teas
Internet-Draft
Intended status: Informational
Expires: May 5, 2020

R. Rokui
Nokia
S. Homma
NTT
K. Makhijani
Futurewei

November 2, 2019

IETF Definition of Transport Slice
draft-nsdt-teas-transport-slice-definition-00

Abstract

This document describes the definition of transport slice in IETF and considerations on implementation (realization) of transport slice. This work is part of work on TEAS WG network slicing Design Team.

Table of Contents

1. Introduction

Network slicing is an approach to provide separate virtual networks depending on requirements of each service. Network slicing receives attention due to factors such as diversity of services and devices, and it is also a fundamental concept of the 5G for applying networks to such various types of requirements (Ref [TS.23.501-3GPP]). However there are other applications which might benefit from network slicing. Following is a list of other applicaitons:

- 5G network slicig
- Wholesale business VPN
- Network sharing among operators
- NVVI connectivity (DCI)

A network slice is composed of several parts such as endpoints and transport connecctivity between them. However, there is no concrete definition of network slices established on transport and how to realize them.

This document describes the definition of transport slice from IETF aspect and considerations on their realization as well.

2. High level architecture of end-to-end network slicing

To demonstrate IETF definition of both E2E network slice and transport slice, consider a typical network shown in Figure 1 where the network operator-Y has various networks of various technologies (e.g. IP, MPLS, Optics, PON, Microwave, 5G RAN, 5G Core etc.). Each network contains one or more nodes of (aka physical or virtual network functions, PNFs or VNFs), which have various capabilities and technologies such as:

- Routers
- Switches
- Application servers
- Firewalls
- 4G/5G RAN nodes
- 4G/5G Core nodes
- etc.

Each node (aka endpoint) in the network might support various technologies such as IP, MPLS, Microwave, 5G RAN, 5G Core etc. For example,

- Network-1 might contains multiple 5G gNBs connected to a few routers as Cell Site Gateways (CSG).
- Network-3 might have one or more L2/L3 routers and switches which are running on top on Optical network.
- Network-2 might have a few nodes of 5G RAN which are connected by PON.

```
         <========================= E2E NS =====================>

         <-OS1-> <-TS1-> <-TS2-> <-OS2->    ...   <-TSn-> <-OSm->

             .--.             .--.                    .--.
            (    )--.        (    )--.               (    )--.
           .'        '      .'        '             .'        '
   [EU-x] (  Network-1  )  (  Network-2  )   ...   (  Network-p ) [EU-y]
           `-----------'    `-----------'           `----------'

   Legends
     E2E NS: End-2-end network slice
     TSy: Transport Slice y
     OSx: Other Slice x
     EU-x: End User-x
```

Figure 1: E2E network slice

To further clarify the concept of the E2E network slice, consider the network operator-Y has various customers (tenants). One of its customers, needs to have a separate independent E2E logical network for specific service (e.g. CCTV, autonomous driving, HD map etc) for specific SLA requirement (e.g. high secure connection with Latency less than 5ms) from End User-x (EU-x) from one side of the network to End User-y (EU-y) to the other side. This E2E logical network is call an "E2E network slice". A typical example of EU-x in 5G is the User equipment such as infotainment unit in the car, CCTV, Car for autonomous driving etc. and a typical example of EU-y in 5G is 5G application server, IMS etc.

As shown in Figure 1 we use the term "E2E network slice" to show this logical i ndependent network from EU-X to EU-Y. It is important to consider that an "E2E network slice" is associated to a customer (tenant) and a service type (e.g. CCTV, autonomous driving etc.). Also there is only one E2E context between EU-x and EU-y. Anything else is not E2E.

For example, customer "City of NY" would like to connect all its CCTV cameras for entire city together. To do so, it asks Operator-Y who has coverage in NY to create a new separate independent logical network with SLA requirement of B/W greater than 10Mbps. In this case, a single E2E network slice (with NS ID 10) will be created by Operator-Y for Customer "City of NY", service type of CCT and SLA of 10Mbps.

It is also possible that customer and service type associate to an E2E network slice to be a wild card. For instance, in above example,

the E2E network slice 10, can be associated not only to service type CCTV but another service "Public Safety", i.e. NS ID 10 is used for two services for City of NY.

3. IETF Definition of Transport slice

The IETF definition of a transport slice is as follows:

A Transport Slice is an abstract network topology connecting different endpoints with appropriate isolation and specific Service Level Agreement (SLA) described in terms of shared or dedicated network resources, level of isolation etc.

In other words, a transport slice is a group of connections which connecting various endpointss in the network to achieve specific SLA for a customer as shown in Figure 2. Examples of the endpoints are any physical or virtual network functions (PNF/VNF) or any network services.

```
         <-------  Transport slice -------->

             .--.               .--.
   [EP11]   (    )- .          (    )- .    [EP21]
           .'         '       .'         '
   [EP12] (  Network-1 ) ... (  Network-p ) [EP22]
    :      `-----------'      `-----------'   :
   [EP1m]                                   [EP2n]

   Legend
     EP: Endpoint
```

Figure 2: Transport slice

Referring to Figure 1, when operator-Y would like to create a specific E2E network slice, it should create one or more of two types of artefacts:

o Transport slice (aka Transport sub-slices or Transport sub-nets)

o Other slice (aka Other sub-slices or other sub-nets)

As shown in Figure 1, an E2E network slice might have one or more of "Transport Slices" and one or more of "Other Slices" of any combinations. One of the critical parts of an E2E network slice is "Transport Slices" which provides various connections with certain SLA between various nodes (aka endpoints).

"The Other Slices" is out-of-scope of current work but in summary they contain various context or personality in the network to support a specific e2e network slice, i.e. The "Other Slices" are referred to as slices created by networks or components where IETF protocols do not strictly apply and operator can choose any method for defining them. For instance, in 5G, the prime example of these slices are:

- o 5G RAN slice (aka RAN sub-slice, RAN sub-net or RAN-NNSI): Contains the context or personality on various 5G RAN network functions (e.g. gNB, eNB, CU, DU etc) in support of specific e2e network slice with certain SLA

- o 5G Core slice (aka Core sub-slice, Core sub-net or Core-NNSI): Contains the context or personality on various 5G Core network functions (e.g UPF, SMF, AMF, etc) in support of specific e2e network slice with certain SLA

Figure 2 demonstrates the definition of a Transport Slice where a single Transport slice provides connectivity between "m" endpoints on left hand side to "n" endpoints on right hand side with specific characteristic for Service Level Agreement (SLA).

Each transport slice has main characteristics:

- o Transport slice definition: Technology agnostic to address a set of connections between various endpoints with certain SLA

- o Transport slice Implementation (aka realization): In addition to its definition, a Transport Slice has an implementation (aka realization) in the operator's network. Unlike transport slice definition, its implementation (aka realization) might be technology specific.

A few examples below demonstrate the idea of transport slice in various scenarios.

3.1. Scenario-1

Figure 3 depicts an example of transport slice connecting two 5G RAN nodes (gNB) to three 5G Core user plan function nodes (UPF). In this case a transport slice 20 is created with SLA of latency 10 [msec] or better between 5G endpoints gNBs and UPFs.

```
                    <--- Transport slice(TS ID:20) --->
                         with SLA of latency, less
                         than 10[msec]

                        .--.              .--.
              [gNB1]   (    )--.         (    )--.    [UPF1]
                      .'Network '      .' Network '
                     (  Midhaul  )    (   Backhaul )  [UPF2]
              [gNB2]  `----------'     `-----------'
                                                      [UPF3]
```

Figure 3: Example of Transport Slice 20 connecting gNBs to 5G Core UPF

3.2. Scenario-2

Figure 4 depicts another example where transport slice 30 is created to connect router ER1 to two firewall endpoints with SLA of 10 [Mbps] or higher bandwidth.

```
                   <--- Transport slice(TS ID:30) --->
                          with SLA B/W 5Mbps

                                .----.
                               (      )----.
                          .---'             '----.    [FW1]
               [ER1]     (          Network       )
                         `-----------------------'    [FW2]

               Legends
                 ER: Edge Router
                 FW: Firewall
```

Figure 4: Example of Transport Slice connecting Router to two firewalls

3.3. Scenario-3

Another example of transport slice is SFC case as shown in Figure 5 and Figure 6 which depict an example with SF1 and SF2 (e.g. DPI, Firewall, WAF, video optimizer, content cache server, NAT/CGN, Load balancer) and the transport slice between ER1 and ER2 traverses these SFs. There are two approaches:

o Approach-1 shown in Figure 5 where Transport slice 40 chains router ER1, SF1, SF2, and router ER2. Transport slice 40 needs lower than 30 ms delay. However, endpoints SF1 and SF2 are implicitly identified during the transport slice implementation. In this case, a single transport slice is created between ER1 and ER2.

o Approach-2 shown in Figure 6 where the transport slice 40 can be broken into transport slices 41, 42, 43. In this case SF1 and SF2 are explicityly identified and as a results three transport slides between following endpoints will be realized:

   * Between endpoints ER1 and SF1

   * Between endpoints SF1 and SF2

   * Between endpoints SF2 and ER2

```
                    <--- Transport slice(TS ID:40) --->
                          with SLA of latency 30ms

                         +-----+
                         | SF1 |
                         + *** +     .----.
                           * *      (      )--.
                           * *     (           )
                           * *  --' Network    '--.
               [ER1]******** *********** *************[ER2]
                         `-------------*-*--------'
                                       * *
                                     + *** +
                                     | SF2 |
                                     +-----+
```

Figure 5: Approach-1: Example of Transport Slice connecting Edge Routers ER1 and ER2 with SFC

```
                  <-- Transport slice (TS ID:40) -->
                        with SLA latency 30ms
                  <- TS41 -> <- TS42 ->  <- TS43 -->
                     SLA        SLA         SLA
                     5[ms]      20[ms]      5[ms]

                        +-----+
                        | SF1 |
                        + *** +     .----.
                          * *      (      )--.
                          * *     (           )
                          * *  --' Network    '--.
               [ER1]******** *********** ************[ER2]
                        `------------*-*--------'
                                     * *
                                   + *** +
                                   | SF2 |
                                   +-----+
```

Figure 6: Approach-2: Example of Transport Slice connecting Edge Routers ER1 and ER2 with SFC

4. Implementation (aka Realization) of Transport slice

In addition to its definition, a Transport Slice has another characteristic which is its implementation (aka realization) in the operator's network. Unlike transport slice definition, which is technology agnostics, its implementation (aka realization) is technology specific. To clarify the concept of transport slice implementation, in following section the implementation of scenarios described above will be described.

4.1. Implementation of Scenario-1

Figure 7 depicts the implementation (realization) of the transport slice 20 of Figure 3. Operator's transport slice controller receives an abstract API to create a transport slice between 5G endpoints gNB1 and gNB2 to 5G Core endpoints UPF1, UPF2 and UPF3 with SLA of 10 [ms].

Since in most cases neither 5G RAN endpoints nor the 5G Core endpoints can support any IP/MPLS/Optics services, the endpoints to implement the transport slice 20 will not be the endpoints passed in. This is one of the most important aspects to consider when implementing the transport slices.

As shown in Figure 7, the implementation of transport slice 20 required the transport slice controller to find out the "best" endpoints which support the realization of transport slice 20 in the network, i.e. endpoints ER1, ER2 and ER3. After that, the implementation of the transport slice 20 will be initiated by creation of various services/tunnels/paths between edge routers ER1, ER2 and ER3. The type of Services/Tunnels/Paths depends on the supported technologies of endpoints ER1, ER2 and ER3.

In this scenario, the end points of transport slice implementation are not those endpoints passed in, i.e.

o Definition of transport slice is between gNB1 and gNB2 to UPF1, UPF2, and UPF3

o Implementation of transport slice is between edge routers ER1, ER2 and ER3

```
                     | Create Transport slice 20 between gNB1 & gNB2
                     | to UPF1 & UPF2 & UPF3  with SLA  latency
                     | 10 ms or better
                     v
    +--------------------------------------+
    |Operator-Y Transport Slice Controller|
    +--------------------------------------+
                     | Implement (aka Realize) transport slice 20
                     | between ER1, ER2 and ER3 with SLA latency
                     | 10 ms  or  better
                     v                          +-----+
                   .-----.                      +UPF1|
   [gNB1] +-----+    (      )---.              /+--=-+
        \ |     |===========================[ER2]+
         \| ER1|  (      Network       )        +-----+
         /|     |===========================[ER3]+-|UPF2|
        / +-----+    `----------------'        +  +-----+
   [gNB2]                                        \
                                                  \+-----+
                                                    +UPF3|
                                                    +-----+
   Legends
     === : Tunnels & Services
     ER: Edge Router
```

Figure 7: Implementation (aka Realization) of Transport slice 20 of Figure-3

4.2. Implementation of Scenario-2

Figure 8 depicts the realization of transport slice 30 of Figure-4. Operator's Transport Slice Controller receives a request to create a transport slice between network functions R1 and firewalls FW1 and FW2 with SLA of 5 [Mbps]. Depends on the underlying network topology, Operator's transport slice controller will implement (aka realize) the transport slice. For example, if both network functions (i.e. R1, FW1, FW2) and network supports segment routing, two Tunnels/Services of type SR can be created (or used) in the network between R1, FW1 and FW2 to realise the transport slice 30. However, if the network just supports RSVP, two tunnels/services of type RSVP can be used to realize this transport slice.

Note that since the network functions ER1, FW1 and FW2 support segment routing, the endpoints of the tunnels in this example are those endpointss passed in, i.e. the endpoints of the both transport slice definiton and its implementation are R1, FW1 and FW2:

- o Definition of transport slice is between network functions R1 to FW1 and FW2

- o Implementation of transport slice is between network functions R1 to FW1 and FW2

We will see in next example that in some scenarios this is not the case and the endpoints of Transport Slice definition might be different from endpoints of its implementation (aka realization of transport slices).

It is very clear that regardless of how transport slice is realized in the network (i.e. using tunnels of type RSVP or SR), the definition of transport slice 30 does not change at all but rather its implementation.

```
                         | Create Transport slice 30 between
                         | ER1 and FW1 and FW2 with SLA 5 Mbps
                         v
          +----------------------------------------+
          | Operator-Y Transport Slice Controller |
          +----------------------------------------+
                         | Implement (aka Realize) transport
                         | slice 30 between R1 and FW1 & FW2
                         | with SLA 5  Mbps
                         v
                       .-----.
       +-----+      .----(       )----.
       |     |=============================[FW1]
       | ER |    (      Network       )
       |     |=============================[FW2]
       +-----+      `----------------'

       Legends
         === : Tunnels & Services of type SR
               or RSVP with SLA 5 Mbps
```

Figure 8: Implementation (aka Realization) of Transport slice 30 of Figure-4

4.3. Implementation of Scenario-3

Figure 9 depicts the implementation (realization) of the transport slice 40 of Figure 5 where a transport slice needed between network functions R1 and R2 across SF1 and SF2. However, the location of SF1 and SF2 are decided internally with a logic in Transport Slice Controller. For example, when SLA requires the high secure transport slice between ER1 and ER2 which in turn results on adding SF2 and SF2 to the implementation of transport slice 40 implicitly by transport slice controller.

Figure 10 shows the implementation (realization) of the transport slice 40 of Figure 6. In this case the location of SF1 and SF2 has been explicitly decided by higher level logic. In this case three transport slices 41, 42 and 43 will be created separately and eventually bind together to form a single transport slice 40 to meet the SLA that delay is lower than 30 ms.

```
                           | Create transport slice between ER1
                           | and ER2 with latency 30 [msec]
                           v
                  +-------------------+             +------------+
                  | Transport slice  |<------------->|    SF      |
                  |   controller     |             |  Manager   |
                  +-------------------+             +------------+
                           | Implementation transport slice 40
                           | between ER1 & ER2 traversing SF1 and SF2
                           | with SLA of latency 30 [msec]
                           V
        <------------------ TS 40 -------------------->
                +-----+
                | SF1 |
                + === +    .----.
                  # #    (      )--.
                  # #  (           )
                  # #  --' Network    '--.
    [ER1]=========   =========  ===================[ER2]
              `-------------- # # --------'
                            # #
                          + == +
                          | SF2 |
                          +-----+

    Legends
      ===== : Tunnels & Services
```

Figure 9: Implementation (aka Realization) of Transport slice 40 of Figure-5

```
                          | Requirements on communication
                          | between ER1 and ER2
                          v
                 +-----------------+            +-----------+
                 |  Orchestrator   | <---------> | SF Manager|
                 +-----------------+            +-----------+
                          | Create transport slice between ER1 and SF1,
                          |        with latency 5 [msec]
                          | Create transport slice between SF1 and SF2,
                          |        with latency 20 [msec]
                          | Create transport slice between SF2 and ER2,
                          |        with latency 5 [msec]
                          v
                +-------------------+
                | Transport slice   |
                | manager/controller|
                +-------------------+
                   |      |     |         Realize TS 41 between ER and
                   |      |     |            SF1 with latency 5 msec
            +------+      |     +-------+ Realize TS 42 between SF1 and
            |             |             |    SF2 with latency  20 msec
            |             |             | Realize TS 43 between SF2 and
            |             |             |    ER2 with latency 5 msec
            v             v             v
       <--TS 41-->   <--TS 42-->   <--TS 43 -->
      <---------------- TS 40 ---------------->
                +-----+
              #=| SF1 |
              # +-----+ .----.
              #     #  (      )--.
              # .---#' Network    '--.
    [ER1]=====#(    #==========#       )#=====[ER2]
                `-------------#------' #
                              #        #
                            +-----+    #
                            | SF2 |====#
                            +-----+

    Legend
    === : Tunnels & Services
```

Figure 10: Implementation (aka Realization) of Transport slice 40 of Figure-6

5. Definition of SLA and Isolation levels

TBD

6. Informative References

[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