teas
Internet-Draft
Intended status: Informational
Expires: October 23, 2020

R. Rokui
Nokia
S. Homma
NTT
K. Makhijani
Futurewei
LM. Contreras
Telefonica

April 21, 2020

IETF Definition of Transport Slice
draft-nsdt-teas-transport-slice-definition-02

Abstract

This document describes the definition of a slice in the transport networks and its characteristics. The purpose here is to bring clarity and a common understanding of the transport slice concept and describe related terms and their meaning. It explains how transport slices can be used in end to end network slices, or independently.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 23, 2020.

Copyright Notice

Table of Contents

1. Introduction

A number of use cases benefit from establishing a transport connectivity according to the assurance of a specific set of network resources. Some such services which might benefit from the transport slices are:

- 5G services (e.g. eMBB, URLLC, mMTC)(See [TS.23.501-3GPP])
- Network wholesale services
- Network infrastructure sharing among the operators
- NFV connectivity (Data Center Interconnect)
- VPNs with specific characteristics

This document defines the concept of transport slices that provide connectivity with specific use of network resources between a number of end points over a shared network infrastructure. Transport slices are created and managed within the scope of transport networks (e.g. IP, MPLS, optical). They are expected to enable a diverse set of applications that have different requirements on communication to coexist on the same network infrastructure.

Transport slices relate to a more general topic of network slicing. It is not the goal of this document to define this broader concept, but in general, it is a methodology to describe the logical partitioning of network resources associated with a service or an application.

2. Terms and Abbreviations

The terms and abbreviations used in this document are listed below.

- E2E NS: End to End Network Slice
- TS: Transport Slice
- TSC: Transport Slice Controller
- EP: Endpoint
- EU: End User
- NBI: NorthBound Interface
- SBI: SouthBound Interface
- SLO: Service Level Objective
- SLA: Service Level Agreement
- MTBF: Mean Time Between Failures
- MTTR: Mean Time To Repair

Author's notes: This list may be non-exhaustive. Also, a light explanation for each term would be required. Or this section may be removed if it isn't needed.

3. Definition and Scope of Transport slice

The basic definition of a transport slice is as follows:

"A transport slice is a logical network topology connecting a number of endpoints and a set of shared or dedicated network resources, which are used to satisfy specific Service Level Objectives (SLO)".

SLOs are used to describe different network resources associated with the service delivered and corresponding parameters necessary to realize the transport slice.

Transport slice should be technology-agnostic, and the means for realization can be chosen depending on several factors such as service requirements, specifications or capabilities of underlying infrastructure. The structure and different characteristics of transport slices are described in the following sections.

4. Transport Slice System Characteristics

The characteristics here describe the properties and main functionality related to different components of a system that supports transport slices.

4.1. Service Level Objectives on Transport Slice

A transport slice is defined in terms of several quantifiable objectives or SLOs. These objectives define a set of network resource parameters or values necessary to provide a service a given transport slice. SLOs need not concern with 'how' will they get implemented or realized in the underlying networks. They may be defined along the dimensions of operations (time, capacity, compute...), reliability and, availability attributes. A non-exhaustive list of characteristics types for transport slice is described below:

- Guaranteed Bandwidth: assurance of minimum or range of the bandwidth requirement. Requested unidirectionally.

- Guaranteed Latency: maximum permissible network delay when transmitting between source and destination endpoints. Requested unidirectionally. The latency is measured in terms of network characteristics (excluding application-level latency). [RFC2681] and [RFC7679] discuss round trip times and one-way metrics, respectively.

- Minimal permissible jitter: packet delay variation (PDV) as defined by [RFC3393] is measured by the difference in the one-way

delay between selected packets in a flow. Minimizing variations in the delay are important for real-time applications, therefore, jitter-prevention parameter provide minimal permissible jitter expectations for a flow.

- Packet loss rate: To specify permissible packet loss rate between two endpoints. For critical networks, this number may be very close to zero. See [RFC7680].

- NF resources: The endpoints in Section 4.2 performance depend on resource allocated to those functions and can be specified in terms of compute, memory and access control objectives points. See [NFVGST].

- Availability: concerns with how quickly network becomes available after a fault. The requirements are specified through Meant time between failures (MTBF), and Mean time to repair (MTTR) to acquire availability metrics.

- Resource redundancy: represents reliability attributes in which a backup or duplicate resources such as path (same SLOs - latency, bandwidth, etc.), network functions (same compute, memory, etc.) To meet no packet loss objective, packet replication maybe necessary to guarantee that at least packets from one path was achieved. However, we should consider this as 'how' aspect of objective and not 'what'.

- Security: The objective of securing a transport slice concern with three attributes: a) end-to-end encryption between source and destination endpoints, this can be seen as the logical link between source and destination end points requiring encryption, b) Authentication and access control (ACLs) objectives to permit data on this transport slice, c) Isolation, is also a characteristic of security, to prevent interference between two or more slices or other flows on the same infrastructure. Isolation is implied by the definition of transport slice itself (logical partitioning...).

- Resolution of guarantee: The above objectives can be resolved in to either hard or soft guarantees. A hard guarantee is the one that is not affected by other traffic. In a soft guarantee, a violation (of the guarantee) may occur in rare cases due to resource interference. In such cases, the guarantee will be maintained by the network controller within a certain tolerance level of that objective. Note that a hard guarantee does not prevent from hardware failures, such as losing a node. Additional protection against such issues is possible, by specifying those characteristics separately (see item "resource redundancy" below).

Note also that the hard and soft guarantees do not say anything about the specific implementation of how these guarantees are achieved. Different implementations might use different techniques, from avoiding oversubsription to dedicating particular links or their virtual fractions to particular transport slices.

- Resource isolation: In some cases it may be necessary to dedicate specific resources to the slice, for instance, for security reasons.

- etc.

The framework [I-D.nsdt-teas-ns-framework] may further specify mechanisms for the performance, assurance and monitoring of these objectives.

Note: Additional objectives may be necessary to capture, such as specifying well defined paths or domains that a slice may transit. A transport slice carries multiple flows between the 2 endpoints, therefore objectives should also say if they are aggregates or on per flow basis and in such case to be specific enough for the system to be able to identify these specifics (subset of flows).

Further description of a set of measurable attributes is captured in [I-D.contreras-teas-slice-nbi].

SLA vs SLO discussion: In defining transport slices, the term SLO instead of SLA is used even though SLAs are more commonly used term by the operators. SLOs are definitive and measurable parameters associated with a service, therefore, network resource and connectivity requirements are defined accurately. In contrast, service level agreements represent contracts for a service between a service provider and a service consumer (or subscriber). Providers then translate SLA into SLO; these translations vary from one service provider to the other. Therefore, all through within the scope of transport slices term SLO will be used.

4.1.1. Isolation discussion

Due to overloading of the term, a further discussion is added to highlight two aspects of isolation, first the resolution of isolation of an objective (as described above) and second, the dedicated use or a hard-separation perspective of the resource.

Providing a hard resolution of guarantee for the characteristics of a transport slice means that the behavior and performance of other transport slices should not impact that slice, even if they run over

the same underlying infrastructure or use logically shared network resources.

In the context of soft resolution of guarantees, since the transport slices are logically partitioned over the shared resources, a certain degree of commitment to the guarantee is expected even when it is not hard. When the shared resource pools begin to become saturated, SLO violations can happen, however, impacting only the performance or operation of service associated with the transport slice.

This degree of isolation can be derived from availability characteristics requested, such as whether a hard or soft guarantee was requested. Requesting a hard guarantee may commit more resources than would be required for a softer limit.

In addition, resource isolation may be applied to ensure dedicated access to a particular node, for instance. In such requests a dedicated allocation to a link, node and/or other resources to create a transport slice for a particular service. For example, a mission-critical service may ask for a dedicated router and/or a link or port for complete isolation from other services.

When realizing a transport slice, a network controller should be responsible for allocating and providing resources according to the specified objectives.

SLO violations can occur for two reasons and corresponding statements apply

- Shared resource interference: i.e. multiple transport slices simultaneously share the same resource, and one of them consume the resource in surplus. If the SLO guarantees are strictly required, then the network controller can be informed of this by requesting a hard guarantee. Note that the terms hard and soft limit are requirement oriented and different from what is specified in, [I-D.ietf-teas-enhanced-vpn]).

- Resource failure or fault occurs, such as a link or node failure. Where it is important to defend against these, the relevant characteristics on resource redundancy (and perhaps some other characteristics on restoration speed and other factors) need to be specified.

  * Restoration isolation: the network is not impacted for a period longer than the given time. For example, failover or the service restoration takes no longer than some number of seconds. This is specified by Availability SLO.

* Protection isolation: the network path is protected with specified backup path. This is specified by Availability SLO.

4.2. Endpoint Variation

Transport slice endpoints are the terminating or originating nodes requiring connectivity with specific SLO. Endpoints may be devices or functions.

4.2.1. Types of Endpoints

There are two types of endpoints based on the functions they perform.

Transport type EP: This type of end point performs forwarding customer payload without any modification. E.g. routers, switches.

Service type EP: This type of endpoint manipulates, processes or modifies the user data payload (based on policies). A non-exhaustive list of service functions includes: firewalls, WAN and application acceleration, Deep Packet Inspection (DPI), server load balancers, NAT44 [RFC3022], NAT64 [RFC6146], HTTP header enrichment functions, and TCP optimizers. The generic term "L4-L7 services" is often used to describe such service functions (SFs).

This document leverages the term Network Function (NF) to represent both types of endpoints in [I-D.ietf-teas-sf-aware-topo-model].

4.2.2. Connectivity Patterns

Endpoints may be connected point to point (P2P), point to multipoint (P2MP), multi-point to point (MP2P), or multi-point to multi-point (MP2MP) based on the topology requested by the customer.

P2P pattern: P2P type of connections are between 2 endpoints like a pseudowire, or a logical link. The interconnections must represent the SLOs as requested by the customer.

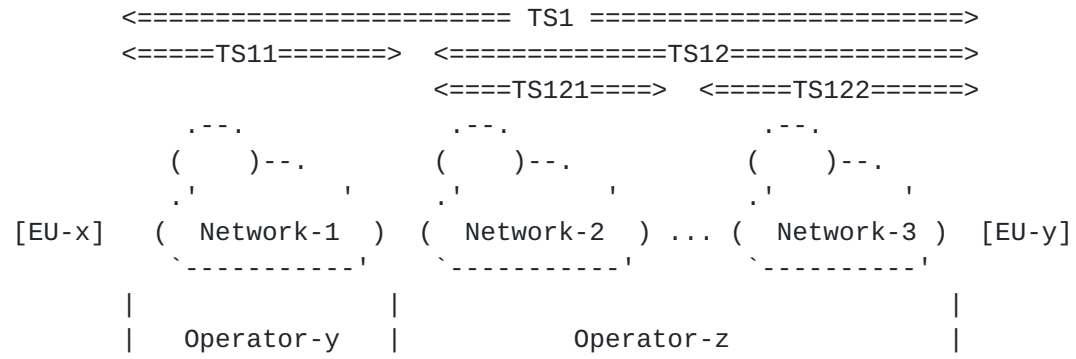
P2MP/MP2P/MP2MP patterns: P2MP/MP2P/MP2MP type connections will interconnect two or more endpoints together (one to many, many to one, many to many), representing an abstract topology or graph. When describing P2MP/MP2P/MP2MP scenarios it should be possible for each logical link to have different SLO than the other link in the same graph.

4.3. Vertical Transport Slice

Transport slice may follow a hierarchical relationship that would provide a vertical structure to it. This is used for building multi-layer slices in which each layer provides an abstraction, as well as an independent monitoring, performance, control and management of the resources. The vertical transport slice characteristic maybe used in 2 forms:

- The Transport slice itself where it represents a hierarchy of abstracted transport slices. In this case, realization will be done just once with a particular technology. Thus, the lowest transport slice in the hierarchy that can not be decomposed further will be one to one mapping to its instance of realized transport slice.

- Each layer (physical, datalink, or IP) has its own set of resources that can be provided to the upper layer as a transport slice. Thus, transport slice at one layer is used by the layer above. This type of multi-layer vertical transport slice associates resources at different layers. For example, an IP transport slice would utilize one or more optical transport slice. In this case, realization will be done for a particular technology at that particular layer. Thus, the lowest transport slice in this type of hierarchy that can not be decomposed further will be an instance of realized physical layer transport slice.



Legend:
TSnnn: Level 3 vertical transport slice nnn
TSnn: Level 2 vertical transport slice nn
TSn: Level 1 transport Slice n

Figure 1: Transport Slice Vertical and Horizontal Composition

Figure 1 shows the transport slice hierarchy. Slices TS11 and TS12 are composed together to form TS1 that is the top level transport

slice definition, TS121 and TS122 collectively define TS12. The SLO for bandwidth guarantee will be shared and latency guarantee will be split into latency in networks 2 and 3. To emphasize the hierarchical structure, consider Network-2 and Network-3 are in the same administrative domain but use different transport technologies SR and L2VPN respectively. Then instead of presenting 2 transport slices, Operator-z can expose only one transport slice TS12 abstracting the underlying transport technology details.

Note: The specification to connect TS121 and TS122 are similar to those connecting TS12 and TS11.

4.4. Horizontal Composition of Transport slice

In contrast, horizontal transport slices enable the composition of multiple realized transport slices. Since transport slices are not necessarily a single encapsulation tunnel and may traverse through different data planes, each realized transport slice will require a stitching, interworking or mapping function. These stitching functions can be viewed as a type of intermediate network function endpoints. For instance in Figure 1, TS11 and TS12 are horizontal transport slices. If we assume that TS11 is an L2 tunnel and TS12 is an SRV6 based path, then a 'Service type EP' (not shown in the figure) is needed for translation.

Author's notes: This service type EP is a new type of transport slice specific service function. We may call it transport slice gateway.

5. Transport Slice Structure

A transport slice has a set of connections and various endpoints to form a logical network. The goal is to achieve specific SLO for a customer as shown in Figure 2. The endpoints may be user equipment, any physical or virtual network functions (PNF/VNF), or any network service for that matter. Similarly, connections may be virtual or physical links of any type of technology.

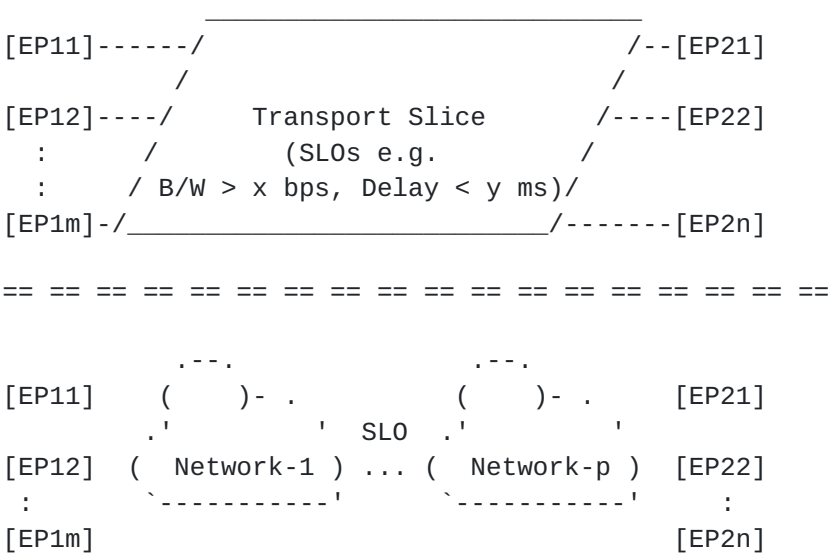


Figure 2: Transport slice

Figure 2 illustrates a case where a single transport slice provides connectivity between any pair of endpoints with specific characteristics for SLO (i.e., assuring bandwidth to at least x bps and transmission delay to be less than y ms). The endpoints may be distributed in the underlay networks, and transport slice can be deployed across multiple network domains. Also, the endpoints on the same transport slice may belong to the same address space.

The "Transport Slices" provides various connections with certain SLO between various endpoints whereas the transport slice realization addresses its implementation using various technologies. In short, the structure of a transport slice involves both definition and its realization.

A transport slice is built based on a request from a higher operations system. The interface to higher operations systems should express the needed connectivity in a technology-agnostic way, and slice customers don't need to recognize concrete configurations based on the technologies (e.g being more declarative than imperative). The request to instantiate a transport slice is represented with some indicators such as SLO, and technologies are selected and managed accordingly.

In the context of network slices, the term sub-slice or slice-subnet comes up in other standard organizations, however, w.r.t. the IP/MPLS based transport networks these terms are all equivalent.

Furthermore, the structure of transport slices may be layered vertically or composed horizontally, i.e. operationally, a transport slice maybe decomposed in two or more transport slices which are then independently realized and managed. This is further described in Section 4.3.

5.1. Stakeholders

A transport slice and its realization involves the following stakeholders and it is relevant to define them for consistent terminology.

Customer or User: A customer is a user of transport slice. Customers may request for monitoring of associated resources or specific changes to them. A user may either directly manage its service by interfacing with the transport slice controller or indirectly through an orchestrator.

Orchestrator: An orchestrator is an entity that aggregates different services, resource and network requirements. It interfaces with the transport slice controllers.

Transport Slice Controller (TSC): It realizes a transport slice in the network, maintains and monitors the run-time state of resources and topologies associated with it. A well-defined interface is needed between different types of transport slice controller and different types of orchestrator. A transport slice operator (or slice operator for short) manages one or more transport slices using the Transport Slice Controller(s).

Transport Network Controller: is some form of network infrastructure controller that offers network resources to TSC to realize a particular transport slice. These may be existing network controllers associated with one or more specific technologies that may be adapted to the function of realizing transport slices in a network.

5.2. Transport Slice Controller Interfaces

The interworking and inter-operability among the different stakeholders is required to provide common means of provisioning, operating and monitoring the transport slices. The following communication interfaces are identified (see Figure 3).

TSC Northbound Interface (NBI): The TSC Northbound Interface is an interface between a higher level system, e.g. 'E2E network slice orchestrator' and the 'Transport slice controller'. It is a technology agnostic interface. Over this NBI, slice characteristics and other requirements can be informed to TSC and current state of a transport slice may be requested.

TSC Southbound Interface (SBI): The TSC Southbound Interface is an interface between 'Transport slice controller' and network controller(s). These interfaces are technology-specific and can utilize many of the existing data models such as L2SM, L3SM, VPN, etc. TSC may request for network resources or request of their current state for SLO assurance.

Note on technology -agnostic vs -specific use: These terms are used in a transport slice's context. A transport slice from customer level in TSC, is not concerned with the underlying network protocol or technology (such as L2VPN, L2VPN, etc.) or corresponding service model (L2SM, L3SM, etc.) representing that protocol. Therefore, for example, both L2VPN, L2SM are technology-specific from a customer of a slice's view. Technology-agnostic simply means representing a transport slice completely as a logical entity.

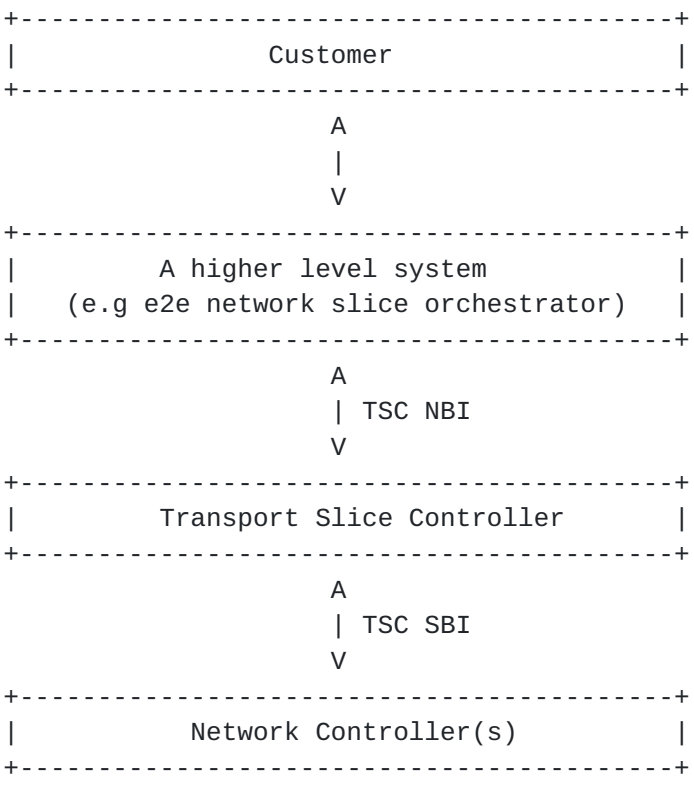


Figure 3: Interface of Transport Slice Controller

5.3. Transport slice Realization

Realization of a Transport Slice is a mapping of underlying infrastructure with its definition. It is technology specific entity that is created and maintained over southbound interfaces. The Network controller(s) export the connectivity and resource mappings to the TSC. The network controller abstracts the details of underlying resources from the TSC.

The realization may be achieved in the form of either physical or logical connectivity through VPNs, a variety of tunneling technologies such as segment routing, SFC, etc. Accordingly, endpoints may be realized as physical or logical service or network functions.

6. Relationship with End-to-End Network Slicing

An end-to-end (E2E) network slice is a complete logical network that provides a service in its entirety with a specific assurance to the customer. A transport slice concerns with those assurance aspects only within the transport networks. Consider Figure 4, where a network operator has an E2E network slice that traverses multiple

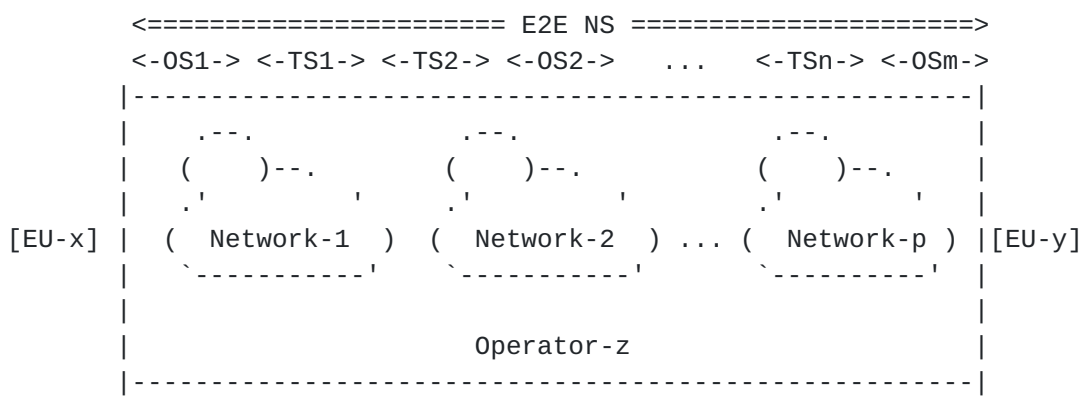
technology-specific networks. Each of these networks might use any number of technologies, including but not limited to IP, MPLS, Fiber-Optics (e.g. WDM, DWDM), Passive Optical Networking (PON), Microwave, etc.

Each of these networks includes multiple (physical or virtual) nodes and may also provide network functions beyond simply carrying of technology-specific protocol data units. The types of nodes used in any of these networks may include:

- Packet/frame processing nodes (e.g., Routers, Switches)
- Application servers
- Service Functions(e.g., Firewall, Loadbalancer)
- Radio Access Network (RAN) components
- Mobile Core components
- Microwave transceivers
- Optical repeaters
- etc.

Each network may support different technologies and an E2E network slice is a combination of these networks. As an example:

- Network 1 might contain multiple 5G RAN nodes connected to a few Cell Site Gateways (CSG) routers.
- Network 2 might have one or more layer-3 routers and layer-2 switches which may run on top of an optical network.
- Network 3 might have a number of 5G RAN nodes connected to Passive Optical Network (PON) switches.



Legend:
E2E NS: End-to-end network slice
TSn: Transport Slice n
OSm: Other Slice m
EU-x: End User-x
EU-y: End User-y

Figure 4: E2E network slice

When an operator-z creates a specific E2E network slice, it may create one or more of transport slices and other slices (application logic or other system functions).

An independent E2E logical network (called E2E network slice) is created for a service (e.g. CCTV, autonomous driving, HD map, etc.) with a specific network SLO requirement e.g. a secure connection with an E2E latency less than 5ms, from End User-x (EU-x) to End User-y (EU-y). EU-x maybe a 5G user equipment such as an infotainment unit in a car, CCTV, or a car for autonomous driving, etc. and EU-y in 5G is 5G application server, IMS, etc.

In Figure 4, "E2E NS" is that logical network with requested SLO between EU-x to EU-y and is associated with a customer and a specific service type.

7. Security Considerations

TBD

8. IANA Considerations

This memo includes no request to IANA.

9. Acknowledgment

The entire TEAS NS design team and everyone participating in those discussion has contributed to this draft. Particularly, Eric Gray, Xufeng Liu, Jie Dong, Jeff Tantsura, and Jari Arkko for a thorough review among other contributions.

10. Informative References

[I-D.contreras-teas-slice-nbi]
Contreras, L., Homma, S., and J. Ordonez-Lucena, "Considerations for defining a Transport Slice NBI", draft-contreras-teas-slice-nbi-01 (work in progress), March 2020.

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Networks (VPN+) Services", draft-ietf-teas-enhanced-vpn-05 (work in progress), February 2020.

[I-D.ietf-teas-sf-aware-topo-model]
Bryskin, I., Liu, X., Lee, Y., Guichard, J., Contreras, L., Ceccarelli, D., and J. Tantsura, "SF Aware TE Topology YANG Model", draft-ietf-teas-sf-aware-topo-model-05 (work in progress), March 2020.

[I-D.nsdt-teas-ns-framework]
Gray, E. and J. Drake, "Framework for Transport Network Slices", draft-nsdt-teas-ns-framework-02 (work in progress), March 2020.

[NFVGST] ETSI, "NFVI Compute and Network Metrics Specification", Febuary 2018, <https://www.etsi.org/deliver/etsi_gs/NFV-TST/001_099/008/02.04.01_60/gs_nfv-tst008v020401p.pdf>.

[RFC2681] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, DOI 10.17487/RFC2681, September 1999, <https://www.rfc-editor.org/info/rfc2681>.

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, DOI 10.17487/RFC3022, January 2001, <https://www.rfc-editor.org/info/rfc3022>.

[RFC3393] Demichelis, C. and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, DOI 10.17487/RFC3393, November 2002, <https://www.rfc-editor.org/info/rfc3393>.

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, DOI 10.17487/RFC6146, April 2011, <https://www.rfc-editor.org/info/rfc6146>.

[RFC7679] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Delay Metric for IP Performance Metrics (IPPM)", STD 81, RFC 7679, DOI 10.17487/RFC7679, January 2016, <https://www.rfc-editor.org/info/rfc7679>.

[RFC7680] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Loss Metric for IP Performance Metrics (IPPM)", STD 82, RFC 7680, DOI 10.17487/RFC7680, January 2016, <https://www.rfc-editor.org/info/rfc7680>.

[TS.23.501-3GPP] 3rd Generation Partnership Project (3GPP), "3GPP TS 23.501 (V16.2.0): System Architecture for the 5G System (5GS); Stage 2 (Release 16)", September 2019, <http://www.3gpp.org/ftp//Specs/archive/23_series/23.501/23501-g20.zip>.

Authors' Addresses

Reza Rokui
Nokia
Canada

Email: reza.rokui@nokia.com

Shunsuke Homma
NTT
Japan

Email: shunsuke.homma.fp@hco.ntt.co.jp

Kiran Makhijani
Futurewei
USA

Email: kiranm@futurewei.com

Luis M. Contreras
Telefonica
Spain

Email: luismiguel.contrerasmurillo@telefonica.com