ETSI GR NGP 011 V1.1.1 (2018-09)

Next Generation Protocols (NGP); E2E Network Slicing Reference Framework and Information Model

Disclaimer

The present document has been produced and approved by the Next Generation Protocols (NGP) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
DGR/NGP-0011

Keywords
network, virtualisation

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Next Generation Protocols (NGP).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document describes an information specification of resources used by services in network slices to provide true resource-assured multi-tenancy across multiple administrative and technology domains. It does not, cover the data plane or hardware aspects of traffic associated with a slice, nor does it alter the core control plane functionality of physical network infrastructure and domains. Any specific language to describe a network slice is out of scope as well.

As such, the topic of network slices encompasses the combination of virtualization, cloud centric, NFV and SDN technologies the primary gap identified is a lack of normalized resource information flow over a plurality of provider administration planes (or domains). Resource requirement of a given network slice can be satisfied in different networks using different technologies; the goal of the present document is to provide a simple manageable and operable network through a common interface while hiding infrastructure complexities. The present document defines how several of those technologies may be used in coordination to offer description and monitoring of services in a network slice.

Please note that the scope does not try to formally define a network slice, instead it relies on background material for the purpose.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] NGMN Alliance: "5G White Paper V1.0".

NOTE: Available at https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf.

[i.2] NGMN Alliance (V1.0): "Description of Network Slicing Concept".

NOTE: Available at https://www.ngmn.org/fileadmin/user_upload/160113_Network_Slicing_v1_0.pdf.

[i.3] IETF RFC 5440: "Path Computation Element (PCE) Communication Protocol (PCEP)".

[i.4] ETSI GS NGP 001: "Next Generation Protocol (NGP); Scenario Definitions".

[i.5] IETF RFC 7665: "Service Function Chaining (SFC) Architecture".

[i.6] IANA: "Path Computation Element Protocol (PCEP) Numbers".

NOTE: Available at https://www.iana.org/assignments/pcep/pcep.xhtml.

[i.7] IETF RFC 2998: "A Framework for Integrated Services Operation over Diffserv Networks".

[i.8] IETF Traffic Engineering Architecture and Signaling (teas) Working Group.

NOTE: Available at https://datatracker.ietf.org/wg/teas/.

[i.9] NGMN White Paper on Security for Network Slicing.

NOTE: Available at https://www.ngmn.org/fileadmin/user_upload/160429_NGMN_5G_Security_Network_Slicing_v1_0.pdf.

[i.10] Recommendation ITU-T Y.3110/3111: "IMT-2020 network management and orchestration requirements & framework".

[i.11] Recommendation ITU-T Y.3112: "Framework for the support of Multiple Network Slicing".

[i.12] Recommendation ITU-T Y.3150: "High-level technical characteristics of network softwarization for IMT-2020".

[i.13] ETSI TS 123 502: "5G; Procedures for the 5G System (3GPP TS 23.502)".

[i.14] 3GPP TR 28.801: "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects: Telecom Management (SA5) - Study on management and orchestration of network slicing/Network slice management".

[i.15] 3GPP TS 28.531: "Provisioning of network slicing for 5G networks and services: Detailed specification of network slice provisioning/Network slice management".

[i.16] 3GPP TS 28.541: "Management and orchestration of networks and network slicing; NR and NG-RAN Network Resource Model (NRM); Stage 2 and stage 3".

[i.17] IETF draft-netslices-usecases-02: "Network Slicing Use Cases: Network Customization and Differentiated Services".

NOTE: Available at https://datatracker.ietf.org/doc/draft-netslices-usecases.

[i.18] draft-ietf-spring-segment-routing-14: "Segment Routing Architecture".

NOTE: Available at https://tools.ietf.org/html/draft-ietf-spring-segment-routing-14.

[i.19] IETF: "Deterministic Networking (detnet)".

NOTE: Available at https://datatracker.ietf.org/wg/detnet/.

[i.20] draft-ietf-ippm-ioam-data-02: "Data Fields for In-situ OAM".

NOTE: Available at https://tools.ietf.org/html/draft-ietf-ippm-ioam-data-02.

[i.21] BBF SD-406: "End-to-End Network Slicing".

NOTE: Available at https://wiki.broadband-forum.org/pages/viewpage.action?spaceKey=BBF&title=SD-406+End-to-End+Network+Slicing.

[i.22] Generic Network Slice Template Version 0.1.

NOTE: Available at https://infocentre2.gsma.com/gp/pr/FNW/NEST/WorkingDocuments/GST%20document%20baselines/GST_Baseline_v0.8_20180712_clean.docx

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

Network Slice (NS): network slice is a description of a service aware logical network that is composed of different physical or virtual network elements, resources and functions

Network Slice Agent (NSA): entity that has complete view and control of its network infrastructure

NOTE: An agent can be a logical component of controller that performs special functions relating to network slices and exports them to network slice provider.

Network Slice Instance (NSI): instance of a type of network slice that has resources allocated to it from underlying network infrastructure and is independently managed and monitored by the tenant

Network Slice Provider (NSP): entity that provides access to network slice instance and resources associated with it

NOTE: Network slice providers coordinate and aggregate network resources from multi-domain, multi-technology networks.

Network Slice Subnet (NSS): subnet represents single or multiple networks under the control of an agent

NOTE: A complete network slice is inter-connection of subnets.

Network Slice Service Profile (NSSP): structure high-level format in which a network slice is described

slice: simplified text to represent 'network slice' in the context of the present document only

tenant: entity that consumes a network slice instance from network slice providers

NOTE: Such tenants do not care about implementation and technology details of the physical networks.

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

5GC	5G Core
API	Application Programming Interface
BBF	BroadBand Forum
DB	DataBase
Diffserv	Differentiated services
EVC	Ethernet Virtual Circuit
FIB	Forwarding Information Base
GSM	Global System for Mobile
GSMA	GSM Alliance
GST	Generic Slice Template
IETF	Internet Engineering Task Force
Intserv	Integrated services
IP	Internet Protocol
IPPM	IP Performance Measurement
MANO	MANagement and Orchestration (of NFV framework)
MPLS	Multi-Protocol Layer Switching
NEST	NEtwork Slice Template
NS	Network Slice
NFV	Network Function Virtualization
NG	Next Generation
NGMN	Next Generation Mobile Networks
NGNS	Next Generation Network Slice
NS	Network Slice
NSA	Network Slice Agent
NSI	Network Slice Instance
NSP	Network Slice Provider
NSS	Network Slice Subnet
NSSP	Network Slice Service Profile
OAM	Operations, Administration and Maintenance
PCE	Path Computation Element
PCEP	Path Computation Element Protocol
QoS	Quality Of Service
RAN	Radio Access Network
RSVP	Resource Reservation Protocol
SDN	Software Defined Networking
SDO	Standards Developing Organization
SFC	Service Function Chaining
SLA	Service Level Agreement
TCAM	Ternary Content-Addressable Memory

TE	Traffic Engineering
TEAS	Traffic Engineering Architecture and Signaling
TED	Traffic Engineering Database
UCL	University College Of London
VDI	Virtual Device Interface
VM	Virtual Machine
VPN	Virtual Private Network
VXLAN	Virtual eXtensible Local Area Networks
WG	Work Group

4 Network Slicing Architecture

4.1 Overview

Network slicing concept allows support of logical networks that are tailored for a specific service or set of services over a shared common network infrastructure for the purpose of efficient utilization of network resources. NGMN white paper [i.1] states that "the intention of a 5G slice is to provide only the traffic treatment that is necessary for the use case and avoid all other unnecessary functionality". In this regard network slice is a framework aimed at providing flexible on-boarding of newer verticals as a consequence of higher definition broadband, machine to machine communication, industrial automation, advanced emergency services and so on [i.17]. These verticals cannot be served cost-effectively by traditional network architectures because of the diversity of requirements. Network slicing techniques abstract several infrastructures and provide a communication framework for verticals to build their own services.

There are several aspects that need to be resolved in terms of efficient resource scheduling, reservation and placement mechanisms at the lower layers. Many of those aspects are either hardware or particular technology related. However, a technology independent generalized reference framework for network slicing is very much needed to demonstrate how information flows for the purpose of alignment of applications and data over dissimilar communication infrastructures.

The Next-gen network slicing (NGNS) framework defined here is a generalized architecture that would allow different network service providers to coordinate and concurrently operate different services as active network slices.

4.2 Informative Background

Network slicing is an end to end paradigm initially discussed in the context of 5G to support new kind of applications that need absolute resource guarantees in terms of latencies, bandwidth, jitter, reliability and privacy. The goal is an ability to use common end to end infrastructure that is capable of delivering diverse services with their corresponding assurance. Network slicing will be expensive, due to its inherently stringent resource assurance demands. Therefore, network slices will be used to implement specific vertical markets and it does not imply to provide QoS to individual streams.

Network slicing is a multi-technology solution that spans across multiple planes. There are several SDO activities that focus on different aspects of the network slicing.

Since SDN and NFV are considered enabling techniques for network slicing, ETSI MANO, NFV ISGs activities are concerned with the orchestration perspective that involves transforming a service using NFV infrastructure.

Many other standard activities at IETF are involved in distribution of services using SFC [i.5], segment routing [i.18] or VPN mechanisms. Some additional efforts such as deterministic networks [i.19] provide data plane centric lower level functionality to meet service assurances of bounded latency and bandwidth requirements.

3GPP network slice management [i.13], [i.14], [i.15] and [i.16] together provide provisioning and resource management of RAN and 5GC slices. BBF has recently begun work on study of network slicing in the context of BBF architecture (SD-406) [i.21]. ITU-T SG13 and SG15 have published several documents [i.10], [i.11] and [i.12] on the topic motivated by IMT-2020 initiative and bulk of the effort is aligned with 3GPP related work in 5G domain.

GSMA NEST is an internal taskforce set to define a common language in the form of GST through which all operators can describe parameters of a given slice type. GST specification [i.22] complements 3GPP slice/service type work by providing its characteristics. The GST defines attributes applicable to slices in a 5G networks such as maximum packet size, terminal density, uplink/downlink bandwidth, reliability and so on. The work will help different operators describe a particular kind of slice in a standardized manner but it still requires a framework for propagation and realization of these attributes which is not part of GST work.

In contrast, proposed framework in NGP is independent of any underlying assumption about the enabling architecture, protocol or methodology. It is also technology independent and concerns with both management and control aspects of network slices. The gap identified here is to look at holistic solution of implementing network slices so that parts of the networks enabling slices in parts can use common information aspects. It is our expectation that architectures from different SDOs can be reduced to details in the present document.

The section 8.7.4 in the NG Scenario definitions [i.4] discusses NG slicing aspects and provides a conceptual view of network slicing coordinators and agents. In the following clauses these components will be described in greater detail. The NGNS information model puts together major components, managed information objects and interfaces that provide clarity about end to end network slicing functionality. It does not apply directly to low-level control and data plane functionality and provides coordination guidelines at the network level.

4.3 High level description

A Network slice is a description of a service aware logical network composed of different physical or virtual network elements, resources and functions. A network slice is an independently managed instance of a logical network; It shares underlying infrastructure with other independently managed instances. Since the infrastructure itself comprises of different interconnected domains, essentially a slice can be seen as concatenated network of subnetworks belonging to different network domains.

The NS methods are aimed at providing custom design of networks suitable for a specific use case (vertical market). Such methods need to be able to translate a service requirement into normalized description of resources across different type of network domains based on NGMN's description of network slicing, 3-layer approach [i.2] and is reproduced below in Figure 1.

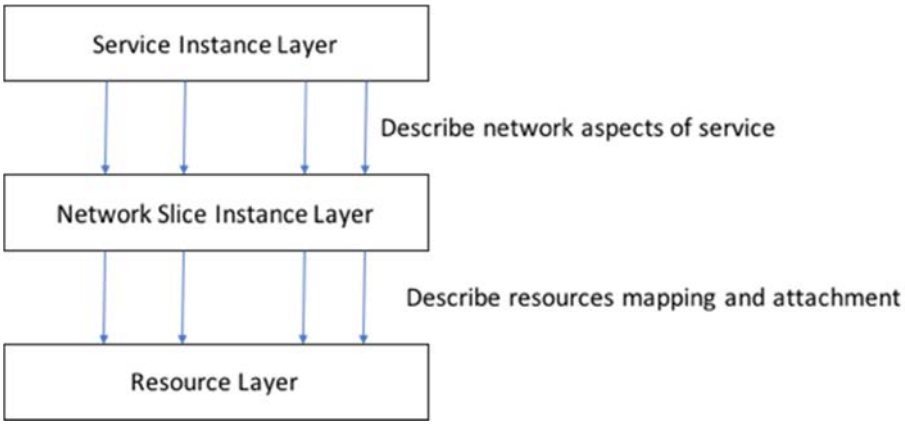

Figure 1: NGMN 3-Layer Network Slice and Service Concept

Network slice supports multi-tenancy for new set of services described in [i.1]. The focus is on the use cases that do not necessarily fit into traditional virtual networking or VPN solutions. They require much higher degree of resource assurance as well as stricter guarantees of those resource availabilities. For example, low-latency communications for V2X, high-throughput for immersive multimedia applications, extremely reliable network for emergency response situations. There are several differentiating aspects among these use cases from traditional isolation techniques, such as:

a) once allocated, the resource may be under the control of the network slice service operator (or tenant) for autonomous control of the resources,

b) absolute guarantees should be met with, even under active contention of resources in other best-effort flows,

c) every flow (per stream QoS) should receive the assured treatment, i.e. two flows within the same slice should not compete with each other.

Aligning with the NGMN's network slicing concept, there are three key areas of consideration for the NS architecture.

1) **Service description** (corresponds to service instance layer): A sketch of services instantiated, independent of any technology or underlying control plane.

2) **Network slice to abstract resource mapping** (corresponds to network slice instance layer).

3) **Resource allocation** (across different networks).

The present document describes a top-down structure and creates an information model corresponding to generalized service aware NS. Through this model, tenant of a slice is able to express service constraints and requirements. The primary contribution of present document is to identify the actors in NS architecture, the data necessary to be exchanged between different actors, how they use it and the methods associated with the information.

4.4 Network slicing design principles

4.4.1 Service Oriented Approach

In order to provide scalability and flexibility in basic network slice architecture follows these principles:

- Abstraction
- Loose coupling
- Reusability
- Autonomy

The subsequent clauses will demonstrate how these are relevant in the network slice architecture.

4.4.2 Network slice abstraction

4.4.2.1 Motivation

Abstractions hide details of underlying implementation and technologies. It is important across heterogeneous (different technologies) access and transit networks. Network slices offer a network through which a consumer can fulfil its service delivery objectives. It should be agnostic of whether a particular technology, topology or routing protocol are used. The slice specification should be well-defined and not adheres to a specific underlying solution.

4.4.2.2 Service lifecycle abstraction

This form of abstraction is dependent upon how much of the service logic is exposed as its capabilities. Each network slice has well-known create, modify, get and delete methods associated for network aspects of the service. The users or subscribers of a slice or internal data are hidden information, and a slice need not expose them.

4.4.2.3 Technology information abstraction

Any information about the underlying technology used within the service would result exposing extra information. It might result in design of a service favouring a particular technology. This takes away, flexibility and reuse aspect of a network slice. Network operations and resources relating to networks can be standardized in logical or abstract form so that they remain independent of whether underlying transport is optical or packet based, MPLS or IP, or whether the topology is L2 or L3. There will still be need for interconnections of two networks with different technologies.

4.4.2.4 Quality abstraction

Quality abstraction relates to the details provided within the service's accompanying service level agreement (SLA). Network slices should only concern with network resource information and only with the quality parameters that impact the communication aspect of the service. Separation of network and service logic policies is also necessary.

4.4.3 Loose coupling

It should be possible for network slices to be added and removed and altered flexibly across multiple administrative domains. A loosely coupled network slice ensures that changes made within one network slice domain has no adverse effects or unanticipated changes within other network domains and other network slice instances for that matter. Interconnections between domains should be clear to help isolate problems in a network slice instance.

4.4.4 Network slice reusability

Many services have similar set of network requirements. The ability to compose a slice and describe its operations in a reusable manner reduces system complexity.

4.4.5 Slice autonomy

A network slice should have control over its own runtime environment. Each slice being an independent entity should not impact the operations or lead service disruptions for other slices. Different type of virtual networks such as VxLAN, VLANs implement autonomy partially through isolation. This implies packets forwarded or identified in a virtual network are visible to other instances. In addition, it should be possible for a tenant operator of a slice to control and manage resources as well as allocate them to different users or flows with in its own network slice.

5 Information Model

5.1 Reference Component Architecture

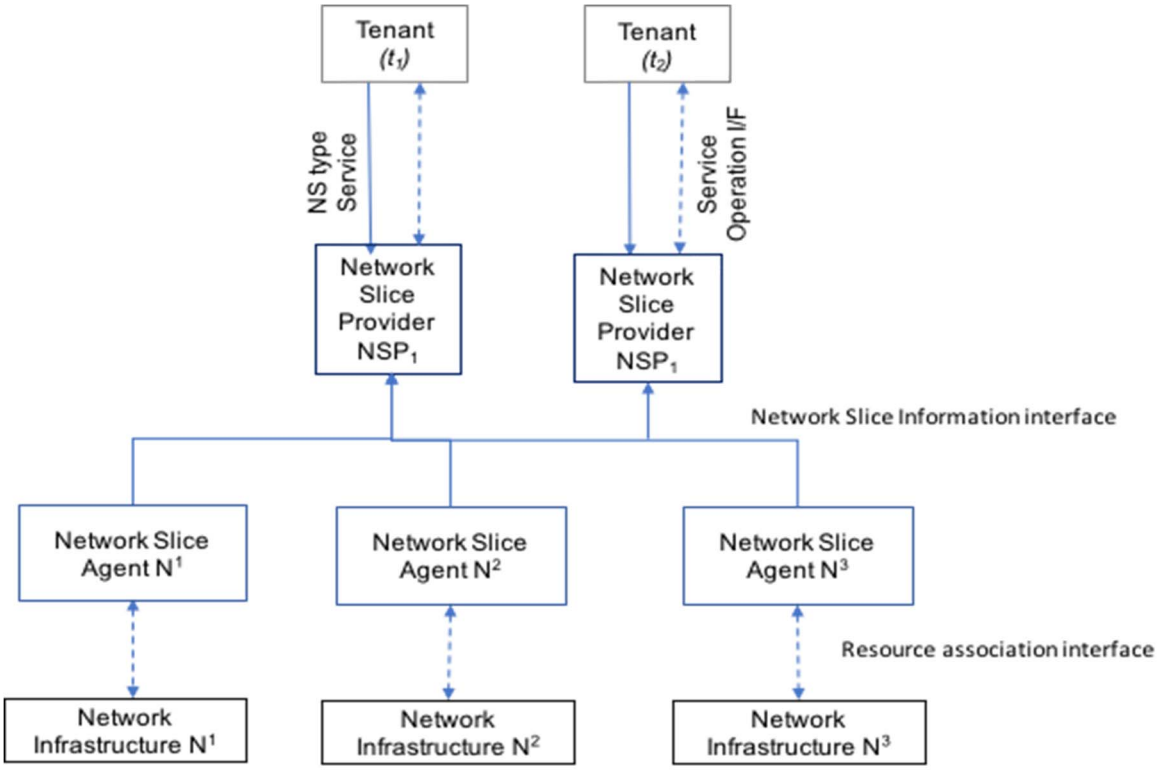

Figure 2: NS Reference Framework

The complete architecture with all the components (or actors) is described in Figure 2. There are three entities interacting for the use of a network slice namely:

1) tenant;

2) Network Slice Provider (NSP); and

3) Network Slice Agent (NSA), associated with network infrastructure provider.

A tenant is a user of a network slice. Tenant creates a service with a particular network slice type. A network slice type distinguishes the kind of network resources need to fulfil service requirements. An NS type is used as a guidance in preparation of computing resources for a slice.

An NSP provides network slice as a service to tenant for control and operations of resources in the service. There may be multiple NSPs (for instance virtual network operators) that may operate and manage its slices and tenants independently.

An NSA is a network slice entity in infrastructure provider's domain. It understands processes and maps NSP's information with in its domain. NSA is capable of extracting topology and operational state of its own network and thereby coordinating with NSP to maintain its own portion of the network slice.

To explain the working of the framework, three most important aspects are:

- **Network slice managed objects:** These are containers of information that should be shared between different components. The description of objects in terms their relationship, scope and role help clearly define the operation of network slices. The state of network slice can be extracted through these objects.
- **Network slice Interfaces:** These are the communications path over which information is collected and distributed. Each interface is associated with well-defined functions.
- **Network slice related functions:** These are the set of functions in the framework and help define complete workings of network slices.

5.2 Network service resource concept

5.2.1 Types of resources

The resources in the networks required by a network slice can be generalized to be of two types:

- Link resources
- Node resources

5.2.2 Link resources

Link resources relate to traffic or path related constraints and metrics. They comprise of bandwidth, delay, cost, packet loss, redundancy etc. It is the expectation from the network that the resources are available at the time allocated request is made, so that the congestion on the path associated with a service cannot happen if that's the requirement. Only conditions that can cause failure to meet service assurances are faults and/or topology changes. This implies the portion of link dedicated for network slice services may not contend with similar service for the life of the connection. Similarly, it is assumed that the requested latency is met with guarantee through new and advanced scheduling techniques in the data plane.

5.2.3 Node resources

Node resources can be generalized in a form of either compute or store. These resources comprise of network functions such as firewall, routers or the service-based logic node, etc. Their behaviour and the data are decided by the tenant and its interpretation is opaque to the NSA or NSP.

5.3 Network slice managed objects

5.3.1 General description

A top-level diagram Figure 3 of different managed objects and a brief description of relationship among them. Different terms are described as follows.

Network slices are logically isolated network over multi-domain, multi-technology physical networks that provide resource guarantees. **Network slice providers** offer different types of network slices as services to **tenants**. Physical networks that participate in a slice are referred to as **network slice subnets**. The network slice entities in network slice subnets is called **network slice agents**. **Resources** are described as nodes that perform specific function or links with constraints such as bandwidth, packet loss and delay variation. **Network slice service profile** is a standard definition of a particular type of network slice. It is described in terms of a **network slice service graph** that consists of **service nodes** and **service edges** that can be customized by tenants. **Network slice instance** is a runtime instance of a service profile with specific parameters and description of resources. **Network slice service context** represents the mapping of network slice service to infrastructure resources while operations are methods associated with the network slice instance.

In the following clauses important data related to these objects is described. Since this is a reference framework the details of data, its format and structure are omitted from present document.

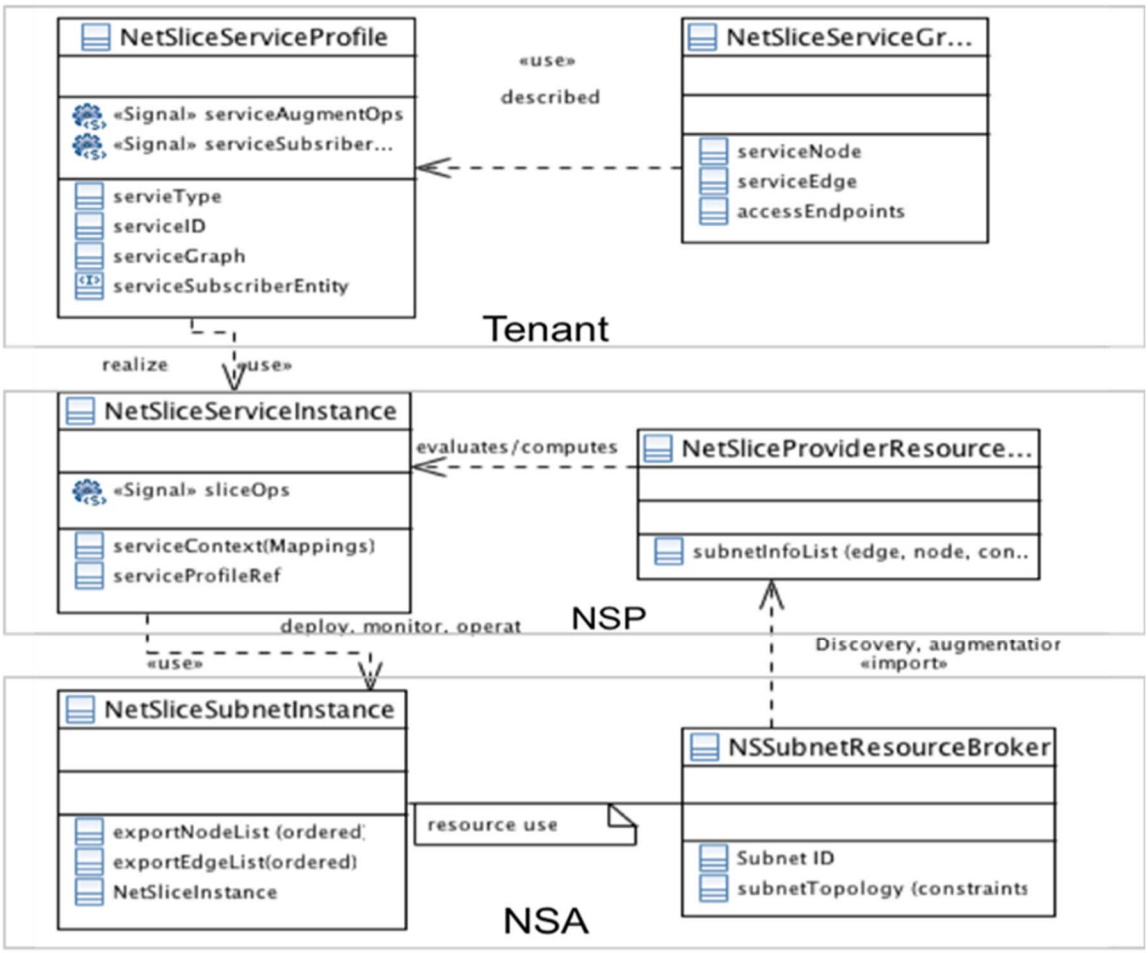

Figure 3: Managed objects in network slice provider and agents

5.3.2 Discovered objects

5.3.2.1 Network slice subnet object

A NS subnet comprises of NS path topology object. This object is a set of links that traverse through subnet with traffic parameters of latency, bandwidth, metrics associated and is exposed to the NSP. A NS subnet node can either be a compute, storage node or a network function. Parameters associated with a node includes its address, next hop, compute and function information. A NS subnet is an abstract representation of the topology with constraints in a domain.

A special kind of NS subnet node is a network function called **slice gateway** [also see i.4]. Its function is to inter-connect slice-subnets. These are exported as nodes to NSPs during discovery process with next hop information enabling connection to subnets under different NSAs. For example, NSA_1 exports a NS subnet node say, $\boldsymbol{gw(ns_1, nh_2)}$, where next hop nh_2 associates to NSA_2. The attributes of network function specify cross-domain slice encapsulation translation method (such as from/to VDI to/from MPLS label or EVC, etc.).

Table 1: Objects in NS Subnet

	Description
Subnet Identifier	Name or Identifier for the network device.
NS subnet Topology	connectivity links with different constraints to transit through subnet.
NS subnet Node list	Internal compute or network node exposed to NSP.

Aggregation of subnet objects happens in NSP aggregated resource database. Without any knowledge of internal topologies, for a given slice requirement, NSP resource database can compute specific path in each subnet. Other than NS subnets, there can simply be interconnection links (point to point logical or actual link) in NSP resource database.

5.3.2.2 NSP aggregated resource database

This is a global and dynamic view of the resources available from which network slice instances are created. Resource database interfaces with both internal network slice instance object and external interface with NSA.

5.3.3 Provisioned objects

5.3.3.1 Ns service profile object

NS service profile database build from NS service profile (Table 1). A NS service object has an identifier and a type as to distinguish between types. These 'NS types' are seed type in NS infrastructure and can be extended for customizing by tenant upon its needs. NS service object is described in terms of a service graph (Table 3). A graph has service nodes to perform specific functions, service edges and service access endpoints.

Table 2: NS Service Profile

	Description
NS Service Profile Identifier	Name or Identifier for the network device.
NS Service type	Type of service well known to NSP.
NS Service Graph	connectivity graph of nodes of interests.
NS Service subscribing entity	Subscription to service is done in a particular format. These entities include subscribers using a service.

Table 3: NS Service Graph

	Description
NS Node	Type of node and is same as those explained in clause 5.2.3.
NS Edge	Type of constraints.
NS Service access end points	Service end points describe the entry and termination points of the services. The endpoints are referred to as service headend and service tail end.

Network slices services are instantiated upon tenant-requests and monitored periodically and/or on demand. A model of minimal management is desired to reduce communication overheads. This may be achieved by using high level operations and offloading performance measurements in the context of service assurance. The operations are delegated to underlying NS agents, which then monitor and notify exceptions to NSP.

5.3.4 Runtime objects

5.3.4.1 NS service context object

Network slice is an instantiation of network slice service profile. Network slice instance coordinates computations of constrained path with the resource database and secures mapping for service graph of the service profile. This is stored in runtime NS service context. From a service graph perspective, a service node may be in a particular subnet and can connect across several subnets. Similarly, an edge segment is an ordered connection of one or more path identifiers and may span across different subnets. By ordering next hops of node or segments, an end to end slice is stitched in NSP. Often, it is not desired by a subnet to expose internal topology. In those cases, an abstract path identifier is provided by the subnet in that Resource database. Further details in clause 9.1.3.

Table 4: Network Slice service context

	Description
Service node segment	A service node may be represented by a single and a group of nodes in a subnet. It maps to network functions, service functions or subnet gateways.
Service edge segment	An edge segment comprises of an ordered list of path ids provided by the resource manager.

Once the path binding is done in the context, it aggregates operations from subnets at runtime for service assurance.

The bindings from network slice happen with subnets through NSAs.

Table 5: Network Slice Instance

	Description
Network slice context	The operation requested by NSP for collection of performance measurement metrics on periodic basis or when the thresholds are met.
Network slice service operations	For fault alarms and statistics and state.

5.3.4.2 NS service operations

Table 6 provides a list of operations associated with a network slice service. Operations are data-driven and cause changes to data and corresponding mappings. 'Post' operation uses NS service profile as an input parameter, the result of the operation is a network slice service instance. In the process several things take place such as path setup for the service and mapping of a network slice on the infrastructure. These processes are explained in the later clause 6.9. Withdraw operation will take a NS service identifier and have it removed from the underlying infrastructure.

Attach and detach operations are in fact runtime operations as they follow after a network slice is instantiated and allow subscribers (or users) of the network slice to be attached at the NS service access endpoints. Attach operations allow subscriber specific policies i.e. finer grained control of tenant specific information. It can be imagined as flows in the aggregated service path. For simplicity, these policies will be associated to service nodes and access endpoints. By limiting attachment with in the mapped network keeps the scope with in the tenant's domain.

Get and Modify operations are also subscribers related and are used to change tenant defined data, retrieve statistics and runtime state.

Altering the service graph should be a restrictive task as it is derived from a template of a service. However, the network slicing essentially, is a dynamic concept. Augment operation could allow certain changes to the service graphs such as:

a) elastic adjustment of capacity and cost to address scale of subscribers as they grow or shrink,

b) replication of exact same service graph and bind to same instance,

c) adding new service nodes in the graph, although this may become disruptive to existing flows.

Apart from these operations, an event interface is used for fault notifications.

Table 6: Network Slice Service Operation

	Description
Post	The operation to instantiate a network slice on infrastructure.
Withdraw	
Attach	Attachment and removal of subscriber policy. This operations associates a flow to the slice posted.
Detach	
Get	The operation to collect runtime statistics about a subscriber.
Modify	a handle to the network slice instance for editing tenant parameters of a service.
Augment	A handle to the tenant for editing constraints associated with a path or node in the service. These requests are made to the NSP resource.

5.3.4.3 NS subnet operations

There is another group of operations that take place in the context of resources as shown in Table 8. NSP Subnet operations are used for maintenance of state between NSP and NSAs as well as overall view of resources available for network slices. At the time of discovery, monitor operations are bound at NSAs for periodic pull of statistics such as delays, bandwidth consumptions. The periodic information may be received as reports from the NSAs.

Subscribe operation is used for registration of events such as outages or exceptions to resource assurance. Augment operation case may be seen to represent two separate mechanisms:

1) to request for specific resources, perhaps as new network slice service profiles are added, by NSP to determine if profile is feasible;

2) adjustment to existing reservations, in context of operations of particular network slice service (see clause 5.3.4.2).

Table 7: NSP Subnet Operations

	Description
Monitor	The operation requested by NSP for collection of performance measurement metrics on periodic basis or when the thresholds are met.
Subscribe	For the notification of fault alarms and statistics.
Augment	Allows an NSP to request adjust constraints for a given path.

5.3.5 Network slice agent objects

5.3.5.1 NS subnet resource broker

Network slice agent interacts with NSP for the control & management of resources and runtime operations of a network slice. Several NSAs are connects to one NSP; an NSA may be viewed as a controller for a particular domain represented as network slice subnet at NSP.

A subnet resource broker object uses its domain specific local control and orchestration to interface with underlying infrastructure. It gathers topology of its own network along with the constraints and exports to the NSP. The exposed topology is expected to filter internal path details and provide subnet entry and exit interfaces. The topology has links with cost, bandwidth, latency, so on constraints. The resource broker may also expose the nodes, it intends to provide as service or network functions. Where the nodes are needed to be placed in topology can be computed on demand (i.e. when the node is needed by a service).

NS subnet resource broker, also coordinates with NSP aggregated resource database (clause 5.3.2.2) for augmentation of resources. Although this is not a trivial task; while bandwidth may be aggregated to certain limit by creating new path with same constraints, adjusting for latency needs an understanding of delays of transmission, processing and propagation at each hop in the network. However, these details are confined to the subnet. The object details are similar to Table 1.

5.3.5.2 NSA service segment

The NSA service segment has description of paths and nodes a network slice service is allocating to in this subnet. It is the mapping of abstract to concrete resource of NSA to a slice instance. It creates an association in resource broker to mark resources, paths that are allocated to a slice instance and monitors them.

5.3.6 NS interfaces

In Figure 4 five interfaces are defined for communication between primary objects in network slice service model. The objects and operations happen over these interfaces are described here.

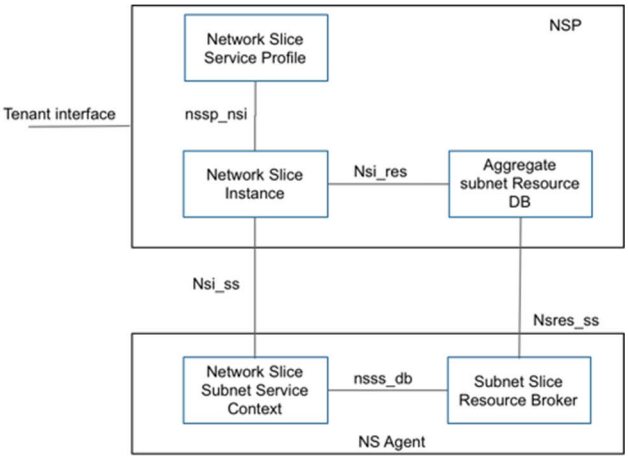

Figure 4: Network slice framework interfaces

a) Network slice service profile to network slice instance (nssp_nsi), is used between network slice service profile to network slice. All transformation from a tenant service to infrastructure resources happens over this interface. Nssp_nsi also carries information of a service. Since the services nodes are leased in the network slice subnets, containers carry opaque operations to be applied.

b) Nsi_res is an internal interface between network slice instances and the aggregated subnet resource db. Network slices can be seen as a client of resource db. A set of accessor functions can be defined to retrieve different information such as runtime state of a path in the subnet, minimal latency paths, a particular service node chain.

c) Nsss_db interface is inside Network slice agent. This is a coordination channel between the service context and subnet slice resource database.

d) Nsi_ss interface carries service instance information between slice instances and the context. All tenant related operations and other functions are relayed thorough slice instances over nsi_ss.

e) Nsres_ss is an interface for the resource coordination between agents and NSPs. The discovery function, state monitoring and topology changes are communicated through these interfaces.

6 High Level Functions

6.1 Network slice functions

The following functions are performed by NSP and NSA:

- Network slice subnet discovery function
- Network slice subnet augment function
- Network slice mapping function
- Resource computation function
- Network slice delegation function
- Report aggregation function
- Service assurance function
- Tenant operated network service function

6.2 Network slice subnet discovery function

The subnet discovery function is required for establishing central repository of resources exported by NSAs to NSP. It enables NSPs to determine constraint-based path for a specified service. A subnet is a logical network under an NSP, providing its information to NSA. Without the prior knowledge (or discovery) NSP cannot create an end to end network slice. The NSAs also subscribe to be notified about changes to resources offered. The architecture is based on repository of resources that helps it compute paths for requested service. In case the path is not available, two options are possible:

a) NSA to report failure to NSP,

b) NSA attempts to pre-empt best-effort flows and make resource available.

The NSAs are not required to export their entire network information but should expose an abstract path information, i.e. entry and exit interface and constraints.

Network slice customization benefits from granularity with which resources can be allocated. For example, when an NSA offers a 10 Gbps bandwidth resource; whether it can allow (or support) allocation at the granularity of 10 Mbps or 1 Gbps should be exported to NSP. Another example is ability to offer latency ranges between say from 50 ms to 10 ms between 2 end-nodes with a granularity of 5 ms (i.e. latency guarantees of 10,15 and so on based on internal path and queue configurations in a domain).

NSAs may use predefined network slice service templates to setup their:

a) domain topology,

b) constraints,

c) statistics and reporting and operations to export those resources accordingly.

The discovery process, builds NSP aggregated resource database. The object involved are NSA network slice subnet and NSP network slice subnet.

6.3 Network slice subnet augment function

Augmentation of resources is done to adjust (increase or decrease) budgets and carried out by NSA; The differentiating motivation for network slicing over other virtual networks comes from the support for dynamic change in service requirements which is possible if and only if underlying network infrastructure is able and/or willing to augment resources on-demand. The augmentations are not constrained to fine-grained (per-flow) or coarse-grained (aggregated), however, when performed at the level of network slice service instances provides better scalability.

Since services have paths associated, resource change can affect those paths depending upon type of resource change:

- Changing a parameter of the service, for example, latency constraint from 15 ms to 10 ms. It may require creation of new path to allocate new resource changes; then move flows from older to the newer path. However, this could cause system instability as each node in path re-/de-provisions itself. It could also lead to packet loss if flows are moved before new path setup is completed.
- Capacity change of service, for example serving 300 flows instead of 100. Often a new path with same constraints can be defined such that newer flows of that slice instance are associated with new path.

Whether the paths are represented as a single logical path to tenant or both paths are exposed is an implementation decision. This function involves interactions between NSA subnet resource db and NSP aggregated resource db.

6.4 Network slice mapping function

The NS mapping function is a mechanism for creating associations between infrastructure elements and logical elements in network slice instances. Before network slice is instantiated on NS subnets, the nodes and edges of network slice service have to mapped to available resources in the aggregated resource database. The network slice mapping function receives edge segment and node segment from clause 5.3.2.2 (aggregated resource database) as a result of computation function in clause 6.5, thus the last step of network slice instantiation is binding physical resources with the logical ones.

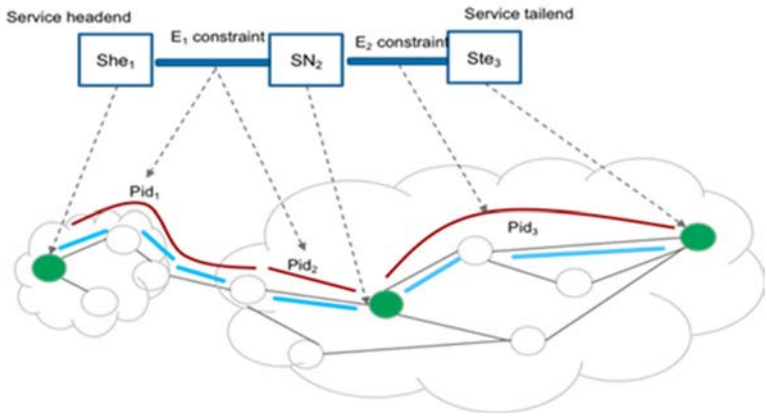

Figure 5: NS service to resource mapping

Figure 5 shows how mapping is achieved between slice subnets and network slice subnets. The resource database provides first subnet with Pid_1, Pid_2 as part of segment that maps to E_1 and Pid_3 maps service edge E_2.

6.5 Resource computation function

Determining which resources are suitable and computation of path for corresponding service graph for a network slice service is the most complex part of the network slice implementation. This operation in NSP can be simplified by certain pre-conditions:

- Each NS subnet independently computes paths based on its topology and knowledge of available resources.
- Computation function need to determine resource-paths (segments in the graph) between service nodes.

- End to end path is concatenation of paths in subnets interconnected at NS gateways.

Computation function happens both at NSP and NSA level. Since a path is independently managed in each subnet, an interesting side-effect is how NSAs choose to use paths in context of slices. E.g. within a subnet an NSA can use shortest path for low latency slices and reroute best-effort services along alternate paths.

Network slicing framework requires a dynamic computation model, which means resource availability information in aggregated resource database should be most updated based on live updates from discovery function and remembering available and allocated resource information.

This function is invoked on network slice aggregated subnet resource database during network slice mapping from newly created network slice instance object in NSP.

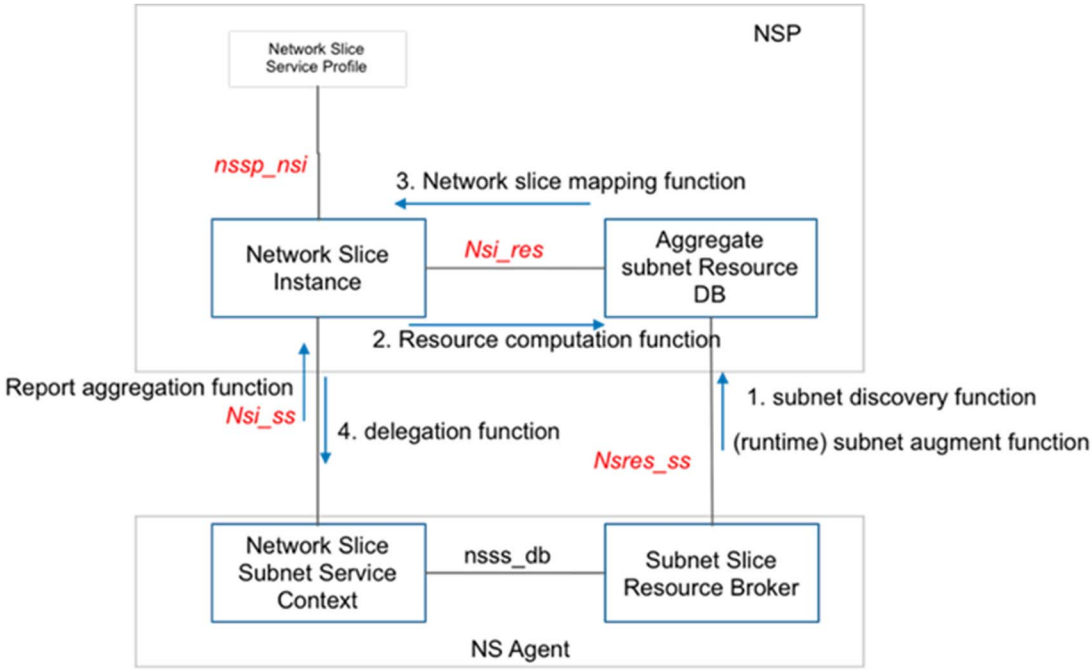

Figure 6: Network slice specific functions

6.6 Network slice delegation function

Delegating to subnets, segments of a network slice instance is the last step in setting up network slice instance and happens after computation and mapping functions are successful. After mapping is complete in NSP, network slice service segments are distributed to NSAs. There are 3 main aspects of this function:

1) For NSA to assign encapsulation of service isolation in its network.

2) To connect data paths from two adjacent subnets on NS gateways.

3) Along with the service mappings, operations are also registered for notifications/queries to/from NSP.

There are 3 possible cases that describe how tenant-flows traverse through its assigned path:

1) End-to-end encapsulation: A tenant flow has end to end encapsulation, negotiated by NSPs with NSAs. This is a limiting case of slicing, that does-not promote multi-technology i.e. all NSAs are expected to have same encapsulation/overlay.

2) Segmented encapsulation: A tenant flow is encapsulated in each subnet using different overlay technique such as MPLS in some, VXLAN in another subnet. Encapsulation mapping of flow to an encapsulation is decided by individual NSAs. The details of encap or tunnel need not be exposed to the tenant. Network slice gateways perform translations when moving across different network domains.

3) A tenant's flows are forwarded natively without any encapsulations, with only egress and ingress port mappings in a subnet with considerations for bandwidth, jitter and latency constraints as prescribed for the slices.

Thus the processing of this function in NSAs involves actual setting of data path for the flows in the network slice instance.

6.7 Report aggregation function

Aggregation is in the context of a network slice instance and provides complete end-to-end monitoring framework by sum of performance and monitoring characteristics from different slice subnets. NSP should be able to collect the following service specific parameters:

- Statistics collection at each segment for accounting.
- Variations in delay and jitter.
- Events such as path changes and link status from different subnets.

The report aggregation helps charging and accounting functions as well. While some reports are simple, for example if bandwidth should remain uniform across all segments, any fluctuation in a segment, or changes in link state etc. are easily reportable; other reports of operational anomalies are more difficult. As an example, in order to verify that latency requirements are met, latency for each segment has to be constantly monitored, this is a hardware assisted activity in which if latency quota exceeds, hardware nodes should raise an alarm to NSA which then gets propagated to NSP. The NSP instructs NSA to collect and report operations that were registered during the delegation function (clause 6.6).

6.8 Service assurance function

While reporting function is closer to accounting and performance for business model, service assurance means each flow in a network slice instance has to be continuously monitored by NSP in order to prove service level objectives are being operationally delivered. Similar to reporting this function is also split at the subnet NSA level. NSAs monitor their own networks by means of technology specific or proprietary tools. Generally, what to monitor is associated with the type of a network slice. In order to provide assurance either or both NSP and NSA can take decisions by forming control feedback loop between NSAs and NSP to resolve exceptions.

A few examples of service assurance are:

- NSA level monitoring of a low latency slice will have latency budget within each subnet and each NSA monitors that for its own subnet. Often it is expected that NSA will resolve anomalies and exceptions, by electing alternate path without changing end points (egress and ingress), making resources available in its original path.
- NSP level assurance: In case of exception (i.e. resource assurance cannot be met in a subnet), NSA informs NSP which in addition to reporting may also trigger resolution steps to amend the situation such as an (switch to alternate path) through feedback loop between NSP and NSA.

As long as network behaves normally no action should be needed, only exceptions, warnings or alarms are sent to NSPs. These activities are performed on resource DB. Upon failure of assurance at subnet level an exception is generated at resource broker database to notify the aggregate resource database in NSP.

Resource assurance may also be done without involvement of NSP or NSA. In this option tenant can use in-band/in-situ/in-transit operations, administration and maintenance (OAM) techniques. It allows a tenant to have full control over its data path (through service graph mapping) and be able to trigger its own OAM requests. In order to support this underlying slice subnets are required to enable in-band OAM capabilities.

6.9 Tenant operated network service function

6.9.1 Tenant operations overview

Each network slice instance is a tenant network and has a tenant-operator associated with it. The tenant expects to operate and manage its own slice. Network slice instance differ from other types of overlay network is in terms of dynamic or runtime control and management of network slice by a tenant-operator. Thus, this function covers details of type of operations and corresponding interface to underlying infrastructure. For the purpose of implementation these may be considered as a standard set of APIs from tenant to infrastructure via NSP (then via NSA) or directly between the tenant and infrastructure. There are three different type of operator driven functions with the following objectives (as in Figure 7) which are described in little more details below:

1) Description of flows attachment to headend/endpoints

2) Dynamic interface to slice specific resources

3) Runtime interface to flow specific monitoring

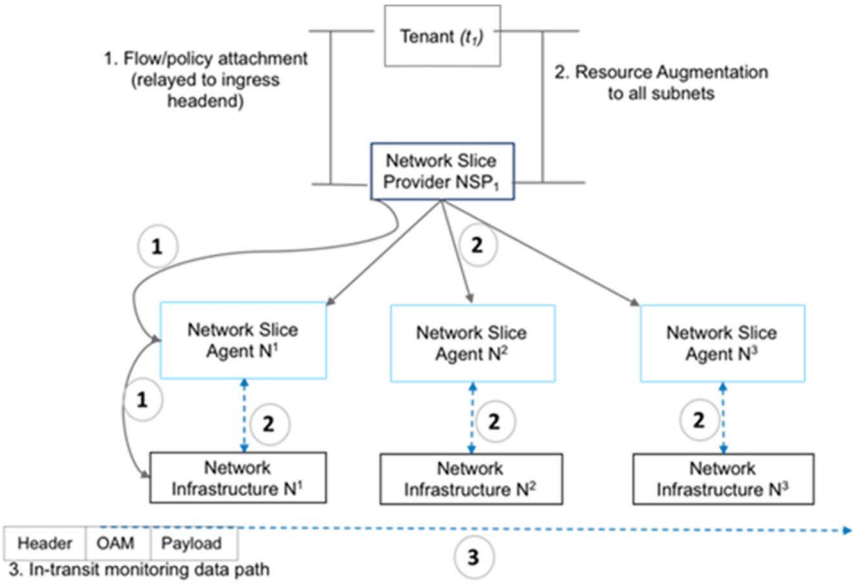

Figure 7: From Tenant to Infrastructure Interface

6.9.2 Service endpoint attachment

The attachment function determines how flows are mapped to a slice. This is typically done at the ingress headend; An API initiated by tenant operator will act on runtime object associated with tenant slice and call attach as defined in clause 5.3.4.2. The attach request traverses through NSP to NSA to install the flows to slice association. This request may comprise of admission control, security and service specific policies on the headend node.

It is the decision of a tenant-operator how they intend to perform identity and address management for their end users or flows.

6.9.3 Interface to slice specific resources

Network slices aim to provide dynamic control to a tenant over the resources associated with its slice instance. The interface is in the form of **tenant → NSP → NSA →network.** This is an API type slow interface and can be used to augment (request more bandwidth, additional path etc.) or query resources (for statistics of different types) as described in clause 5.3.4.2. Many other aspects require aggregated reporting and are used for accounting purposes and may be collected form NSAs via NSP, appropriate get APIs to ask for specific counter can be provided.

6.9.4 Tenant runtime OAM template

The mapping of a slice service graph sets up path and constraints over which different data flows in a tenant network get created. To perform runtime monitoring and management of these flows, a tenant specific OAM template is suggested. A template provides a format and standard specification to describe selectively what particular policies, statistics, monitoring and reporting of the ongoing flows of its subscribers are necessary. The template can specify runtime or in-transit OAM, telemetry, diagnostics. Network slices are often associated with precise and accurate latency, reliability and bandwidth requirements and this is best possible when each flow offers ability to be monitored individually. Since there is no direct interface between NSA and tenant, an enhanced data-path centric approach such as in-transit telemetry maybe be considered. A tenant can inject a certain telemetry request at the head end towards its own telemetry collection node (a VM) and physical nodes from each domain process those to provide relevant data.

7 Network Slice Enablement

7.1 Mechanisms for service assurance

7.1.1 Methods of assurance

Several candidate technologies already exist that can be used to enable or create a sliced network. The important network functions and capabilities related to network slice life-cycle management are service assurance, sliced instance path and data plane of a network slice. These are discussed as below.

Service assurance is achieved through reservation of resources at the network slice set up time, subsequent monitoring of those and then augmentation of slices if demanded by the tenant. The traditional tools available for reservations are:

- Classical Quality of Service Markings
- Traffic Engineering
- Constraint based Path Computations

7.1.2 Quality of service

Service assurances can be performed on network level and/or on per node basis. The Intserv [i.7], f) is a distributed per flow resource reservation technique and Diffserv works at an aggregated flow granularity with scheduling and shaping of flows happen based on Diffserv markings. The challenges with Intserv's complexity and scale and Diffserv's non-uniform treatment of flows can make it hard to implement end to end QoS.

In case of network slices, resource management is segmented and is only maintained at an abstract or logical level at NSP, therefore, a quantitative representation (i.e. amount in units and granularity) of resource should be provided from NSP to NSA. Diffserv implementation cannot provide the scalability that slices need, therefore, is not a solution option in this architecture. On the other hand, Intserv uses RSVP protocol as decentralized resource allocating mechanism; due to the impact on flows when RSVP path changes and the nature of cautious trust between the domains, RSVP in its existing form may not be suitable for the deployment of slices. It may still be used as a part of the solution for creating traffic engineered paths with in a domain with additional association of tenant operations and control on that path and an interface to NSP.

With regards to per-flow QoS in a slice, it can be safely said that all flows of a given service in a slice will have similar resource requirements. Thus, if resources for a single flow are represented by $(R(f))$, allocations for η flows are computed as a multiplier of resource per flow i.e. $\eta \times (R(f))$.

7.1.3 Traffic Engineering relevance

Traffic Engineering (TE) is a broad subject-domain essentially, dealing with alternative resource constraints-based path setup mechanisms. TE satisfies many aspects of the slicing resource assurance. TE solutions serve as useful tools in deploying some aspects of slices. The corresponding work at IETF, TEAS WG [i.8] is relevant where RSVP-TE or SDN based traffic engineering techniques such as PCE are deployed at the network infrastructure. RSVP-TE mandates MPLS data plane for path setup and thus only applies to slice-subnets that choose to support MPLS. TE solutions reflect operator centric control of networks and as a result there is no direct notion of operational support from tenant services. The process of an end to end TE-tunnel setup has complexity and lack of dynamic OAM support are among few gaps that prevent TEAS defined framework to be used as is for network slices.

In contrast, goal of network slicing architecture is to provide an infrastructure agnostic method for carving simple sliced-networks from the customer's point of view. Since network slices here is referred to as a service graph and mainly focuses on vertical markets, it can be assumed that the topologies will be less complex than a typical traffic-engineered network.

7.1.4 Path computation relevance

Path computation element (PCE) is used to compute constrained based topology and manage traffic engineering database (TED). It uses PCEP [i.3] to communicate with network infrastructure nodes. The PCE works in a hierarchical way so it is possible to support parent PCE on NSP and child PCE on NSAs. The PCE offers ability to determine paths; the remaining functions such as what resources should be exported are performed separately. PCE can determine which physical resources can be mapped to a slice-request and compute topology associated with the network slice. While some resources are defined as traffic specifications [i.6], additional constraints may need to be added. In figure 8, it is an example to demonstrate how PCE could be applied to the architecture. The roles between PCE at NSA and NSP could be child and parent respectively.

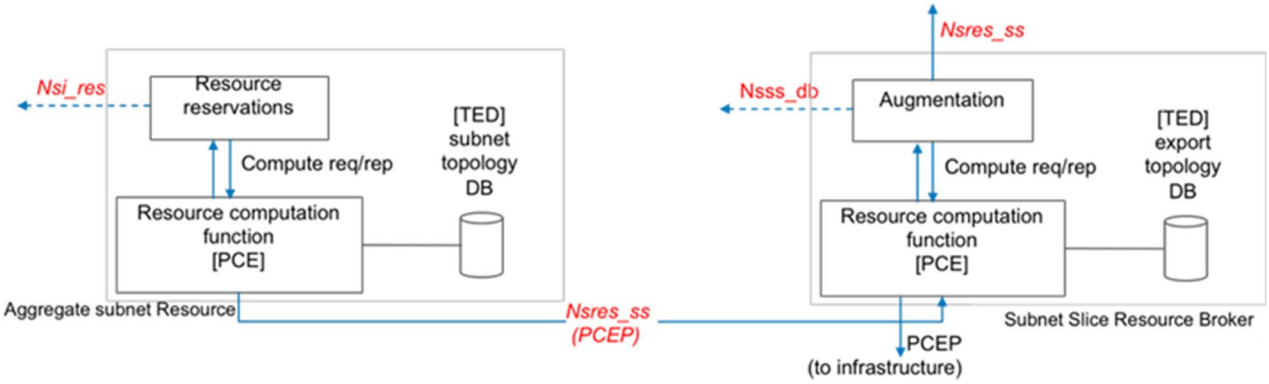

Figure 8: Relevance of PCE in Slicing architecture

7.2 Mechanisms for OAM

Report aggregation function allows to accumulate different slice parameters related to runtime state through offline means and is useful for common charging specific information, in order to monitor health of a particular flow in a network slice instance, in-transit OAM can serve as an enabler. This is discussed in clause 6.9.4. There is been active work in IP performance measurement (IPPM) WG, defining in-situ OAM data fields [i.20]. The data-fields defined in this draft can be used for IP based flows with in a network slice.

7.3 Data path enablement

7.3.1 Enabling approaches

The architecture does not mandate a particular data plane enabling technology, neither is it a proponent of a single data plane encapsulation. There are two approaches to implement user planes for network slice services:

- Existing IP based infrastructure
- Next-generation lowest-level sliced infrastructure

7.3.2 Existing IP based Infrastructure

7.3.2.1 IP Based Modes

Network slicing is a relevant concept for interconnection over large scale, multiple domain networks. The end to end interconnectivity and isolation can further be provided in two modes:

- An end to end encapsulation
- Segmented encapsulation

7.3.2.2 End-to-end encapsulated mode

In an **end-to-end encapsulated** mode, a tenant flow is encapsulated in a single end to end tunnel with same overlay/isolation technique (L3VPN, L2VPN, MPLS, etc.). Admittedly, it is a very simple scenario from data forwarding purposes but it does not meet a technology agnostic approach, e.g. in order to support MPLS based VPNs, all slice subnet may support MPLS and could distribute labels to all NSAs. This may not always be desirable by all NSAs. Not only does it lead to overall increased complexity of the control and management of the system comprising of NSP, NSAs and corresponding infrastructure, it also causes scale related issues, e.g. how many different types of services can be supported.

7.3.2.3 Segmented encapsulated mode

This mode supports different encapsulation per segment (or subnet) allowing each NSA to independently select and manage its overlay/isolation method. Then the requirement is how to resolve forwarding of sliced data from one subnet to the other? This is done by defining and propagating standard segment stitching procedures between NSA and NSP. In this architecture network slice gateways are proposed as special subnet nodes that will perform the stitching function to translate/strip/impose encapsulations.

Architecture makes no assumption about network slices being end to end encapsulated Therefore, in a multi-domain network slicing system stitching of different technologies is necessary by use of network slice gateway components (see clause 7.3.4.

7.3.3 Next-Generation Sliced Infrastructure

A long-term goal of network slices allows for innovation of next-generation of protocols suitable to carry slice aware data. A possible approach may be forwarding data in a slice over media access layer network based on encapsulation or control header defined by new protocol known only in that slice. In this mode, tenant runs its own control protocol for distribution of forwarding rules and flows are forwarded over infrastructure without any IP encapsulations. One way to visualize this is through a low-level virtual data-fabric with egress and ingress logical port mappings. Such an approach requires partitioning of resources in hardware-based manner. With low level partitioning or abstraction of hardware that creates a sliced-fabric with all forwarding logic determined by a network slice.

This approach however needs to be studied further and not covered in the present document which will require identifying data objects to describe an abstract fabric and forwarding properties of ingress and egress ports.

The proposed architecture is extensible to support new object definitions.

7.3.4 Network Slice Stitching Gateways

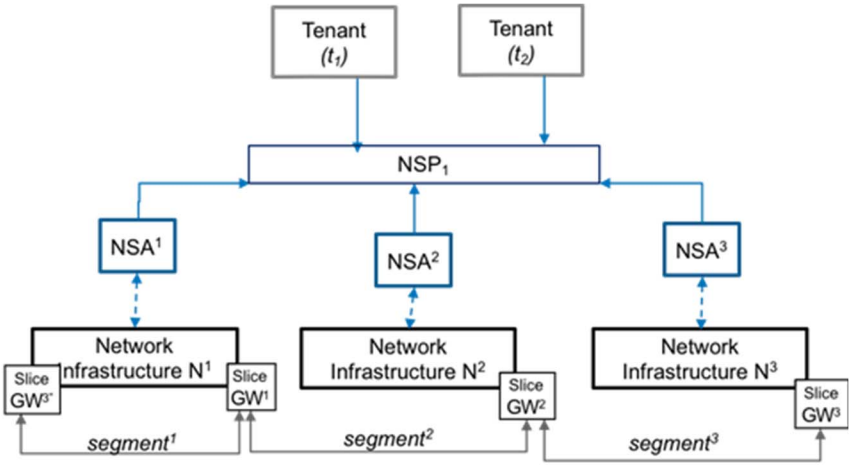

Figure 9: Network slice gateways for interconnections

Network slice comprises of multiple subnets. As data flows through a slice it traverses through different subnets, the data path in each subnet is represented by segment (e.g. segment[1], segment[2] segment[3]). When connected together these segments form end to end data path of flows belonging to a slice. A segment can be independently isolated with in a subnet using any domain specific option. The network slice gateways therefore, are logical nodes where the interconnection of segments of slices take place. The gateways are functionally significant to the data forwarding aspects in many ways. For example, the gateways can do domain specific ingress-, egress- resource monitoring, profiling and accounting. In this manner gateways can help understand which domain is under-performing or is unable to meet service level objectives.

8 Security Considerations

8.1 NGMN security guidelines

The security aspects of network slices and recommendations are thoroughly covered in [i.9] and those of concern from architecture perspective are impersonation attacks against various network slice components, unintentional or malicious exhaustion of resources, and isolation of network slice resources. Besides these other vulnerabilities can very well occur but present document assumes that they will be addressed as per the network-design and best practices known for security of virtualized infrastructures (such as, security of NFV, virtual network functions, etc.).

Network slices run over leased infrastructures, therefore, security of the network from malicious tenants; protection of tenant data from each other and encryption of tenant user data are all important aspects. From this architecture's perspective, protection of sensitive data with in a slice has to be carefully managed because each slice subnet is an independent security domain in itself. The vulnerabilities and strengths of security mechanisms will vary across subnets, therefore, all communication channels - control, management or data forwarding should be protected for each tenant. This architecture recommends three separate aspects of security as follows.

8.2 Protection and privacy of tenant data

This clause concern with protecting network slice user (or data) plane. The present document recommends that end to end protection of tenant data flows should be triggered by the tenants themselves as part of security policies. Encryption through different type of crypto algorithms can be applied through tenant policies. A tenant has own admission control policies to allow only tenant-data that are associated at ingress attachment point. Furthermore, the architecture allows resource allocations and changes through controlled and secured NSP to NSA interface. This allows ability to authenticate, authorize and verify resource-specific requests at each component.

8.3 Tenant resource isolation

Resource allocated to a tenant's slice should not be shared with other tenants in order to protect sensitive data. This is partly related to trust on NSAs. It is the responsibility of NSA to ensure resource allocated once are not re-allocated to other tenants. The concept of prescribed path of a tenant's slice allows clear mappings between a tenant and resources.

In addition to allocated resources, there is an implied consumption of hardware resource tables that are not generally exported for example Ternary content-addressable memory (TCAM) and (FIB) memory. The platforms based on this architecture may incorporate checks to ensure that:

a) these resources are not disproportionately used, a network slice instance,

b) are isolated from other instances.

8.4 Protection against impersonation attacks

Key issues 1 and 2 in [i.9] discuss the notion of establishing trust with NSP. In the proposed architecture, resource discovery function should ensure that NSP to NSA communication control channel is secured. This way, NSAs export resource information to only authenticated NSP. Furthermore, the architecture expects from NSAs to not to export physical resource component specific information directly, but provide a logical mapping of those elements in order to protect internal details and direct control of components.

9 Integration Example

9.1 Generic purpose service slice

9.1.1 Scenario description

This example demonstrates the use of slicing architecture to manage and control a tenant network in integrated, unified way typically deployed as Virtual Private Network (VPN) service. The goal is fairly simple, to provide connectivity from heterogeneous accesses (technology independent) and leverage network slice architecture for common control.

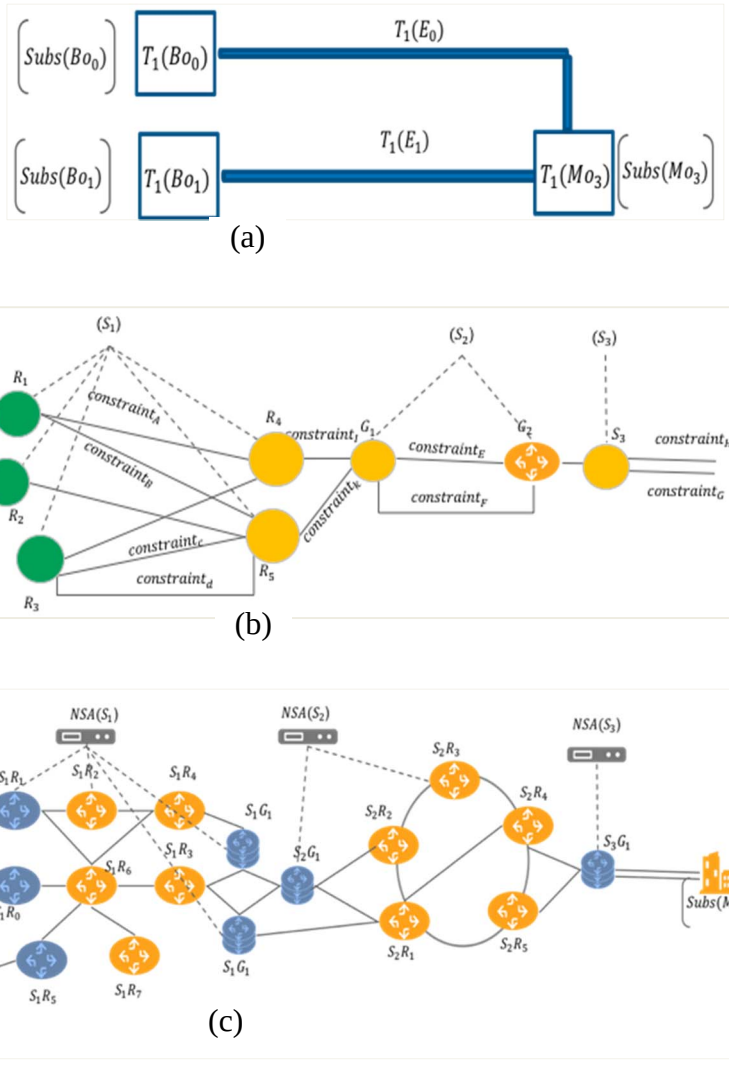

Figure 10: (a) Tenant's View, (b) NSP view, (c) Physical Network View

Figure 10 (a) shows a **tenant view**, it requires the connectivity as a tenant network service and it may have additional constraints. This is represented as a network service profile. VPN connectivity from two branch offices Bo_0 and Bo_1 and main site Mo_3 is desired by tenant T_1 so that constraints of E_o implying bandwidth 5 Mbps, E_1 per user and implying low latency of 1 sec are met.

A **network slice infrastructure view**, Figure 10 (c) is the actual physical network connections across different network domains shown here as S_1, S_2 and S_3. Each of these domains may have any type of complex connectivity infrastructure that may get used for both general-purpose and network slice-based services. The tenant's branch offices Bo_0 and Bo_1 are connected to subnet **S_1**. The Main office Mo_1 connected to S_3.

The **logical network slice provider view** (NSP view) is a reduced and exported view of topology independent connectivity is seen as in Figure 10 (b).

9.1.2 Network slice bootstrap

Bootstrap process is initialization of resources in NSP resource data base, this has been explained earlier. Once NSA and NSP are authenticated, a push-based protocol that allows NSP and NSA authentication and discovery functions can be deployed to build working constrained topology at NSP.

- Different domain administrations announce how much of their resources they intend to utilize for network slice infrastructure and accordingly export logical constrained path mappings to the NSP, which views the entire network as in Figure 10 (b). This is the outcome of **slice subnet discovery function**.
- The exported resource mappings are held at respective NSAs (S_1, S_2, S_3) level databases (e.g. resource or node $R_1 \in$ *slice subnet* S_1).

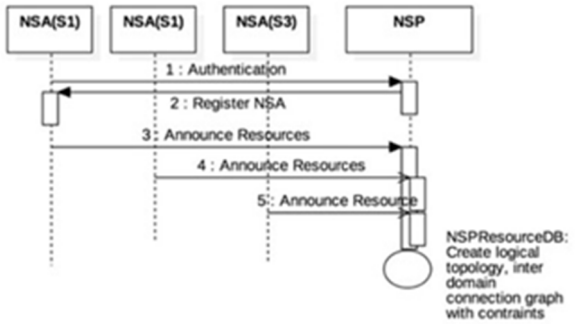

Figure 11: Network slice Bootstrap

9.1.3 Network slice onboarding

To implement such a service through network slices the following functions are used: **Network slice mapping function**, **Resource computation** function and **Network slice delegation function**. considerations are made:

- Tenant creates a service graph of network slice service profile and deliver to a Network slice provider. The NSP would determine head ends or attachment points for the 3 sites based on the geography of locations and determines the path based on the attachments and resource constraints, path from branch offices to main office. For example, path $R_1 \rightarrow R_5 \rightarrow G_1 \rightarrow G_2 \rightarrow S_3$ will have available 5 Mbps bandwidth for Bo_0 to Mo_0. This will be an outcome of Resource computation function.
- Network slice mapping function determines and maintains a mapping in NSP as follows: $T_1(Bo_0) \rightarrow R_1, T_1(Mo_0) \rightarrow S_3, T_1(E_o) \rightarrow R_5 \rightarrow G_1 \rightarrow G_2$ on this network slice instance basis.
- NSP resource database has knowledge of which node belongs to which slice subnet and perform corresponding delegations to NSA of each domain. It can further register for error notifications if the NSA fails to allocate exported resources. The consequence of this step is that network slice service is in effect and instantiated.

9.1.4 Slice operation and management

A key capability differentiation of network slice from other network services is that it provides:

- Tenant level: an interface to network slice tenant to control, operate and monitor resources as discussed in clause 6.9, that allows in-band monitoring capabilities to get statistics and state from the data path directly.
- NSP level: In clause 6.8 slice monitoring for service assurance is described primarily relying on slice from different domains. NSP level service assurance helps to isolate a particular subnet that is unable to meet network slice requirements and make end to end path changes.
- The aggregated report as described in clause 6.7 additionally helps statistics collection (Report aggregation function) primarily used for accounting of resource usage.

Annex A:
Authors & contributors

The following people have contributed to the present document:

Rapporteur:
Kiran, Makhijani, Huawei Technologies

Other contributors:
Kevin Smith, Vodafone

John Grant, Ninetiles

Alex Galis, UCL

Xavier Defoy, Interdigital

Annex B: Bibliography

IETF draft-defoy-coms-subnet-interconnection-02: "Interconnecting (or Stitching) Network Slice Subnets".

NOTE: Available at https://tools.ietf.org/id/draft-defoy-coms-subnet-interconnection-02.html.

Annex C:
Change History

Date	Version	Information about changes
January 2018	0.1	Pre-draft
February 2018	0.2	Skeleton first review for NGP 10. Covered Functions
February 2018	0.3	Addressed initial review comments
April 2018	0.4	Stable draft, Improved gap analysis, added Tenant related operations
July 2018	0.5	Added security recommendations section and integration example

History

Document history		
V1.1.1	September 2018	Publication