

E2E Network Slicing Requirements

Version 1.0

21 July 2022

This is a Non-binding Permanent Reference Document of the GSMA

Security Classification: Non-confidential

Access to and distribution of this document is restricted to the persons permitted by the security classification. This document is subject to copyright protection. This document is to be used only for the purposes for which it has been supplied and information contained in it must not be disclosed or in any other way made available, in whole or in part, to persons other than those permitted under the security classification without the prior written approval of the Association.

Copyright Notice

Copyright © 2022 GSM Association

Disclaimer

The GSM Association ("Association") makes no representation, warranty or undertaking (express or implied) with respect to and does not accept any responsibility for, and hereby disclaims liability for the accuracy or completeness or timeliness of the information contained in this document. The information contained in this document may be subject to change without prior notice.

Compliance Notice

The information contain herein is in full compliance with the GSM Association's antitrust compliance policy.

[Note to editor: Include one of the following statements only; delete the other]:

This Permanent Reference Document is classified by GSMA as an Industry Specification, as such it has been developed and is maintained by GSMA in accordance with the provisions set out in GSMA AA.35 - Procedures for Industry Specifications.

OR

This Permanent Reference Document has been developed and is maintained by GSMA in accordance with the provisions set out in GSMA AA.34 - Policy and Procedures for Official Documents.

Table of Contents

1 Introduction 3
1.1 Overview 3
1.2 Scope 3
1.3 Definitions 4
1.4 Abbreviations 4
1.5 References 5
1.6 Conventions 7
2 High-level Architecture 7
2.1 High-level Reference Architecture 7
3 Proposed Technical Requirements 8
3.1 General 8
3.1.1 Co-operation of application layer and transport layer 8
3.1.2 SLA assurance in E2E 9
3.1.3 Deterministic and operator controlled UE behaviour 9
3.1.4 Multi-aspect resource optimization 11
3.1.5 Resource reservation and feasibility check 12
3.1.6 Automatic provisioning 12
3.1.7 Isolation 13
4 Pathways for a phased rollout of NSaaS 15
5 Conclusion 18
Annex A Collaboration with external organizations 19
A.1 Gap Analysis 19
A.1.1 Co-operation of application layer and transport layer 19
A.1.2 SLA assurance in E2E 19
A.1.3 Deterministic and operator controlled UE behaviour 19
A.1.4 Multi-aspect resource optimization 20
A.1.5 Resource reservation and feasibility check 21
A.1.6 Automatic provisioning 22
A.1.7 Isolation 22
Annex B Document Management 23
B.1 Document History 23
B.2 Other Information 23

1 Introduction

1.1 Overview

Network slicing is a concept of running multiple logically customized networks on a common infrastructure, for different customers in different industries and for different required functions, based on an agreed Service Level Agreement (SLA). Network slicing was outlined in the NGMN 5G White Paper [1] as a vision of 5G capabilities that drive value creation. After that, it has become one of the key capabilities specified by 3GPP to be supported in 5G systems [2]. Network slicing allows mobile operators in the 5G era to provide customized network services to their customers using a common network infrastructure. It is expected to have great commercial potential for the operators.

However, what should be taken into account is that, from the customer's perspective, the technical requirements and specifications are specified not only for mobile networks, as specified in 3GPP, but also for transport networks, management systems, and (user) devices, i.e., from an End-to-End (E2E) perspective. These network and technology domains must be well coordinated.

The GSMA has published a white paper on this topic [3]. It describes the network slicing architectural blueprint designed from an E2E perspective, spanning different technology domains (e.g., devices, access networks, core networks, transport networks, and network management systems). In addition, this whitepaper describes the technical aspects required for network slicing, and the gaps between the ongoing work and what is expected to be achieved by external organizations such as Standards Development Organizations (SDOs) and fora, as a snapshot of the current state of the art per 3GPP Release 17.

The previous work is continued in the present Permanent Reference Document (PRD). This document describes the technical requirements for E2E network slicing. It also maps the necessary specifications to specific external organizations to achieve the requirements, and identifies gaps between this PRD and these specifications. This mapping creates the potential for liaising between GSMA and external organizations asking them to fill the gaps.

1.2 Scope

This document provides technical requirements for E2E network slicing with the aim to align all technologies and their specifications for the correct creation of E2E network slices. The document provides network slicing requirements for operators from an E2E perspective and relates different architectures.

This document covers the following areas:

- Technical requirements for E2E network slicing
- Architectures, functions, and roles
    - Reference architectures envisioned for E2E network slicing
- Standardization and external organizations such as SDOs and forums, specifically:

- Gap analysis of standards: This PRD includes an analysis of the gaps in current standards and identifies the SDOs that are suitable to complete the E2E network slicing architecture.
- Review SDO progress and submit liaison statements to ensure that the E2E system is defined consistently across these organizations.

This version of the document focuses on technologies that are expected to be specified at the time of a completion of 3GPP Release 18. However, it also includes requirements to enable Network Slice as a Service (NSaaS), which is the ultimate goal of operators as mentioned in the GSMA white paper [3].

1.3 Definitions

Term	Description
E2E network slicing	Slicing concept for mobile network which include UE, RAN, CORE and Transport.

1.4 Abbreviations

Term	Description
3GPP	3rd Generation Partnership Project
5GS	5G System
AMF	Access and Mobility Management Function
AN	Access Network
API	Application Programming Interface
B2B2X	Business to Business to everything
CN	Core Network
CNF	Containerized Network Function
CU	Central Unit
DNN	Data Network Name
DRB	Data Radio Bearer
DU	Distributed Unit
E2E	End-to-End
E2EO	End-to-End Orchestrator
GA	General Availability
GBR	Guaranteed Bit Rate
GST	Generic network Slice Template
IMS	IP Multimedia Subsystem
KPI	Key Performance Indicator
MARO	Multi-Aspect Resource Optimization

Term	Description
MNO	Mobile Network Operator
NBI	Northbound Interface
NF	Network Function
NFV	Network Functions Virtualization
NFVO	Network Function Virtualization Orchestration
NRM	Network Resource Model
NS	Network Slice
NSaaI	Network Slice as Identifier
NSaaS	Network Slice as a Service
NSC	Network Slice Customer
NSP	Network Slice Provider
NSSAI	Network Slice Selection Assistance Information
OAM	Operation Administration and Maintenance
OSS	Operation Support System
PDU	Protocol Data Unit
PoCS	Proof of Concepts
PRB	Physical Resource Block
PRD	Permanent Reference Document
RAN	Radio Access Network
RRC	Radio Resources Connections
SDN	Software Defined Networking
SDO	Standards Development Organization
SLA	Service Level Agreement
SLS	Service Level Specification
SMS	Short Message Service
RCS	Rich Communication Service
TN	Transport Network
UE	User Equipment
URLLC	Ultra-Reliable and Low Latency Communications
URSP	UE Route Selection Policy
VNF	Virtualized Network Function

1.5 References

Ref	Doc Number	Title
	NGMN 5G White Paper	NGMN 5G White Paper (2015) https://www.ngmn.org/wp-

Ref	Doc Number	Title
	(2015)	content/uploads/NGMN_5G_White_Paper_V1_0.pdf
	3GPP, TS23.501	System architecture for the 5G System (5GS) https://www.3gpp.org/DynaReport/23501.htm
	GSMA, NG.127	E2E Network Slicing Architecture Version 1.0, June 2021. https://www.gsma.com/newsroom/wp-content/uploads//NG.127-v1.0-2.pdf
	RFC 2119	"Key words for use in RFCs to Indicate Requirement Levels", S. Bradner, March 1997. http://www.ietf.org/rfc/rfc2119.txt
	RFC 8174	Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words https://www.rfc-editor.org/info/rfc8174
	GSMA NG.116	Generic Network Slice Template. Version 6.0, November 2021 https://www.gsma.com/newsroom/wp-content/uploads/NG.116-v6.0.pdf
	3GPP TS28.530	Management and orchestration; Concepts, use cases and requirements https://www.3gpp.org/DynaReport/28530.htm
	Sensors	Wichary, T.; Mongay Batalla, J.; Mavromoustakis, C.X.; Żurek, J.; Mastorakis, G. Network Slicing Security Controls and Assurance for Verticals. Electronics 2022, 11, 222. DOI: 10.3390/electronics11020222. https://www.mdpi.com/2079-9292/11/2/222
	GSMA, White Paper	An Introduction to Network Slicing. White Paper 2017. https://www.gsma.com/futurenetworks/wp- content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf
	NGMN Alliance	Security Aspects of Network Capabilities Exposure in 5G v1.0. White Paper 2018. https://ngmn.org/wp-content/uploads/Publications/2018/180921_NGMN-NCEsec_white_paper_v1.0.pdf
	Sensors	Ordonez-Lucena, J.; Ameigeiras, P.; Contreras, L.M.; Folgueira, J.; López, D.R. On the Rollout of Network Slicing in Carrier Networks: A Technology Radar. Sensors 2021, 21, 8094. DOI: 10.3390/s21238094. https://www.mdpi.com/1424-8220/21/23/8094
	HEXA-X project	https://hexa-x.eu
	GSMA, White Paper	Network slicing Use case Requirements, 2018. https://www.gsma.com/futurenetworks/wp-content/uploads/2018/07/Network-Slicing-Use-Case-Requirements-fixed.pdf
	3GPP TS28.533	Management and orchestration; Architecture framework https://www.3gpp.org/DynaReport/28533.htm
	3GPP TS28.541541	Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and stage 3 https://www.3gpp.org/DynaReport/28541.htm
	3GPP TS24.526526	Telecommunication management; Study on the Self-Organizing Networks (SON) for 5G networks https://www.3gpp.org/DynaReport/24526.htm

Ref	Doc Number	Title
	3GPP TR 28.861	Study on the Self-Organizing Networks (SON) for 5G networks https://www.3gpp.org/DynaReport/28861.htm
	3GPP TS28.531	Management and orchestration; Provisioning https://www.3gpp.org/DynaReport/28531.htm

1.6 Conventions

"The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 **Error! Reference source not found.** and clarified by RFC8174 **Error! Reference source not found.**, when, and only when, they appear in all capitals, as shown here.

2 High-level Architecture

2.1 High-level Reference Architecture

This PRD's primary goal is to provide technical requirements for E2E (End-to-End) network slicing to align all technologies and their specifications for the correct creation of E2E network slices. For communication service customers, it will enable comprehensive network slicing services.

Considering an E2E perspective, multiple actors such as network operators and service providers may also need to interact with each other to provide E2E communication services and to share resources. For this reason, a common way of the interworking is particularly important to be specified so as to be consistent with standard specifications in each technical domain.

In order to satisfy the above, the E2E reference architecture in this PRD (Permanent Reference Document) is based on the architecture illustrated in GSMA Whitepaper NG.127 [3]. Figure 1 shows the architecture in GSMA Whitepaper NG.127, which is a high-level architecture with three different stratums - infrastructure stratum, network and application function stratum, and O&M stratum. The detailed architecture specification of technology domains in each stratum is defined by different SDOs (Standards Development Organization) and fora, per GSMA Whitepaper NG. 127. As a general principle, it is important to fit within an established ecosystem. Therefore, existing and established standards should be reused as much as possible. If reuse is deemed to be insufficient, proposed additional features are described in Section 2.2 of this PRD.

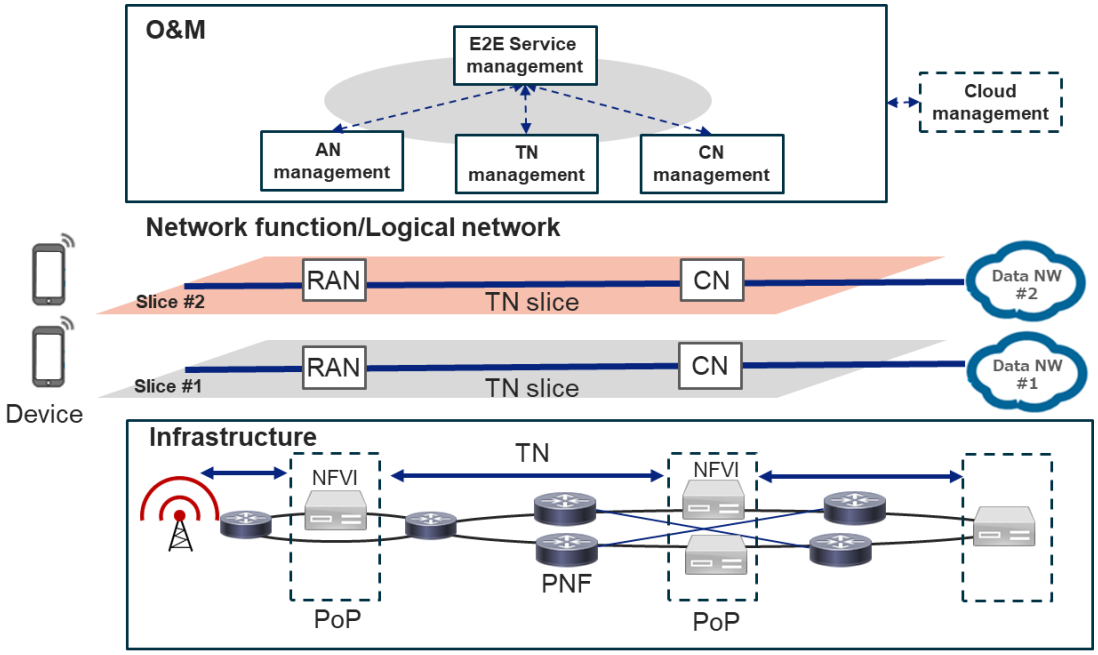

Figure 1 Overall Network Slicing Architecture example for mobile network [3]

3 Proposed Technical Requirements

3.1 General

Requirements relating to E2E network slicing are addressed in this PRD, while GSMA Whitepaper NG.127 [3] indicates technical aspects expected to be achieved by E2E network slicing to guide the requirements. In addition to those requirements addressed in this PRD, there are other network slicing related requirements for each domain specified by other SDOs and fora.

The following sub-sections state the requirements to support E2E network slicing.

3.1.1 Co-operation of application layer and transport layer

To fulfil SLA for E2E network slicing, Radio Access Network (RAN), Core Network (CN) and Transport Network (TN) have to co-operate. TN, as well as RAN, and CN should identify and differentiate a traffic for certain slice in order to fulfil SLA for E2E network slicing. Thus, all factors in E2E network slicing should co-operate to archive an SLA for E2E network slicing. RAN and CN already share common identifier, NSSAI, which is standardized at 3GPP. But RAN and CN should also share the identifier with TN.

There are two possible technical solutions to co-operate

Solution 1: Automation of an identifier(s) translation between RAN and CN and TN.

Solution 2: Depend on operator's deployment.

One of those solutions requires to archive SLA (Service Level Agreement) which is assigned to E2E network slicing.

3.1.2 SLA assurance in E2E

A customized network requires performance requirements to be a fundamental concept for network slicing. An SLA is a commitment of provisioned network services between an operator and a consumer. The consumer declares communication service(s) requirements to the operator. Network performance attributes such as throughput, latency and reliability could be part of the technical specification of an SLA. GSMA PRD NG.116 [6] specifies such attributes.

In order to assure an SLA with each consumer, the following are required.:

- To maintain high quality of E2E network services, including network and applications based on customer requests.
- To provide E2E service for B2B2X (Business to Business to everything) partners with coordinated network and applications.
- Orchestration to achieve E2E alignment in order to reflect the slice-specific requirements of network domains based on the E2E SLA.
- To reserve appropriate amounts of resources (e.g., radio rand compute resources) and to deploy network functions at appropriate places in all network domains.

3.1.3 Deterministic and operator-controlled UE behaviour

Currently, the network can provide the User Equipment (UE) with configuration of the network slices the UE can use. In addition, UE Route Selection Policies (URSP) can offer the operator the opportunity to associate applications to connectivity provided by a network slice and the Data Network Names (DNNs) that are supported in the network slice.

However, it is for UE implementation on how it decides when to register with the network slices (it may do so based on the configured network slices list, or trigger registration only at detection of the need to use a network slice). Likewise, it can deregister with a network slice or tear down a PDU (Protocol Data Unit) session that are unused at a time of its discretion. There is therefore a need for the network to improve the control it has on the UE behaviour.

Mobile Network Operators (MNOs) may want to enforce the registration to network slices only when the need to use them is detected (e.g., to minimize the number of registered users, or to enable the choice of the optimal AMF(Access and Mobility Management Function) (used for the network slices) or based on configuration (e.g., to reduce the times a UE register and deregisters with a network slice). This degree of network control is missing and also it is not possible to specify the time a UE should de-register from a network slice after the last PDU session stops using it (i.e., after the last PDU session established in the network slice by the UE is torn down).

Similarly, MNOs may want certain PDU sessions to be established at all times and some instead to be just established when usage is detected/requested by applications. This degree of control is missing, and also it is not possible to specify the time a UE should release a PDU session after the applications that trigger its establishment stop using it (i.e.,

all the applications that were bound to the connectivity are shut down or do not require a specific connectivity). Given that it is a PDU Session that make use of resources, keeping a PDU session established for longer than needed can cause unnecessary resource shortage. On the other hand, tearing a PDU session down too soon may cause more Session management signalling if this PDU session is needed again within a short lapse of time. Hence, it is important that the network takes additional measures to control the establishment and release of PDU Sessions, which in turn can also allow detection of when there is a need to register or deregister from a network slice.

Lastly, it is important to associate specific standardized categories of applications traffic with the specific connectivity defined by an operator to serve these. These standardized categories support needs to be defined.

Standardized traffic categories should be defined to accommodate connectivity requirements that are shared by different applications traffic, as well as criteria related to the geographical scope, that will be defined as part of the NEST. The following provides a (non-exhaustive) list of categories of traffic that present specific requirements:

- IMS Voice + Video comprising voice, video telephony and multimedia communications over IP networks. Voice, Video and SMS (Short Message Service) over IMS (IP Multimedia Subsystem), as well as RCS (Rich Communication Service), are included in this traffic category.
- Internet, defined for Internet data traffic with wide availability but no critical requirements on latency or data rates.
- IoT and machine to machine type of traffic, characterized by low data rates and requiring low latency occasionally or in scenarios of enterprise/private network environments with reduced geographical scope
- On demand downlink streaming, characterized as downlink high data rates content and low latency, delivered and consumed in a continuous manner from a source, with little or no intermediate storage in network elements.
- On demand uplink streaming, characterized as uplink high data rates content and low latency, send in a continuous manner from a source, with little or no intermediate storage in network elements.
- Vehicular communications, i.e., vehicle-to-infrastructure, vehicle-to-network, vehicle-to-vehicle, vehicle-to-pedestrian, vehicle-to-device, vehicle-to-grid, characterized by low latency, high reliability and high availability.
- Real time interactive traffic, characterized by bidirectional variable data rates as well as low latency requirements, for example gaming, AR/VR.
- Unified communications traffic, that comprise communications through a single service, for instance instant messaging, VoIP and video collaboration through the same user interface. Expecting consistent latency, jitter, and variable data rates.
- Background traffic, comprising processes running in the background for e.g., firmware/software updates over the air, with no critical requirements from latency or data rates perspective
- Location-based traffic, requiring highly reliable user and control plane signaling.
- Critical Communications, with low to very low latency requirements, variable data rates and high availability and prioritization.

The following points are required:

- The 3GPP system should enable the network control of the UE behavior to address the issues described above. The UE should receive from the network the intended control information.
- Standardized Traffic categories for the Traffic Descriptor in URSP should be defined in GSMA (e.g., GSMA PRD NG.113) and these should be supported by 3GPP.

3.1.4 Multi-aspect resource optimization

The resource management in current mobile networks is handled in a network dimension-specific way, i.e., there are specific procedures and parameters for optimizing resource utilization for each dimension of network resources. The network dimension in this context represents the type and scope of network resource, for example, the resource types include physical radio resources, or virtualization and cloud resources, or the different network management scopes like individual network slice instances. The different resource optimization mechanisms (e.g., as illustrated by Figure 2) have visibility of resource utilization limited to specific dimensions, e.g., only to the RAN resources or to the virtual infrastructure resources, but not to both simultaneously. However, the 3GPP management system responsible for RAN slice subnet instances will have an interface towards RAN Network Function (NF) management, but also an interface towards Network Functions Virtualization (NFV)/cloud management in case parts of the network is virtualized, and in this case example, Os-Ma-Nfvo Network Function Virtualization Orchestration) (is used. Furthermore, the means that resource management is limited to a specific dimension. In Figure 2, for example, the number labels indicate the different areas where resource related actions may be performed, specifically: changing radio parameters and policies at (1) and (2); changing slice configurations at (2) or (5); changing virtualization resource configurations by undertaking load balancing among Virtualized Network Functions (VNFs) at (3) or scaling the virtual resources at (4).

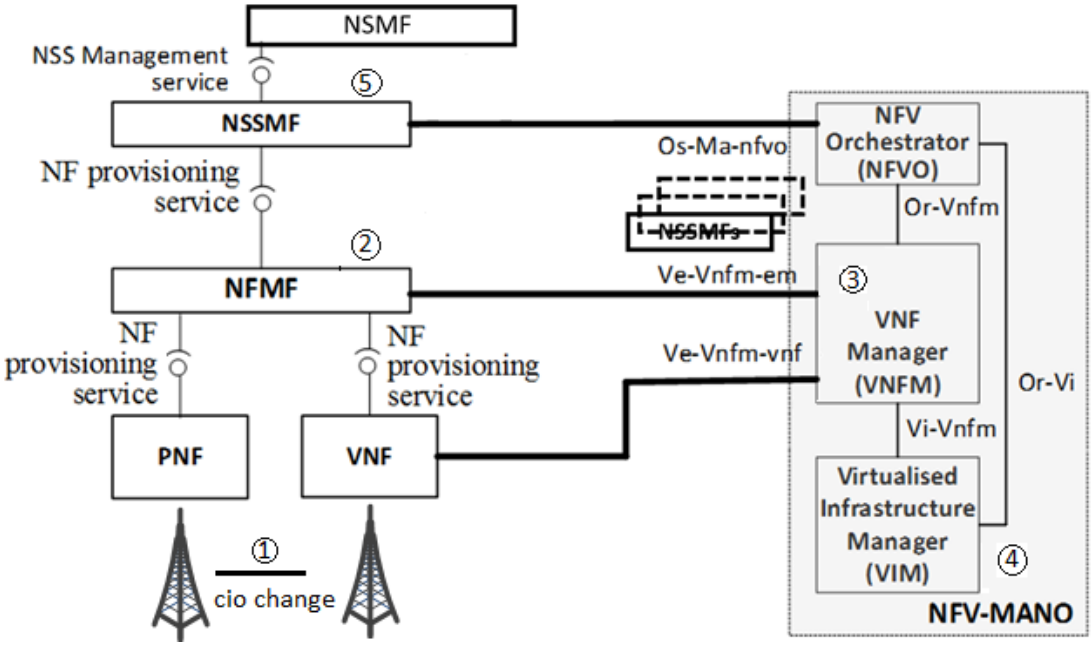

Figure 2 Resource optimization (Load balancing) options labelled 1 through 5.

The impact of dimension-specific actions to other potentially influenced dimensions and the benefits of cross-dimensional decisions are not considered. Such multi-dimensional awareness becomes critical especially with advent of virtualization support in network deployment where network functions in a network slice can be implemented on different infrastructure domains, e.g., partially on physical and partially on virtualized resources such as in case of Central Unit/ Distributed Unit (CU/DU) split of gNB. Besides, multi-dimensional awareness can be leveraged to maximize the available resources, e.g., if core network latency drops from 100ms to say 60 ms, the extra 40ms can allow the Operation Administration and Maintenance (OAM) to use RAN functions that would otherwise violate the E2E latency budget (this is already in scope for cross-domain management as part of network slice management in 3GPP). Moreover, as network slicing enables the cases of network function sharing, the same network function can belong to different administrative domains corresponding to different managed network slice/subnet instances. Therefore, it is of high importance to be able to manage such interdependencies among the different network dimensions in order to perform adequate resource management.

The following is required for the provider of Multi-aspect Resource Optimization (MARO) to jointly optimize resource utilization across multiple network dimensions. The provider of MARO needs

- to consume network resource-related provisioning, assurance and supervision services on the different network dimensions including Radio resources and virtual resources; the deployment environments including virtual machines or cloud instances, as well as the slices and subnets utilizing these functions, and
- to offer provisioning services on the resource parameters configuration.

3.1.5 Resource reservation and feasibility check

It is required to achieve alignment with an E2EO (End-to-End Orchestrator) (in order to perform network slice provisioning and modification without failure. According to 3GPP TS 28.530 [7], there are use cases of provisioning including feasibility checks, and the derived requirements for a 3GPP management system.

3.1.6 Automatic provisioning

The lifecycle phases of a network slice include service design, provisioning, deployment, operation, and removal. In order to provide customers network slicing services, a network service provider is required to map customer requirements into typical values of GSMA PRD NG.116 [6] Generic network Slice Template (GST) attributes and to deploy resources and network functions properly so as to satisfy customer’s SLAs (Service Level Agreement). As network slicing is spanning over several network domains, E2E orchestration is a key capability to provide network slice service offerings from the perspective of lifecycle management. E2E orchestration can control management functions in each domain to provide lifecycle management service, such as provisioning, and can make services available faster to the Network Slice Customer (NSC) by reducing offering time. Toward the goal of NSaaS, automatic provisioning function in an E2E manner is one of the expected

functions to bring benefits, e.g., cost advantages, by removing manual interventions in the lifecycle management. process.

The following points are required.:

- External interface to receive requests from customers/3rd party directly and to provide network services agilely to customers.
- Each automatic provisioning function in E2E manner, e.g., automatic activation, automatic slice creation, and closed loop.

3.1.7 Isolation

Isolation in network slicing is a multi-faceted problem, articulated into three separate dimensions: (i) *performance*, ensuring that Service Level Specification (SLS) is always met on each NS (Network Slice) instance, regardless of workloads or faults from other running instances; (ii) *security,* ensuring that any type of intentional attach occurring in one NS instance have no impact on any other running instance; and (iii) *management*, ensuring that each slice instance can be operated as a separate network partition, with an independent lifecycle management. These three dimensions, originally discussed in GSMA PRD NG.127 [3], impose requirements on both Network Slice Providers (NSPs) side and on Network Slice Customers (NSC) side, as elaborated below.

Performance dimension:

- The **NSC** shall be able to express the SLS for a service to be hosted in a network slice. To that end, the NSC may use a NEST together with other deployment related info (e.g., on-prem execution of some NFs, asynchronous replication of applications and data across different regions for high availability support, etc.).
- The **NSP** shall be able to segregate resources of an NS instance from other NS instances. To that end, the NSP may leverage mechanisms providing means to split the infrastructure into a set of partitioned resources (resource chunks) and allocate partitioned resources into different slices.

Note: This segregation impacts all resource domains, including radio resource domains (i.e., cell resources such as RRC (Radio Resources Connections), non-GBR (Guaranteed Bit Rate) (DRBs (Data Radio Bearer), PRBs (Physical Resource Block), etc.), compute resource domain (i.e., IaaS/PaaS resources allowing for VNF/CNF (Containerized Network Function) execution), and transport resource domain (i.e., WAN resources).

Security dimension:

- The **NSC** shall be able to express that wants its management data to be safe and separated from the rest of NSCs. This management data includes business related data (e.g., charging information, subscriber profile) and operation related data (e.g., performance assurance (PM)/fault supervision (FM)/provisioning (CM) data, log, trace, MDT, policy, analytics reports).
- The **NSP** shall provide means to guarantee the management data is securely stored, and accessible only for authorized NSC. Depending on the criticality of the data, the

NSP may decide on different security management solutions for data protection, including integrity, confidentiality and privacy protection solutions, even for the same NSC. This means that the NSP shall provide each NS instance with appropriate mechanisms preventing unauthorized entities to have read and write access to slice-specific management data and be able to record any of these attempts. Examples of these mechanisms can be found in the survey reported in [8].

Management dimension:

- The **NSC** shall be able to perceive the NS slice as a self-contained, dedicated network.

- The **NSP** shall provide means to allow for multi-tenancy support (controllability separation in the network), based on the definition of separate yet tailored management spaces for different NSCs. Each management space shall be provisioned with only the configuration and monitoring capabilities that the particular NSC needs to consume from their NS instance(s). The activation / de-activation of certain capabilities allows NSP to regulate how much control the NSC can take over allocated NS instance(s).

- The **NSP** shall have the ability to allow for the on-demand and controlled exposure of capabilities to the NSC, according to the settings of associated management space.

Note: The importance of capability exposure in slicing environments has been already highlighted in [9][10]. The NSP can leverage this feature to offer NSaaS in multiple forms, from *provider-managed slices* (i.e., the provider is in charge of the slice operation, while the customer can merely use the network resources of the provider slice, without any further capability of managing or controlling it) to *tenant-managed slices* (i.e., the customer takes full control of the slice, and the provider just segregates the infrastructure as required for that purpose), with some variants in between. By regulating the exposure, the NSP can define the visibility and the degree of control the NSC can take over the slice. **Error! Reference source not found.**3 shows the logic behind the capability exposure concept.

- The **NSP** shall be able to make capabilities available for consumption through Application Programming Interfaces (APIs) adhered to three main principles:
    - Open. Offered APIs need to leverage as much as possible on standard-based or de-facto APIs, following industry recommendations.
    - Global. Offered APIs should allow every NSC to have a uniform and consistent service experience across global footprint, with the effortless portability of application across different operator platforms (relevant in mobility/roaming scenarios) and easy service replicability (in case there exists trans-national slices).
    - User-friendly. Offered APIs need to be abstracted out of internal APIs, to hide telco complexity and make them easy to use (consume) by NSC, especially those with no telco expertise/background expertise.

- The **NSP** shall have mechanisms to manage interactions between NSC and NSP, including API registry & discovery, access control, and authorization.

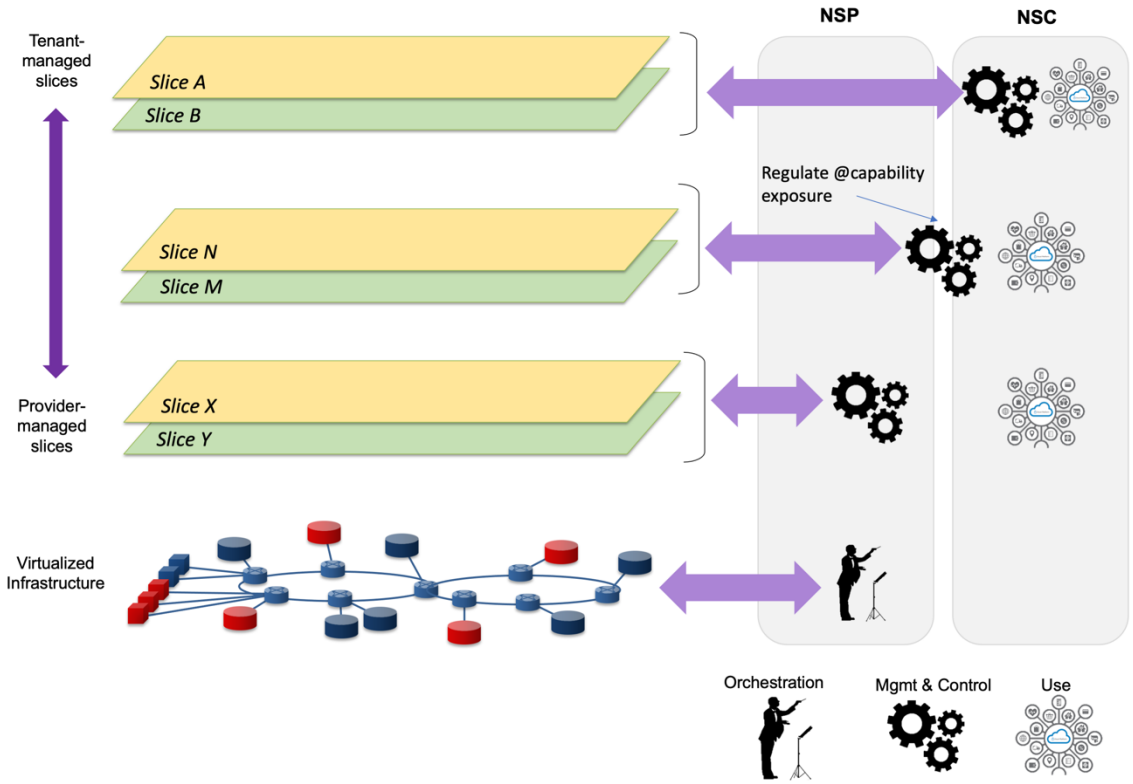

Figure 3 Capability exposure in network slicing environments.

4 Pathways for a phased rollout of NSaaS

Network slicing is an E2E solution, covering all the technology domain spanning from UE to the data network. The provisioning and operation of the different slices, and the lifecycle management of hosted services, is done with an Operating Support System (an OSS) composed of a number of management domains, including vertical management domains (e.g., RAN, CN, TN management domains) and horizontal management domains (e.g., MANO, E2E network & service management domains) [3]. However, the reality proves that the maturity level of slicing varies across all these technology and management domains. In fact, the degree of penetration of slicing features in the different technology domains is not the same. For example, while the core network has incorporated network slicing support since the first 5G release (3GPP Release 15), the transport network does not support any native slicing feature yet, and first solutions have only recently been integrated into the radio access network. The main reason why the maturity level varies across technology domains (and their corresponding management domains) is mainly due to the existing lack of coordination among SDOs in the standardization arena, with a high number of participating SDOs. In the current landscape, each SDO addresses a portion of the E2E problem, developing slicing specifications for this portion under assumptions that do not necessarily match the assumptions made by other standard bodies, which typically address other portions. A clear example of this mismatching can be observed on the priorities that different SDOs set in relation to which slicing features need to be worked out in each release. In fact, these priorities are quite different across SDOs, both in time and scope.

The above-mentioned issue, together with the inherent integration issues existing in carrier networks (e.g., multi-vendor solutions, brownfield facilities), outlines the main challenges that operators need to work out to enable the full promise of network slicing, which is NSaaS. However, solving these issues towards this ultimate goal may require years. In the meantime, operators need to look for workarounds to start commercializing and monetizing network slicing. In this regard, the only realistic option for operators is to introduce slicing capabilities progressively, starting with early stage slicing and to use a vision of what is desired in the long term to guide progress and focus; in other words, operators shall outline a phased NSaaS rollout. For this activity, it is needed to identify all solutions enabling network slicing and position them into a multi-faceted common space, by means of a **technology radar**. The technology radar allows modelling this common space, consisting of:

- **Multiple dimensions**: identify the criteria based on which solutions can be profiled. A given managed domain (RAN, CN, TN, OSS) can have one or more dimensions.
- **Multiple rings**: represent different timelines, associated to the General Availability (GA) of solutions.

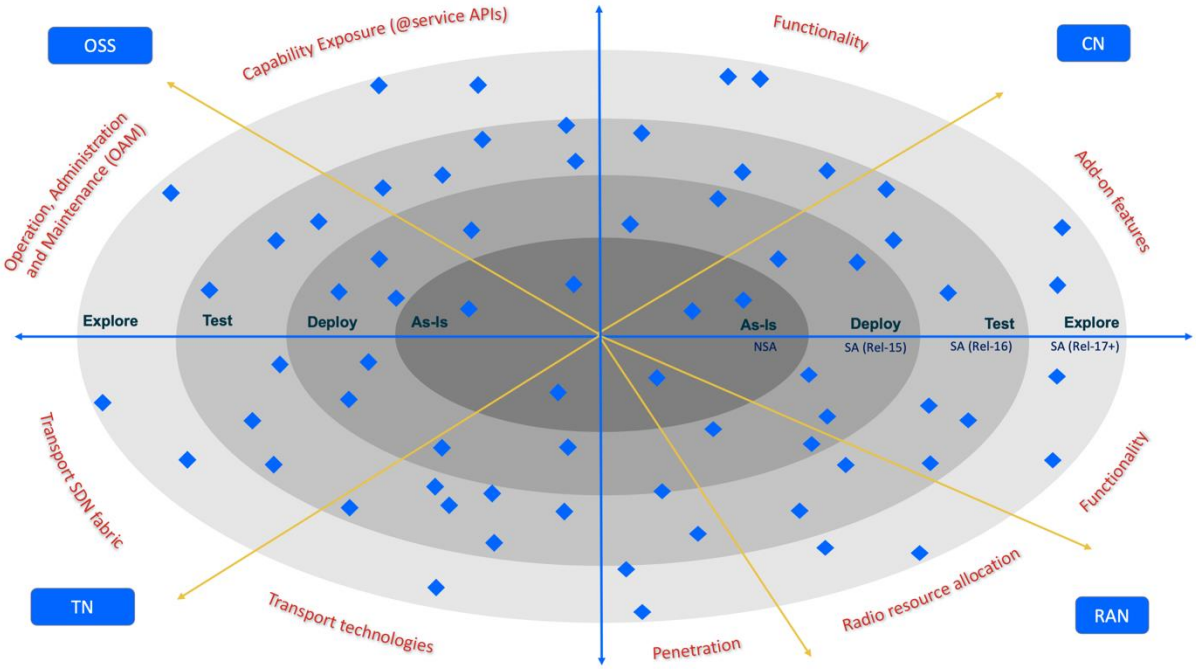

Figure 4 Technology radar toward NSaaS.

Error! Reference source not found.4 illustrates an example of how a technology radar could look like [11]. As seen, it captures solutions impacting all operator managed domains into four rings:

- As-is ring represents solutions that are available in today's carrier networks. These solutions are typically associated with technologies that operators have high confidence in, with low risk and recommended to be available across the entire service footprint. In terms of 5G roll-out strategy, this corresponds to 5G NSA (Non-Standalone).
- Deploy ring covers the slicing solutions that can be applied in early 5G SA (Standalone) networks, based on 3GPP Release 15 standards. Some operators have already started to activate their SA networks, while some others expect to get them

operationally ready within next year. With this timing in mind, we can say that this ring captures proven slicing solutions that operators may integrate in the short-term.

- Test ring: captures slicing solutions that are much more focused on satisfying requirements from uRLLC (Ultra-Reliable and Low Latency Communications) (and mIoT services. Associated with brand new Rel-16 features, these solutions have great potential but are unproven in production networks, hence it is worth operators investing in prototyping efforts in order to evaluate their performance and impact. This evaluation is typically done with commercial trials, either bilateral or multi-vendor, and different Proof of Concepts (PoCs). The upgrade towards Rel-16 is expected within the next 2–3 years; this means that test ring represents slicing solutions that might be available in the medium term.
- Explore ring includes slicing solutions that are foreseen in the long run, starting in the next 4–5 years. These solutions, tied to features from 3GPP Rel-17 onwards, promise to provide great potential, though their impact and commercial availability is still far from crystal clear. The role of the operator is to keep track of their evolution through exploratory activities such as the ones done in research and innovation projects, e.g., HEXA-X [12].

In addition, the solutions are profiled into different dimensions, as follows:

- CN domain dimensions: functionality and add-on features. The *functionality* dimension deals with the discussion on how to use CN functions for the construction of different network slices, scouting different deployment options for these slices depending on the isolation and business requirements of hosted services. On the other hand, the *add-on features* dimension refers to the set of value-added solutions that complement and extend baseline slice functionality. The network operator can optionally make use of these solutions to either (i) provision enriched services to the customer, i.e., new revenue streams; or (ii) streamline internal network operation, i.e., OPEX savings.
- RAN domain dimensions: functionality, radio resource allocation and penetration. The *functionality* dimension provides a deep dive on the applicability of open RAN principles on NR protocol stack functions to design and configure RAN slices, going from monolithic solutions towards more flexible, service-tailored composition patterns. The *radio resource allocation* dimension discusses the availability of solutions to segregate and dispatch cell radio resources to competing RAN slices, so that their targeted Key Performance Indicators (KPIs) are met. Finally, the *penetration* dimension refers to the penetration of RAN slicing technology within the operator's footprint.
- TN domain dimensions: transport technologies, and transport Software Defined Networking (SDN) fabric. The *transport technologies* dimension captures the protocol encapsulation and data plane solutions across the different network segments, allowing for the realization of soft and hard slicing, with some solutions in between. On the other hand, the *transport SDN fabric* dimension refers to the control and management plane aspects, discussing the application of programmability and automation through SDN.
- OSS domain dimensions: OAM, and capability exposure. On the one hand, the *OAM* dimension refers to the set of activities related to network slice life cycle management (model-driven fulfilment and data-driven assurance operations). On the other hand, the *capability exposure* dimension touches on the need for NSPs to make network slice capabilities available for consumption to NSCs, through open, global and user-friendly service APIs.

With this baseline template for the technology radar, the operator is in position to define their own roll-out strategy for NSaaS. To that end, the following steps are needed.

First, the operator should fill the radar out with network slicing solutions. The position of each solution in the radar can be done according to three different criteria: (i) the technology maturity of the solution, which is related to the readiness of the corresponding standards; (ii) the roadmap of commercial products, which specifies when the features associated with the solution will be available; and (iii) the relevance for the customers, which determines the prioritization of the solution over others.

Upon completion of the first task, the operator should then define its own go-to-market strategy [13], based on deciding which solutions will be made available, when, for which customers, and under which business models.

5 Conclusion

This document provides, network slicing requirements for operators from the E2E perspective are provided.

In order to realize E2E network slicing, it is essential factor to harmonize not only all technologies and specifications but also standardization progress among several SDOs and fora. It is expected that the requirements described in this document will be taken into account, as a guideline for developing technical specifications by each SDO and forum. Annex A of this document addresses specifically gaps between existing specifications and the network slicing requirements in this PRD.

Annex A Collaboration with external organizations

A.1 Gap Analysis

Technical requirements are addressed in Section 3 of this PRD. In order to achieve the requirements, identification of gaps between them and existing specification in the external organization is of importance. The following sub-sections state the observed gaps for outreach between GSMA and the external organizations.

A.1.1 Co-operation of application layer and transport layer

Regarding transport network, IETF is on the way to specify the IETF Network Slice framework in their Internet-Draft “draft-ietf-teas-ietf-network-slices” and "draft-contreras-teas-3gpp-ietf-slice-mapping". As mentioned above, harmonization with 3GPP domain is an important factor. GSMA needs to encourage IETF to work on this standardization with harmonization perspective.

A.1.2 SLA assurance in E2E

3GPP has documented management system in 3GPP domain such as architecture framework and management services in TS 28.533 [14] as well as the 5G network resource model (NRM) in TS 28.541 [15]. For SLA assurance in E2E, it is required to clarify and align specifications between 3GPP and other organizations (i.e., IETF and O-RAN).

Regarding transport network, IETF is on the way to specify the IETF Network Slice framework in their Internet-Draft “draft-ietf-teas-ietf-network-slices” and the transport controller Northbound Interface (NBI) of Network Slice Controller in their several Internet-Drafts. As mentioned above, harmonization with 3GPP domain is an important factor. GSMA needs to encourage IETF to work on this standardization with harmonization perspective.

A.1.3 Deterministic and operator-controlled UE behavior

Identified gaps regarding the requirements on deterministic and operator-controlled UE behaviour are as the followings:

- 3GPP specifications need to be updated to improve the UE behavior control to address the issues identified in 3.1.3 (including Standardized Traffic Categories support)
- GSMA lacks the identification and definition of standardized traffic categories.

The following text indicates steps for the requirements on this.

- when the UE is registered with the network, it receives the necessary control information by HPLMN and if applicable and allowed by the VPLMN. 3GPP to define the best suitable technical option.

- (Example solution should for Standardized Traffic Categories support) the URSPs TD definition is augmented in 3GPP TS 24.526 [16] as a "Connection Capability" codepoint is associated with each Standardized Traffic Category. Applications that require specific connectivity defined for a specific Standardized Traffic Category request connectivity from the UE by indicating the related required Connection Capability.

Editor's note: 3GPP could be tasked to find the best technical option to support standardized traffic categories once GSMA identifies them.

A.1.4 Multi-aspect resource optimization

Identified gaps regarding the requirements on MARO are as the followings:

- 3GPP has documented the use case description as summarized in 3GPP Release 16 TR 28.861 [17]; 3GPP also has input data for the different 3GPP domains, but details like resource parameters used for configuration are not described for this use case.
- 3GPP has not agreed to include this use case into normative specification. Hence, lacking is a normative specification of the management services and procedures related to the provision of MARO. Note that the specific solution requirements are not in scope of GSMA to specify, but GSMA should identify the requirement to optimize simultaneously across more than one dimension at a time if there are areas missing so far and encourage 3GPP to work towards solutions address this problem space.

Also, the input that would come from non-3GPP domains may need to be evaluated, possibly the candidate inputs expanded.

The following text indicates example steps for the requirements on this.

The following is an example of possible steps; this should not be seen as a solution description. The provider of MARO takes network status inputs from multiple domains and derives the policies for network configuration and optimization. This may be specific to a single network slice, or it may be applicable to multiple network slices.

The provider of MARO consumes PM, FM and CM services across different domains. The information retrieved may consists of network resource-related data (both physical and virtual resources), which may be characterized as, but not restricted to:

- The information from one or multiple network slices, slice subnets and/or infrastructure domains (physical and virtual)
- For each domain/subnet/slice, information on allocated resources and their utilization, event history (e.g., recent overload situations), history of actions taken (e.g., MLB decisions, scaling of resources)

The provider of MARO then:

- Analyses the collected data for resource optimization across different infrastructure domains for either a single network slice or multiple network slices
- Correlates the data across network slices or slice subnets. Note that although from network planning some initial mapping on resource allocation across domains is known, correlation of performance in the different domains is still necessary. For example, even given CU/DU split with known resource allocation in each domain the relative performance in each domain (CU-virtual resource utilization, vs. DU physical load) must be correlated.
- Calculates/generates network resource provisioning policy improvements. Non-exhaustive list of examples of policy improvements may be:
    - Adjusting the policy on scaling of virtual resources based on the observation that RAN MLB (on physical resources) was performing well but the virtual resources were not adequately provisioned.
    - Change the characteristics of transport links based on observed transport network performance bottleneck.
    - Change in the network slice or subnet resource provisioning for inter-slice resource optimization

Exposes obtained policy improvements as a service toward other network functions.

The policy improvement service exposed can be either:

- Directly applied by the provider of MARO by utilizing the configuration management service of the related network function
- Consumed by the provider of domain-specific control which:
    - takes into account all the existing policy improvement requests related to that particular domain/resource
    - analyses potential conflicts among policy improvements
    - resolves identified conflicts among policy improvements and derives the actual policy improvement to be applied within that specific domain
    - applies the resulting policy improvement through the configuration management of involved network function

A.1.5 Resource reservation and feasibility check

3GPP is in progress to document the use case and the procedure of reservation and checking feasibility in 3GPP Release 17 TS 28.531 [18]. In TS 28.531, Network Slice Management Service Provider (NSMS_Provider) sends reservation requests to Network Slice Subnet Management Service Providers (NSSMS_Providers), not feasibility check. The NSMS_Provider may send feasibility check request with reservation requests. However, from operator perspective, the feasibility check operation should be performed before reservation operation and be defined separately from it.

A.1.6 Automatic provisioning

No specific gaps are identified in this version.

A.1.7 Isolation

No specific gaps are identified in this version.

Annex B Document Management

B.1 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0		New PRD first version	NG, TG	Masaharu Hattori, KDDI

B.2 Other Information

Type	Description
Document Owner	NG-ENSWI
Editor / Company	Masaharu Hattori KDDI

It is our intention to provide a quality product for your use. If you find any errors or omissions, please contact us with your comments. You may notify us at prd@gsma.com

Your comments or suggestions & questions are always welcome.