SPRING                                                        S. Agrawal
Internet-Draft                                                    Z. Ali
Intended status: Standards Track                           C. Filsfils
Expires: April 25, 2019                                  Cisco Systems
                                                             D. Voyer
                                                          Bell Canada
                                                             G. Dawra
                                                             LinkedIn
                                                                Z. Li
                                                  Huawei Technologies
                                                     October 22, 2018

### SRv6 and MPLS interworking
### draft-agrawal-spring-srv6-mpls-interworking-00

Abstract

   This document describes SRv6 and MPLS/SR-MPLS interworking and co-
   existence procedures.

Requirements Language

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in RFC 2119 [RFC2119].

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 April 25, 2019.

Copyright Notice

Table of Contents

# 1.  Introduction

   Many of the deployments require SRv6 insertion in the brownfield
   networks.  The incremental deployment of SRv6 into existing networks
   require SRv6 to interwork and co-exist with SR-MPLS/ MPLS.

   There are various SRv6 and SR-MPLS/ MPLS interworking scenarios
   possible.  They can be classified into the following four categories.

   o  SRv6 over SR MPLS/ MPLS (6oM)

   o  SR MPLS/ MPLS over SRv6 (Mo6)

o  SRv6 to SR-MPLS/ MPLS (6toM)

o  SR-MPLS/ MPLS to SRv6 (Mto6)

These scenarios cover various cascading of SRv6/ MPLS network, e.g.,
SR-MPLS/MPLS <-> SRv6 <-> SR-MPLS/MPLS <-> SRv6 <-> SR-MPLS/MPLS,
etc.  The draft addresses all these possible interworking scenarios.

In addition, the draft also addresses migration and coexistence of
the SRv6 and SR-MPLS/ MPLS.  Co-existence means a network that
supports both SRv6 and MPLS in a given domain.  This may be a
transient state when brownfield SR-MPLS/ MPLS network upgrades to
SRv6 (migration) or permanent state when some devices are not capable
of SRv6 but supports native IPv6 and SR-MPLS/ MPLS.

## 1.1.  Terminology

SID: A Segment Identifier which represents a specific segment in
segment routing domain.  The SID type used in this document is IPv6
address (also referenced as SRv6 Segment or SRv6 SID).

Node k has a classic IPv6 loopback address Ak::/128.

A SID at node k with locator block B and function F is represented by
B:k:F::

A SID list is represented as <S1, S2, S3> where S1 is the first SID
to visit, S2 is the second SID to visit and S3 is the last SID to
visit along the SR path.

(SA,DA) (S3, S2, S1; SL) represents an IPv6 packet with:

-IPv6 header with source address SA, destination addresses DA and SRH
as next-header

-SRH with SID list <S1, S2, S3> with SegmentsLeft = SL

Note the difference between the <> and () symbols: <S1, S2, S3>
represents a SID list where S1 is the first SID and S3 is the last
SID to traverse.  (S3, S2, S1; SL) represents the same SID list but
encoded in the SRH format where the rightmost SID in the SRH is the
first SID and the leftmost SID in the SRH is the last SID.  When
referring to an SR policy in a high-level use-case, it is simpler to
use the <S1, S2, S3> notation.  When referring to an illustration of
the detailed packet behavior, the (S3, S2, S1; SL) notation is more
convenient.

domain: Without loss of the generality, domain is assumed to be
instantiated by a single IGP instance or a network within IGP if
there is clear separation of data plane.

**2**.  **Interworking Procedures**

This documents refers to interworking as a stitching of SRv6 domain
and SR-MPLS/ MPLS domain.  Special stitching procedures are performed
on routers which acts as border between such domains.  Border routers
need to support both SRv6 and SR-MPLS/ MPLS.  Interworking is
applicable when SRv6 domains are deployed and need to interop with
existing SR-MPLS/ MPLS backbones or access networks.

This draft proposes two ways to stitch heterogeneous domains: a
controller based solution and a BGP signaling based approach.  The
PCE based solution is applicable to both best effort as well as
deployments where tight SLA guarantees are required (e.g., ODN like
deployments scenarios).  The BGP signaling covers the best effort
case.  Specifically, the draft proposes the following two ways to
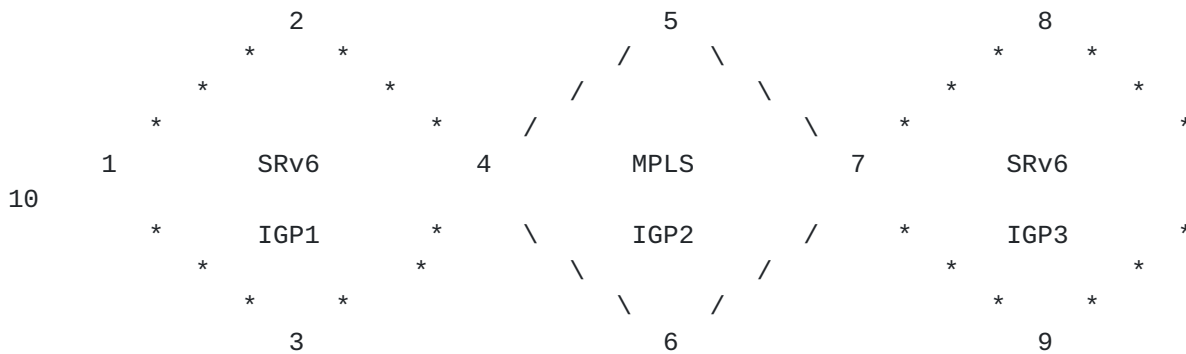stitch heterogeneous domains end to end:

o  Stitching using a Controller: An SDN based approach like Multi-
   Domain On Demand Nexthop (ODN) case for SLA contract end to end
   across heterogeneous domains.  Path Computation Element (PCE) can
   act like the controller.  These procedures can be used when
   overlay prefixes have SLA requirement signaled through a color
   community.  These procedures can also be used for the best effort
   services.

o  Stitching using BGP Inter-Domain Routing.  BGP 3107 procedures
   advertising PE locators/Loopbacks for best effort end to end
   connectivity.  These procedures are applicable in deployments
   where an SDN controller is not deployed.  These procedures can be
   used when overlay prefixes don't have SLA requirement

In summary the draft covers the following SRv6/ MPLS interworking
scenarios.
- Carrying SRv6 over SR-MPLS (controller stitches domains).
- Carrying SRv6 over SR-MPLS (BGP stitches domains).
- Carrying SR-MPLS over SRv6 (controller stitches domains).
- Carrying SR-MPLS over SRv6 (BGP stitches domains).
- SRv6 to SR-MPLS translation (controller stitches domains).
- SRv6 to SR-MPLS translation (BGP stitches domains).
- SR-MPLS to SRv6 translation (controller stitches domains).
- SR-MPLS to SRv6 translation (BGP stitches domains).
- Cascaded domains (controller stitches domains).
- Cascaded domains (BGP stitches domains).

While the number of interworking scenarios is large, the few building
blocks outlines in this draft address all of them.  For the same
reasons, without the loss of generality, the building blocks are
illustrated using the SRv6 over SR-MPLS example of Figure 1 but the
procedure equally applies to the other deployment scenarios.

```
                2                          5                          8
            *       *              /          \              *          *
         *               *          /              \            *              *
      *                     *       /                    \      *                      *
    1           SRv6            4            MPLS            7           SRv6
10
         *        IGP1        *      \        IGP2        /      *        IGP3          *
           *               *          \              /          *              *
             *       *                  \        /              *        *
                3                          6                          9
```

Example Network Scenario(6oM)

                              Figure 1

## 3.  Building blocks for domain stitching

## 3.1.  Stitching heterogenous domains using a Controller

This procedure provides a best-effort path as well as a path that
satisfies the Service Level Agreement (SLA), across multiple domains.
A PCE may act as an SDN controller.  In that case, based on the SLA,
the PCE computes and programs end to end path.  The PCE is also aware
of interworking requirement at border nodes, as each domain feeds
topological information to the controller through BGP LS feeds.
Intermediate domain of different data plane type is represented by
Binding SID (BSID) of head end type in SID list.  The intermediate
domain BSID is programmed at domain entry border node with SID list
through domain and exit node SID as last segment.  In summary, an
intermediate heterogenous domain is replaced by a BSID of the data
plane nature of headend.  The procedure work for all of deployment
model mentioned above.

## 3.1.1.  Illustration

The procedure is illustrated using the example of Figure 1.

When a service prefix (e.g., vpn or evpn) is received on head end
with SLA (color extended community), the head- end (Node 1) node
requests a PCE for a path to egress node that can satisfy the SLA.
This is because Node 1 does not know how to compute the traffic
engineered path through the multi-domain network to node 10.  Node 1

requests SR PCE to compute a path to node 10 providing optimization objective, constraints(eg: low latency).  The PCE computes low latency path via node 2, 5 and 8.  The PCE identifies the end-to-end path is not consistent data plane and kicks in interworking procedures at the border node(4).  It programs a policy at 4 that "Stitches" domains using SRv6 End.BM BSID.The PCE installs SRv6 End.BM BSID at node 4 with segments are node 5, 7.  SR PCE responds back to node 1 with SRv6 segments via node 2, End.BM at node 4, node 8 and node 10.

The data plan operations for the above-mentioned interworking example are described in the following:

o  Node 1 performs SRv6 function T.Encaps.Red with VPN service SID
   and SRv6 Policy (BLUE,10):
   Packet leaving node 1 IPv6 ((A:1::, B:2:E::) (B:10::DT4, B:8:E::,
   B:4:BM-BLUE-7:: ; SL=3))

o  Node 2 performs End function
   Packet leaving node 2 IPv6 ((A:1::, B:4:BM-BLUE-7::) (B:10::DT4,
   B:8:E::, B:4:BM-BLUE-7:: ; SL=2))

o  Node 4 performs End.BM function
   Packet leaving node 4 MPLS (16005,16007,2)((A:1::, B:8:E::)
   (B:10::DT4, B:8:E::, B:4:BM-BLUE-7-:: ; SL=1))

o  Node 7 performs a native ipv6 lookup on due PHP behavior for 16007
   Packet leaving node 7 IPv6 ((A:1::, B:8:E::) (B:10::DT4, B:8:E::,
   B:4:BM-BLUE-7:: ; SL=1))

o  Node 8 performs End(PSP) function
   Packet leaving node 8 IPv6 ((A:1::, B:10::DT4))

o  Node 10 performs End.DT function and lookups IP in vrf and send
   traffic to CE.

## 3.2.  Stitching heterogenous domains usin BGP inter domain routing

For providing services across domains, edge node locators need to be reachable.

-Locators are advertised by edge nodes in the BGP ipv6 unicast address family (AFI=2,Safi=1) to border nodes.  These locators are also advertised in its local IGP domain.

-On border nodes these prefixes are like any IPv6 global prefixes. These will be advertised in BGP IPv6 LU[AFI=2/SAFI=4] session using

3107 procedures in label core.  It could be summary prefix for all
locators in that domain.

-Remote domain border router advertising locator over SRv6 domain
need to attach SRv6 SID in prefix SID attribute.  SRv6 SID in this
case will be End function of advertising border node.

-Ingress node learns remote locators over BGP ipv6 address
family[AFI=2, SAFI=1].  These locators have prefix SID attribute
containing SRv6 SID.  This SRv6 SID is End function of advertising
border node and helps to tunnel traffic to border node in remote
domain.

-If locators are leaked into remote IGP and no tunneling of traffic
will be needed in remote domain.  Hence attaching SRv6 SID on remote
border nodes can be avoided.

These procedures work for any of deployment model mentioned above.
Below are some important aspects for Mo6, 6toM, Mto6
-Loopback address are advertised in BGP label unicast session to
border node when advertised from MPLS domain.  These are also
advertised in local IGP.
-Border nodes advertises prefix over SRv6 domain in BGP IPv4/IPv6
session.  It attaches prefix SID attribute with SRv6 SID.  This SRv6
SID maps to label received in prefix update.  -Remote border node
allocates local label to advertise prefix in MPLS domain to ingress
node.  This local label maps to received SRv6 SID in prefix sid
attribute of prefix.

### 3.3.  6toM and Mto6 considerations

For 6toM and Mto6 BGP inter domain or ODN multi domain stitching will
work if SRv6 edge nodes are capable of handling vpn/service label.
In 6toM scenario, ingress node should be able to encap vpn label and
then perform T.Encap function with SRv6 SID associated with prefix
nexthop.  In Mto6 case, traffic will be received with SRv6 SID and
vpn label below it on egress PE.  So egress SRv6 capable node should
be able to process vpn labels after decapsulating SRv6 SID and when
next header is 137 in IPv6 header.

Service information encoded by SRv6 PE will be in SRv6 Service SID
and MPLS PE will be vpn label/service label or just IP payload for
internet.  If SRv6 PE do not support vpn label, then we need some
special handling to translate SRv6 service SID to vpn label and vice
versa at border nodes.  This will be detailed in future versions

## 4.  FRR handling

Failure within domain are taken care by existing FRR(TILFA, rLFA, LFA etc) mechanisms.

Failure of border nodes are to be addressed in a future version of the document.

## 5.  Migration and co-existence

These procedures would be detailed in a future revision

## 6.  BGP based services interworking and migration

SRv6-based VPN (SRv6-VPN)/EVPN service information is encoded in SRv6 SIDs specifically END.DT*/END.DX*/END.DT2.  MPLS-based VPN service information is encoded in labels.  This requires special consideration during Migration and Interworking.  Will be discussed more detail in future versions

## 7.  IANA Considerations

None

## 8.  Security Considerations

## 9.  Acknowledgements

## 10.  Normative References

[RFC2119]  Bradner, S., "Key words for use in RFCs to Indicate
           Requirement Levels", BCP 14, RFC 2119,
           DOI 10.17487/RFC2119, March 1997,
           <https://www.rfc-editor.org/info/rfc2119>.

Authors' Addresses

Swadesh Agrawal
Cisco Systems

Email: swaagraw@cisco.com


Zafar ALI
Cisco Systems

Email: zali@cisco.com

Clarence Filsfils
Cisco Systems

Email: cfilsfil@cisco.com


Daniel Voyer
Bell Canada
Canada

Email: daniel.voyer@bell.ca


Gaurav dawra
LinkedIn
USA

Email: gdawra.ietf@gmail.com


Zhenbin Li
Huawei Technologies
China

Email: lizhenbin@huawei.com