configuration and the plans/work flows for how to instantiate and control the Network Slice Instance during its life cycle. A Network Slice Blueprint enables the instantiation of a Network Slice, which provides certain network characteristics (e.g. ultra-low latency, ultra-reliability, value-added services for enterprises, etc.). A Network Slice Blueprint refers to required physical and logical resources and/or to Sub-network Blueprint(s).

Sub-network Instance: A Sub-network Instance is a run-time construct and comprises of a set of Network Functions and the resources for these Network Functions.

- The Sub-network Instance is defined by a Sub-network Blueprint.
- A Sub-network Instance is not required to form a complete logical network.
- A Sub-network Instance may be shared by two or more Network Slices.
- The resources comprise physical and logical resources.

Sub-network Blueprint: A description of the structure (and contained components) and configuration of the Sub-network Instances and the plans/work flows for how to instantiate it and control the Sub-network Instance during its life cycle. A Sub-network Blueprint refers to Physical and logical resources and may refer to other Sub-network Blueprints.

Physical resource: A physical asset for computation, storage or transport including radio access

- Network Functions are not regarded as Resources.

Logical Resource: Partition of a physical resource, or grouping of multiple physical resources dedicated to a Network Function or shared between a set of Network Functions.

Network Function (NF): Network Function refers to processing functions in a network.

- This includes but is not limited to telecom nodes functionality, as well as switching functions e.g. Ethernet switching function, IP routing functions
- VNF is a virtualized version of a NF (refer to ETSI NFV for further details on VNF).

5 GUIDELINES FOR APPLICABILITY ACROSS MULTIPLE ADMINISTRATIVE DOMAINS

5.1 General

The NGMN 5G white paper, Version 1.0 published, in February 2015 provides several types of partnerships and sharing options as new business models. As delineated in the paper: *"5G will foster innovation by flexible exposure of the network's value creation capabilities. This will enable partner-based propositions and allows for faster development and launch of these partner services at the benefit of all."*

In this section, the term "provider" refers to an entity that provides a service. The term "provider" is used as a generalized reference to include the term "operator", to accommodate the broader context of the 5G vision. Examples of such directions include Asset Providers (e.g. IaaS, PaaS, NaaS, Network sharing) and Partner Service Providers (e.g. operator service offering enriched by partner, partner service offering enriched by operator).

The network slicing concept described in this document provides a framework for broad applicability across a variety of usage scenarios that are envisioned to suit emerging and diverse business models in the industry. The scope of a network slice instance may include networks and devices. The network slicing concept is both a relevant and significant enabler for the sharing of a network slice instance, where more than one administrative domain is involved. For example, roaming is one scenario where more than one administrative domain is engaged over a network slice instance.

> *NOTE: An administrative domain is universally defined in IETF RFC 1136: "A collection of End Systems, Intermediate Systems, and subnetworks operated by a single organization or administrative authority. The components which make up the domain are assumed to interoperate with a significant degree of mutual trust among them, but interoperate with other Administrative Domains in a mutually suspicious manner". This generalized definition encompasses both wired and wireless network resources, virtualized or otherwise.*

A network slice instance may be wholly statically defined, e.g., as in fixed-access business or residential service, or partially dynamic, e.g., as in roaming mobile devices which may be connected to a statically-defined service chain, or fully constructed on demand.

Even when a network slice instance is statically defined, the necessary resources may be more or less abstract, e.g., as transport tunnels over a layered infrastructure network, or as VNFs located somewhere in a cloud. The actual physical resources, together with their configuration, may thus vary over the course of time, including on-demand allocation or scaling.

Various forward-looking business models that engage multiple administrative domains may be envisioned in the industry. An administrative domain refers to the scope of jurisdiction of a provider. A provider may obtain service capabilities from 3rd parties to enrich the services it provides to its end customers. A provider could also benefit from offering its spare capabilities or resources to a 3rd party. A network service can be a single user connectivity service, NaaS (Network as a Service) such as a service instance, a network slice instance or a subnetwork slice instance offering for a business vertical that utilizes forward-looking business models, or IaaS (Infra-structure as a Service).

A network slice blueprint is a template that describes how a network slice instance is to be created. The network slice blueprint may include resources or service capabilities from other providers with which an SLA exists.

The notion of a partnership between two providers is qualified in terms of the one which is hosting the service, and the one whose service is being hosted. A formalized description, of the roles that qualify the behaviour of a provider, is as follows:

P-Hosted: A service provider that provides services to e.g. end customers , which is allowed to negotiate with another provider (*P-Hosting)* based on a trust model, for the establishment of a hosted network slice instance or a hosted sub-network instance using functions and resources from the hosting domain.

Note: The necessary resources, in the hosting domain, are allocated based on a configured SLA between *P-Hosted and P-Hosting,*

P-Hosting: A service provider, which is allowed to negotiate with another provider (*P-Hosted)* based on a trust model, for providing the usage of functions and resources in the hosting domain towards the hosted domain.

Note: The necessary resources, in the hosting domain, are allocated based on a configured SLA between *P-Hosted and P-Hosting,*

Different types of partnerships and sharing may be envisioned, with a variety of distinctions:

- Various levels of functional exposure are considered, as envisioned in Section 4.5.2 of the NGMN 5G whitepaper:
  "…5G should provide an abstraction layer as an interface, where all types of in-networking functionality (control plane and data plane related) can be exposed to the application layer functions and/or service providers based on a service level agreement. Application/Service provider will then be able to use sub-set of the network capabilities in a flexible, configurable and programmable manner, and to use network resources depending on their service preference."
- Automated real time negotiation, as well as manual acquisition which implies different considerations.
- Static or dynamic configuration of a partnership
- Partnerships or agreements, may be based on one or more bilateral agreements for realizing any set of multiple partnerships:
  - A bilateral partnership or agreement is typically based on an SLA (Service Level Agreement) between two parties, where each of the two participating providers are enabled to provide the necessary resources for the realization of a service instance or a network slice instance.
  - For scenarios where the services of a broker are leveraged, there would be a pair of bilateral SLAs in place, where the broker behaves a trusted mediator for the realization of a service instance or a network slice instance.

5.2 Scenarios with multiple service providers

In general, there are two general categories of scenarios where network services need to be provided across multiple service providers:

- *Roaming scenario:* Individual users move from one provider (i.e. Home network provider) to a network managed by another provider (i.e. Visited network provider). The services that a user requires while roaming needs to be specified in the SLA between the two providers. In this case the two providers, with an SLA, would be the P-Hosted (Home network provider), and the P-Hosting (Visited network provider), with the corresponding behaviours required to support the inbound roamers (e.g. using a service instance or network slice instance) by P-Hosting.

- *Business verticals*: When a business vertical service user's request cannot be met by the capabilities of a single provider, the provider may harness the necessary capabilities from another provider, based on an SLA between the two providers. In this case the two providers, with an SLA, would be the P-Hosted (Home service provider), and the P-Hosting (Third-party service provider), with the corresponding capabilities required by the P-Hosted from the P-Hosting.

5.3 Informative illustration

The model of a bi-lateral partnership between a P-Hosting and P-Hosted is illustrated in Figure 3. In this illustrative example, an exposure of service capabilities, in the P-Hosting domain, exhibits available capabilities (e.g. API etc.) for the P-Hosted domain.

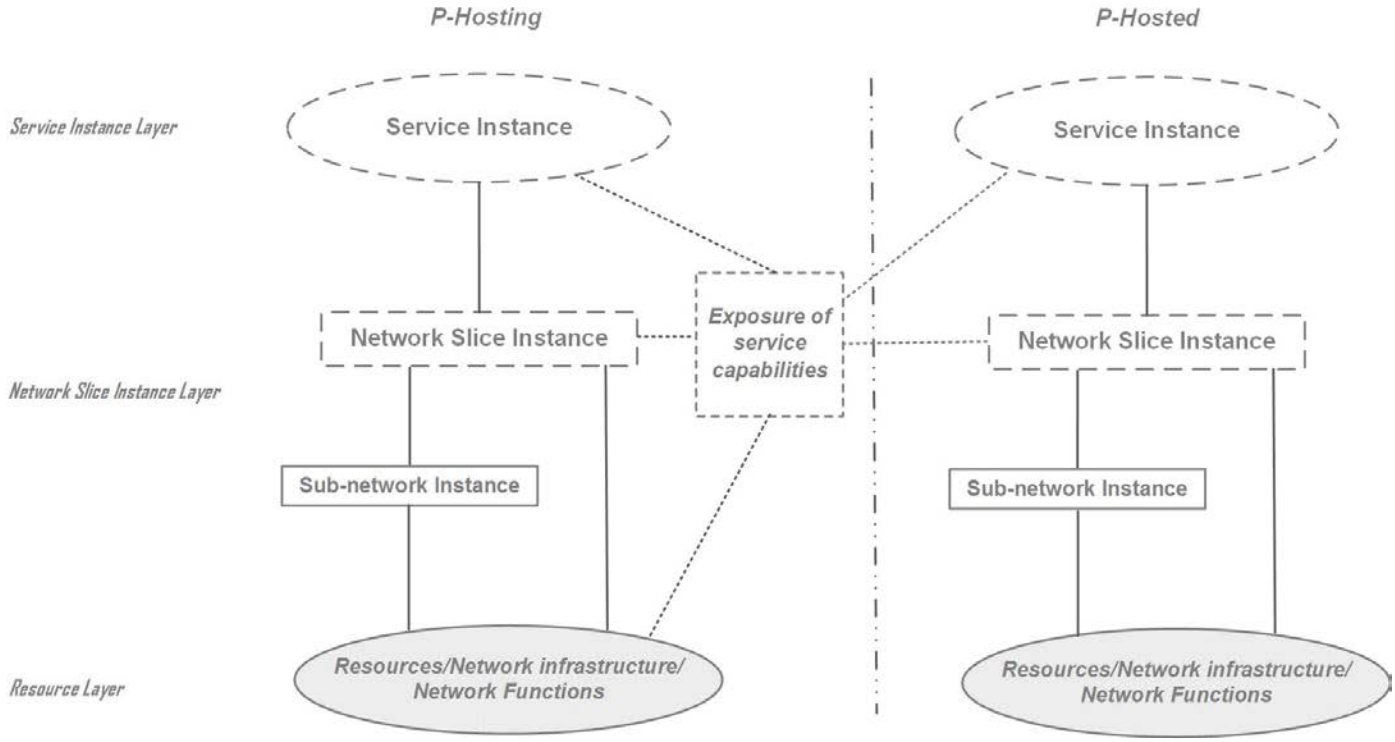

Figure 3: Model of a partnership between providers for the instantiation of a network slice

5.3.1 Service partnership behaviour

The service partnership behaviour between a P-Hosting and a P-Hosted, illustrated in Figure 3, is based on an SLA between the two, and is characterized as follows:

a) P-Hosting supports the resources for the realization and scaling of a service instance (for a business service) or a network slice instance to satisfy a service requested by P-Hosted.

b) P-Hosting supports the configuration information required for a realization of a service instance or a network slice instance, to satisfy the service requested by P-Hosted. The configuration information may consist of P-Hosting controlled network functions, parameters etc. associated with the service request made by P-Hosted. The P-Hosted requests the service level features needed and the P-Hosting creates (or scales an existing) the service instance (for the business service) or a network slice instance to be offered P-Hosted.

c) P-Hosted should be able to utilize P-Hosting resources, as its own based on the requirements of a requested service instance or a network slice instance (e.g. IMS profile, delegated HSS etc.). P-Hosted is given indirect control by allowing access to P-Hosting functions that are invoked under the control of the P-Hosting orchestration function.

d) P-Hosting should be able to offer resources to one or more P-Hosted entities, with adequate isolation among the resources that are offered to different P-Hosted entities.

e) Since P-Hosting serves the end customers of P-Hosted, a service request from P-Hosted may carry, among other requirements, for example geographical and temporal traffic distribution, types of application and their service requirements including network KPIs.
f) The realization of a service instance or a network slice instance in the hosting domain, based on the needs of P-Hosted may be accomplished through the use of APIs.

The partnership behaviour may be summarized as follows

- A provider realizes its service, after the appropriate network slice instance has been configured
- A network slice instance may be based on abstract resources, some of which may be acquired from other providers.
- A provider that contracts resources or service capabilities from another provider takes on the role of user of those resources or services.
- A provider may operate their own internal network slice instances in any way they deem appropriate, subject to conformance with SLAs associated with other providers.

6 ABBREVIATIONS

3GPP	3rd Generation Partnership Project
IaaS	Infrastructure as a Service
M2M	Machine to Machine
MNO	Mobile Network Operator
NaaS	Network as a Service
NF	Network Function
NSI	Network Slice Instance
OTT	Over The Top
PaaS	Platform as a Service
SDO	Standards Development Organisation
VNF	Virtual Network Function