### Service Function Chaining with Stitched Tunnels
### draft-hegde-spring-sfc-stitched-tunnel-00

Abstract

   The term "service function chaining" is used to describe the
   definition and instantiation of an ordered list of instances of such
   service functions, and the subsequent "steering" of traffic flows
   through those service functions.This document describes transport
   mechanisms that use stitched tunnels to achieve end-to-end service
   chaining.  This document specifies two different types of transport
   encodings, one based on SR-MPLS and another based on IP tunnels.

Table of Contents

## 1.  Introduction

Service function chaining requires an ordered set of service
functions to be executed on the traffic.  The set of service
functions could be virtualized functions or physical appliance based
functions or in some cases a mix of both.  Traffic needs to be
steered to the set of service functions in an ordered manner.  Based
on the the type of traffic, the order of the service functions and
the type of service functions that need to be executed may differ.
Service functions may not be aware of the transport encodings and
hence the transport encodings may have to be removed before passing
the packet(s) to the service functions.

Section Section 3 describes basic concepts of using stitched tunnels
to steer the traffic through the service functions.
Section Section 4 describes the transport encodings using SR-MPLS
based tunnels [RFC8402].  Section Section 5 describes the transport
encodings using IP Tunnels [I-D.saad-teas-rsvpte-ip-tunnels].

This document uses terminology defined in [RFC7665].

## 2.  Terminology

 This document uses the following terminology

  o  Service Function Orchestrator (SFO): Service Function Orchestrator is
     responsible for defining the Service Function Chains and the traffic
     types where the particular service function chains are applicable.
     SFO is also responsible for making all the required configurations  to
     realize the Service Chain.


                      Figure 1: Terminology

**3**.  **Concept of Stitched Tunnel**

   Consider a data center environment with virtualized service functions
   as shown in (See Figure 2).  DCGW1 and DCGW2 are Data center Gateway
   devices that connect the data center to the WAN.  SP1 and SP2 are
   spine devices.  TOR1, TOR2 and TOR3 are Top-Of-the-Rack switches that
   connect to the servers. a1,b1, c1 etc are VLAN interfaces that
   connect to the servers.  A1, B1, C1 etc are the service function
   instances deployed in the servers.

```
+----------------------------------------------------------------+
|               +------+                    +------+             |
|               | DCGW1|                    | DCGW2|             |
|               +------+                    +------+             |
|                  |                           |                |
|               +-----+                     +-----+             |
|               | SP1 |----+ +---------| SP2 |                  |
|               +-----+    | |         +-----+                  |
|                  | |     | |          | |                     |
|        +--------+ |      | |          | +--------+            |
|        | +--------|------|-|-----------+         |            |
|        | |        +------|-|-------------------+ |            |
|        | |               | |                   | |            |
|      +------+          +------+              +------+         |
|      | TOR1 |          | TOR2 |              | TOR3 |         |
|      +------+          +------+              +------+         |
|       / | \             / | \                / | \           |
|      a1 b1 c1          a2 b2 c2             a3 b3 c3          |
|      +---------+       +----------+         +----------+      |
|      | A1 B1 C1|       | A2 B2 C2 |         | A3 B3 C3 |      |
|      +---------+       +----------+         +----------+      |
|                                                                |
+----------------------------------------------------------------+
```
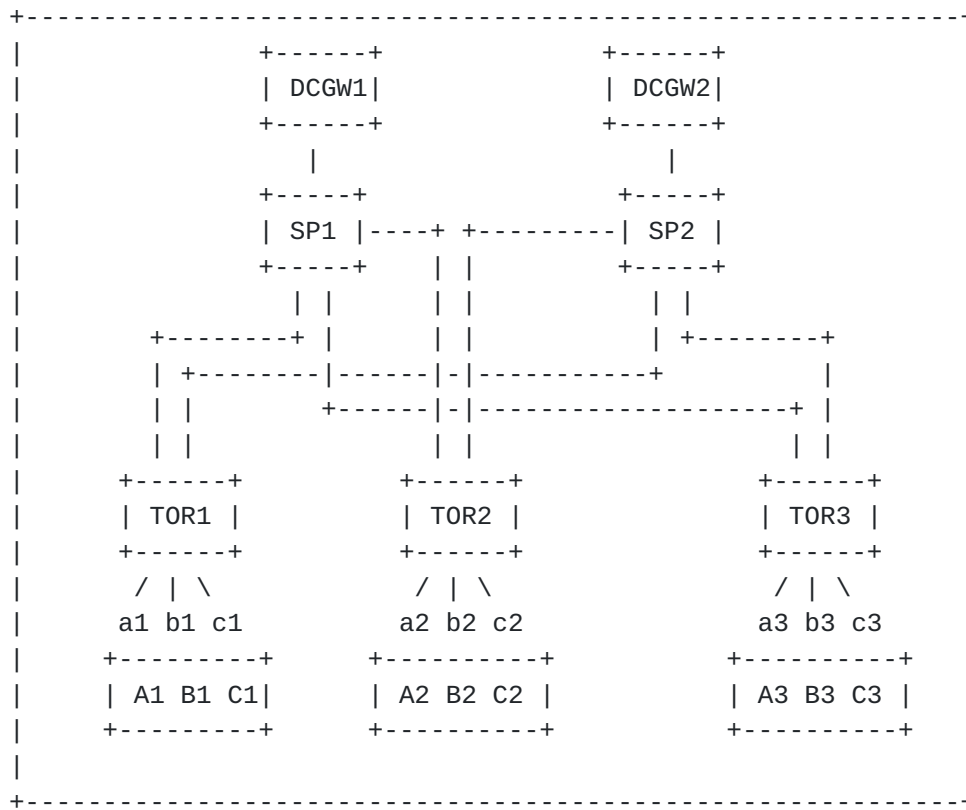
                 Figure 2: Virtualized Service Functions

   Let us assume certain traffic that originates at A1 and destined to Z
   (destination attached to remote WAN device not shown in the diagram),
   needs to visit service functions deployed at A2 and A3.  Traffic that
   originates at B1 and destined to Z needs to visit service functions
   B2 and B3.  The stepwise procedures described below provide the basic
   constructs involved in creating a stitched tunnel solution for
   steering traffic through the service functions.

   Building Tunnels: The traffic that originates at A1 and destined to Z
   needs to visit A2 and A3.  To achieve this, two transport tunnels are
   built.  The first tunnel Tunnel1 from TOR1 to TOR2 exiting on a2
   VLAN.  As the tunnels ends at the TOR2, the transport encapsulation
   is removed and the original packet is passed to A2.  Another tunnel
   Tunnel2 is built from TOR2 to TOR3 exiting on a3 VLAN.

   Similarly for the traffic originating at B1 and destined to Z, two
   more tunnels are built.  Tunnel3 from TOR1 to TOR2 exiting on b2 VLAN
   and Tunnel4 TOR2 to TOR3 exiting from b3 interface

   These separate tunnels will be stitched together using traffic
   classification rules as described below

   Traffic classification and steering: The traffic is classified at the
   TOR1 based on either 5-tuple or based on other fields in in the
   packet such as DSCP bits.  In the above example, the traffic steering
   may be based on source and destination address.  Traffic from source
   A1 and destination Z will be steered into Tunnel1.  On TOR2, traffic
   steering policies for source A1 and destination Z will steer the
   traffic into Tunnel2.

   Service Function Orchestrator (SFO): SFO is responsible for deploying
   the service function instances.  SFO builds tunnels from TOR1 to TOR2
   and TOR2 to TOR3 etc as required by the service function chain.  SFO
   is also responsible for defining the traffic steering policies and
   configuring them on appropriate traffic entry points.  For
   simplicity, the example in this section shows all virtualized service
   functions, but the concepts equally apply to service functions
   deployed on physical devices or a mix of physical and virtualized
   service functions.

   Bidirectionality: Certain service functions require the order of
   service chaining to be preserved for the return traffic.  In the
   stitched tunnel based solution, the transport encodings are
   completely independant and are not aware of the service functions.
   The SFO should ensure the traffic steering policies on traffic entry
   points to ensure correct order of service functions for the reverse
   traffic.  For example, the traffic with source Z and destination A1,
   should be steered to TOR3 , exiting on a3 VLAN and on TOR3, same
   policy should steer the traffic into a tunnel from TOR3 to TOR2
   exiting from VLAN a2.

   Looping service Functions: [RFC7665] describes possibility of looping
   service functions that require to be applied repeatedly.  In the
   solution based on stitched tunnels, this needs to be orchestrated by
   the SFO and traffic should be steered into the right tunnel to
   redirect to the service function that needs to be applied.  For
   example, if the service function at A2 need to be re-applied after
   servicing A3, the traffic steering policy at TOR3 should steer the
   traffic through a tunnel from TOR3 to TOR2 existing out of VLAN a2.

   Meta data handling: The trasnport encodings in stitched tunnel
   solution is completely independent of meta data handling.  There may
   be SFC encapsulations as described in [RFC8300] or other kinds of
   packet encapsulations.  Transport encodings will see these
   encapsulations as if it was original packet and hand it over to the
   service functions.

4.  **SR-MPLS based Tunnels**

   Segment Routing (SR) [RFC8402] is an architecture based on the source
   routing paradigm.  SR can be used with an MPLS or an IPv6 data plane
   to steer packets through an ordered list of instructions, called
   segments.  [I-D.ietf-spring-segment-routing-policy] describes a
   mechanism to use a stack of segments to steer the traffic along the
   explicit path.  The Tunnel1 and Tunnel2 described in Section 3 can be
   built using the segments.  For example, the Tunnel1 may be built from
   TOR1->SP1->TOR2->a2.  The path may be represented using Node-SID or
   Adj-SID as specified in [RFC8402].  At every segment end-point, the
   segment will be removed and traffic will be forwarded as per the next
   segment. on TOR2, the segment for a2 should be an Adj-SID.  In Data
   center networks that deploy IGPs, this could be an Adj-SID for the
   the passive IGP interface.  When the traffic hits TOR2, the passive
   Adj-SID for a2 will be removed and traffic will be sent on a2 V-LAN.
   As the transport encodings are completely removed from the packet
   before sending to the Service Function, there is no special handling
   required for SR-aware and SR-unaware service functions.

5.  **IP Tunnels**

   [I-D.saad-teas-rsvpte-ip-tunnels] is a solution that describes the
   use of RSVP (Resource Reservation Protocol) to establish Point-to-
   Point (P2P) Traffic Engineered IP (IP-TE) Label Switched Path (LSP)
   tunnel(s) for use in native IP forwarding networks.  The solution
   introduces one or more reserved local IP prefixes, referred to as
   Egress Address Blocks (EABs) per egress router that are dedicated for
   RSVP to establish IP-TE LSP(s) tunnels.  In
   [I-D.saad-teas-rsvpte-ip-tunnels], EABs are managed by the egress
   router.  To facilitate service chaining, EAB management would be the
   responsibility of the SFO along with the aforementioned flow steering
   mentioned above.  Further, in draft-saad-teas-rsvpte-ip-tunnels RSVP
   is responsible for path establishment from ingress router to egress
   router for a given IP-TE tunnel.  In this solution, the SFO would not
   only serve to calculate the explicit path of the IP-TE tunnel but
   also to program the per node forwarding state for each IP-TE tunnel.
   Extensions to the Path Computation Element Protocol (PCEP) as defined
   in [I-D.ietf-teas-pce-native-ip] and
   [I-D.ietf-pce-pcep-extension-native-ip]could be leveraged.  The
   Tunnel1 and Tunnel2 described in Section 3 can be built using per ToR
   per VLAN EAB IP-TE tunnels.  For example, the Tunnel1 may be built
   from TOR1->SP1->TOR2->a2.  The path may be represented using the EAB
   associated with A2 vlan a2.  At every IP-TE tunnel end-point, the IP
   tunnel Encapsulation header will be removed and traffic will be
   forwarded accordingly.

6.  Backward Compatibility

    TBD

7.  Security Considerations

    TBD

8.  IANA Considerations

    NA

9.  Acknowledgements

    TBD

10.  Contributors

11.  References

11.1.  Normative References

   [RFC7665]  Halpern, J., Ed. and C. Pignataro, Ed., "Service Function
              Chaining (SFC) Architecture", RFC 7665,
              DOI 10.17487/RFC7665, October 2015,
              <https://www.rfc-editor.org/info/rfc7665>.

   [RFC8402]  Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L.,
              Decraene, B., Litkowski, S., and R. Shakir, "Segment
              Routing Architecture", RFC 8402, DOI 10.17487/RFC8402,
              July 2018, <https://www.rfc-editor.org/info/rfc8402>.

11.2.  Informative References

   [I-D.ietf-pce-pcep-extension-native-ip]
              Wang, A., Khasanov, B., Fang, S., Tan, R., and C. Zhu,
              "PCEP Extension for Native IP Network", draft-ietf-pce-
              pcep-extension-native-ip-09 (work in progress), October
              2020.

   [I-D.ietf-spring-segment-routing-policy]
              Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and
              P. Mattes, "Segment Routing Policy Architecture", draft-
              ietf-spring-segment-routing-policy-09 (work in progress),
              November 2020.

   [I-D.ietf-teas-pce-native-ip]
              Wang, A., Khasanov, B., Zhao, Q., and H. Chen, "Path
              Computation Element (PCE) based Traffic Engineering (TE)
              in Native IP Networks", draft-ietf-teas-pce-native-ip-15
              (work in progress), December 2020.

   [I-D.saad-teas-rsvpte-ip-tunnels]
              Saad, T. and V. Beeram, "IP RSVP-TE: Extensions to RSVP
              for P2P IP-TE LSP Tunnels", draft-saad-teas-rsvpte-ip-
              tunnels-01 (work in progress), November 2019.

   [RFC7498]  Quinn, P., Ed. and T. Nadeau, Ed., "Problem Statement for
              Service Function Chaining", RFC 7498,
              DOI 10.17487/RFC7498, April 2015,
              <https://www.rfc-editor.org/info/rfc7498>.

   [RFC8300]  Quinn, P., Ed., Elzur, U., Ed., and C. Pignataro, Ed.,
              "Network Service Header (NSH)", RFC 8300,
              DOI 10.17487/RFC8300, January 2018,
              <https://www.rfc-editor.org/info/rfc8300>.

Authors' Addresses

   Shraddha Hegde
   Juniper Networks Inc.
   Exora Business Park
   Bangalore, KA  560103
   India

   Email: shraddha@juniper.net


   Colby Barth
   Juniper Networks Inc.

   Email: cbarth@juniper.net