BGP Working Group Y. Liu
Internet-Draft B. Song
Intended status: Standards Track ZTE Corporation
Expires: July 11, 2021 January 7, 2021

BGP Extensions for Services in SRv6 and MPLS Coexisting Network
draft-ls-bess-srv6-mpls-coexisting-vpn-01

Abstract

This document proposes a method to achieve VPN/EVPN in a network where SRv6 and SR-MPLS/MPLS coexist, including extensions of BGP.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The incremental deployment of SRv6 into existing networks require SRv6 to interwork and co-exist with SR-MPLS/MPLS.

Currently [I-D.agrawal-spring-srv6-mpls-interworking] and [I-D.pzm-bess-spring-interdomain-vpn] discuss about the SRv6 and MPLS interworking method.

In the progress of upgrading some network, some of the legacy devices that support only MPLS/SR-MPLS will coexist with the new devices capable SRv6 for a long time. The co-existence scenario also need to be further addressed.

This document proposes a method to achieve VPN/EVPN in a network where SRv6 and SR-MPLS/MPLS coexist, including extensions of BGP.

2. the Co-existence Scenario

```
                  +----R1----R2----+
                  |                | +----+CE21
                  |                | |
        CE1+----+PE1             PE2+
                  |                | |
                  |                | +----+CE22
                  +----R3----R4----+
```

Figure 1: Reference Topology 1

```
                    +----R1----R2----+
                    |                |
          CE1+----+PE1               |
                                   PE2+----+CE2
          CE3+----+PE3               |
                    |                |
                    +----R3----R4----+
```

Figure 2: Reference Topology 2

As shown in Figure 1 and Figure 2, R3 and R4 are capable of SRv6, R1 and R2 are legacy devices which only support SR-MPLS/MPLS.

In Figure 1, PE2 is connected to different services with different SLA requirements. Different SLA requirements may correspond to different forwarding paths, these paths may be SRv6 capable, or may pass through the devices that only support SR-MPLS/MPLS.

In Figure 2, to reach for the same service, the underlay path from PE1 to PE2 support SRv6 forwarding, while the path from PE3 to PE2 passes through the devices that only support SR-MPLS/MPLS.

Existing solutions include the following:

1)The egress PE allocates the MPLS label and SRv6 SID for the same service and advertise them separately through different routes, which have different priorities, the ingress PE then selects the route of higher priority.

For example, in Figure 1, an end-to-end VPN IPv4 BGP peer relationship and a IPv6 BGP peer relationship are established between PE1 and PE2.

After PE2 receives a VPN route from its VPN instance, PE2 advertises a copy of this route to the VPN IPv4 BGP peer and applies for an MPLS label. PE2 then advertises another copy to the VPN IPv6 BGP peer, with the route carrying a SRv6 SID. PE1 receives two VPN routes with the same prefix, one with an IPv4 next hop and the other with an IPv6 next hop. The route with the IPv4 next hop recurses to the MPLS tunnel, and the route with the IPv6 next hop recurses to the SRv6 tunnel. If routes with IPv4 next hops are of higher priority, the MPLS tunnel is chosen, otherwise the traffic reaches the PE2 through the SRv6 tunnel.

The disadvantage of this method is that only one route can take effect at the same time and the method is not flexible enough. In figure 1, if the best path from CE1 to CE21 is a MPLS tunnel while

the expected path from CE1 to CE22 is an SRv6 tunnel, this method cannot meet such requirements easily.

2)If the underlay path attribute corresponding to each service is predictable, the egress PE allocates either MPLS labels or SRv6 SIDs for each service based on the underlay path attribute. That is, the engress PE advertises only one kind of BGP route for a particular service prefix, either with MPLS labels or the SRv6 SIDs.

Once the path attribute of underlay is changed, for example, the device that only supports MPLS forwarding is upgraded to support SRv6, the configuration on PE should also be changed accordingly.

Based on the above scenarios, this document proposes a method:

The egress PE allocates MPLS label(s) and SRv6 SID(s) for the same service and signals them within the same BGP overlay service route.

After receiving the BGP advertisement, the ingress PE should add the prefix with the MPLS label and SRv6 SID information to the RIB.

When encapsulating packets, the ingress PE selects whether to use MPLS label or SRv6 SID according to the attribute of the underlay path.

If there is a route reflector in the network, it must support the extended BGP message too.

Currently, the MPLS-based VPN/EVPN service information is encoded in the MPLS Label field of the corresponding NLRI, and the SRv6-based VPN/EVPN service information is encoded as SRv6 service SIDs such as END.DT*/END.DX*/END.DT2 with BGP Prefix-SID attribute [RFC8669] extended to carry SRv6 service SIDs information [I-D.ietf-bess-srv6-services]

But how does the egress PE indicate in the BGP advertisement that a service supports both MPLS and SRv6 identification is not clearly described.

3. BGP extensions

3.1. Extended SRv6 Service TLVs

For the convenience of understanding and reading, the two methods of notifying SRv6 SID in [I-D.ietf-bess-srv6-services] are described briefly below.

In the first method, SRv6 Service SIDs are encoded as a whole in the SRv6 Services TLVs. In this case, the MPLS Label field(s) of the corresponding NLRI is set to Implicit NULL.

The second method is called Transposition Scheme of encoding, where the SRv6 SID Structure Sub-Sub-TLV describes the size of each part of the SRv6 SID and also indicates the offset of variable part along with its length in SRv6 SID value. The function and/or the argument part of the SRv6 SID is encoded in the MPLS Label field of the NLRI and the SID value in the SRv6 Services TLV carries only the locator part with the SRv6 SID Structure Sub-Sub-TLV.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   TLV Type    |         TLV Length            |M| RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//  SRv6 Service Sub-TLVs                                       //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: Extended SRv6 Service TLVs

This document introduces a M-flag in the RESERVED field of SRv6 Services TLVs as shown in figure 3, when set, it indicates that this service supports both MPLS label and SRv6 service SID identification.

If the advertisement message carries multiple SRv6 Service TLVs at the same time, for example, in the EVPN scenario, the M-flag of the these TLVs must be set to the same. If not, the advertisement MUST be discarded.

The MPLS-based VPN/EVPN service information is always encoded in the MPLS Label field of the NLRI.

If the Transposition Scheme of encoding is needed, the egress PE MUST allocate SRv6 service SIDs with the function and/or the argument part same as the MPLS VPN label.

Otherwise, SRv6 SIDs and MPLS labels can be of independent values, and SRv6 Service SIDs are encoded as a whole in the SRv6 Services TLVs.

The allocation of SRv6 SIDs and MPLS labels for VPN/EVPN on egress PEs is an implementation thing, and it is outside the scope of this document.

More processing details will be further discussed.

3.2. Dual-Stack VPN Capability

[RFC5492] defines the "Capabilities Optional Parameter". A BGP speaker can include a Capabilities Optional Parameter in a BGP OPEN message. The Capabilities Optional Parameter is a triple that includes a one-octet Capability Code, a one-octet Capability length, and a variable-length Capability Value.

This document defines a Capability Code for dual-stack VPN capability.

If a BGP speaker has not sent the dual-stack VPN capability in its BGP OPEN message on a particular BGP session, or if it has not received the dual-stack VPN capability in the BGP OPEN message from its peer on that BGP session, that BGP speaker MUST NOT send on that session any UPDATE message that includes the extended SRv6 service TLVs.

4. Illustration

The reference topology is show in Figure 2. PEs support both SRv6 and SR-MPLS capabilities.

Take IPv4 VPN as an example, PE2 assigns an MPLS label vpn2 and an SRv6 service SID(eg, END.DX4) sid2 for CE2, and the function part of the SID is vpn2.

Label field of IPv4-VPN NLRI is encoded as specified in [RFC8277] with the Label Value set to vpn2.

If Transposition Scheme of encoding is used, the locator part of the SRv6 Service SID is encoded in the SRv6 L3 Service TLV with the M-flag set to 1.

PE1 and PE3 learn through M-flag that CE2 has both MPLS and SRv6 identification, and obtain the corresponding MPLS label and SRv6 SID carried in the BGP update messages.

When a service prefix is received on PE1, by looking at the local forwarding table, PE1 finds that the service is related to an MPLS label and an SRv6 SID, and the corresponding path is a segment list consisting of SR-MPLS SIDs , such as <Label 1, Label 2>. PE1 then encapsulates the payload packet with an MPLS label stack <Label 1, Label 2, vpn2>.

Similarly, PE3 finds out that the underlay path is based on SRv6 such as <SID3, SID4>, then it encapsulates the payload packet in an outer IPv6 header with the segment list <SID3, SID4, sid2>.

5. Operation

If the underlay between PEs support IPv6 forwarding, including SRv6 and IPv6-MPLS, it is simple to implement dual-stack VPN using the above extensions. The PEs advertises the BGP route with an IPv6 next hop. Once whether the forwarding path is based on SRv6/IPv6 or IPv6-MPLS is decided, the subsequent processing is based on the existing BGP procedure.

Another scenario is that the legacy devices only support IPv4-based MPLS forwarding. In this case, the PEs should support IPv4/IPv6 dual stack and using an IPv4 next hop when advertising VPN routes. If the MPLS tunnel is chosen, the packet forwarding procedure is unchanged.

When providing SRv6-based best-effort connectivity to the egress PE, the ingress PE encapsulates the payload in an outer IPv6 header where the destination address is the SRv6 Service SID associated with the related BGP route update. The reachability of SRv6 service SID should be provided by other means, such as IGP or BGP advertisement and the forwarding is independent of the IPv4 next hop in the BGP VPN route.

If the BGP route received at an ingress PE is colored with an extended color community and is expected to be steered over a SRv6 Policy, there're two options:

a) Use color-only steering method regardless of the next hop of the BGP route and the endpoint of SR policy [I-D.ietf-spring-segment-routing-policy] section 8.8.1.

b) Steer on an SR Policy by the matching of the BGP route's next-hop N and color C with an SR Policy defined by the tuple endpoint N and color C. Then the endpoint of the SRv6 Policy should be configure as an IPv4 address.

Note that it is not stipulated in [I-D.ietf-spring-segment-routing-policy] that the endpoint of an SRv6 Policy must also be an IPv6 address.

6. Security Considerations

TBD

7. IANA Considerations

TBD

8. References

8.1. Normative References

[I-D.ietf-bess-srv6-services]
Dawra, G., Filsfils, C., Talaulikar, K., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "SRv6 BGP based Overlay services", draft-ietf-bess-srv6-services-05 (work in progress), November 2020.

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., Rosen, E., Jain, D., and S. Lin, "Advertising Segment Routing Policies in BGP", draft-ietf-idr-segment-routing-te-policy-11 (work in progress), November 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.

[RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <https://www.rfc-editor.org/info/rfc5492>.

[RFC8277] Rosen, E., "Using BGP to Bind MPLS Labels to Address Prefixes", RFC 8277, DOI 10.17487/RFC8277, October 2017, <https://www.rfc-editor.org/info/rfc8277>.

[RFC8669] Previdi, S., Filsfils, C., Lindem, A., Ed., Sreekantiah, A., and H. Gredler, "Segment Routing Prefix Segment Identifier Extensions for BGP", RFC 8669, DOI 10.17487/RFC8669, December 2019, <https://www.rfc-editor.org/info/rfc8669>.

8.2. Informative References

[I-D.agrawal-spring-srv6-mpls-interworking]
Agrawal, S., Ali, Z., Filsfils, C., Voyer, D., and Z. Li, "SRv6 and MPLS interworking", draft-agrawal-spring-srv6-mpls-interworking-03 (work in progress), August 2020.

[I-D.pzm-bess-spring-interdomain-vpn]
Zhang, Z., Peng, S., Mirsky, G., and Y. Wang, "SRv6 and MPLS interworking for VPN service", draft-pzm-bess-spring-interdomain-vpn-02 (work in progress), August 2020.

Authors' Addresses

Liu Yao
ZTE Corporation

Email: liu.yao71@zte.com.cn

Song Bing
ZTE Corporation

Email: song.bing@zte.com.cn