Workgroup: BESS
Internet-Draft:
draft-lp-bess-vpn-interworking-00
Published: 2 January 2022
Intended Status: Informational
Expires: 6 July 2022
Authors: Y. Liu S. Peng
ZTE ZTE

A Label/SID Allocation Method for VPN Interworking

Abstract

This document analyzes the SRv6-MPLS service interworking solution and offers an MPLS label or SRv6 SID allocation method for label/SID saving and better scalability purpose.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Inter-AS option B is one of the ways in which the PE routers from different ASes can exchange the MPLS VPN routes with each other as described in [RFC4364]. This method uses BGP to signal MPLS VPN labels between the AS boundary routers.

[I-D.ietf-bess-srv6-services] defines procedures and messages to carry SRv6 Service SIDs and form VPNs in SRv6. SRv6 Service SID refers to an SRv6 SID associated with one of the service-specific SRv6 Endpoint behaviors on the advertising PE router, and its usage is similar to MPLS VPN label to some extent.

In the progress of network upgrading, some of the legacy devices(e.g, PEs) that only support MPLS VPN will coexist with the new devices capable of SRv6 VPN for a long time. For service connectivity, services over the SRv6 PE need to interwork with that over MPLS PE.

A common method for service interworking is similar to inter-AS option B. ASBRs needs to translate between MPLS VPN labels and SRv6 service SIDs, i.e, when an ASBR receives an SRv6 VPN route from the egress PE, it needs to allocate an MPLS VPN label for it and advertise the corresponding MPLS VPN route to the ASBR in the MPLS domain.

This document analyzes the SRv6-MPLS service interworking solution and offers an MPLS label or SRv6 SID allocation method for label/SID saving and better scalability purpose.

2. Conventions used in this document

2.1. 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 [RFC2119].

2.2. Terminology and Acronyms

PE: Provider Edge

CE: ;Customer Edge

AS: Autonomous System

ASBR: Autonomous System Border Router

RD: Route Distinguisher

RR: Route Reflector

3. Service Interworking between SRv6 and MPLS

```
                                                +-----------------------
                                                |
 PE1-----------------------ASBR1-----------ASBR2---------------------


 |          MPLS            |               |          SRv6
 |<------------------------>|<------------->|<----------------------
 |          IBGP            |     EBGP      |          IBGP
```

Figure 1: Reference Multi-Domain Network Topology

As shown in Figure 1, PE1 is a legacy device that only supports MPLS-based services. PE2 and PE3 support SRv6-based services. ASBR2, as a node in the domain with new features, support both MPLS and SRv6.

On PE2, SRv6 service SID SID-21 is allocated, per VPN instance, for VPN prefix 1 and prefix 2 in VRF1, SRv6 service SID SID-22 is allocated, per VPN instance, for VPN prefix 2 in VRF2. On PE3, SRv6 service SID SID-31 is allocated, per VPN instance, for VPN prefix 2 in VRF1. Route distinguisher RD1 is configured for VRF1 and RD2 for VRF2.

The SRv6 VPN route carrying the corresponding service SID and RD with it propagates to ASBR2.

On ASBR2, the received SRv6 service SID needs to be replaced by a new MPLS label, then the new MPLS VPN label propagates to ASBR1.

On ASBR1, the received MPLS VPN label needs to be replaced by a new MPLS label, then the new MPLS VPN label propagates to PE1.

3.1. Label Allocation Method

On ASBR2, MPLS VPN label can be allocated per <RD, service prefix>.

However, there may be scalability issues. The 128-bit encoding space in SRv6 is much more larger compared with the 20-bit label space in MPLS, which means SRv6-based VPN allows massive number of sites(with different prefixes) to access. If the MPLS label is allocated per prefix in this case, the number of MPLS labels may be not enough. A label-saving allocation solution is needed for better scalability.

In order to save label resources, the MPLS label can be allocated per <SRv6-SID, next-hop> sets.

If this method is used, on ASBR2, MPLS lalel L21 is allocated for SID-21 on PE2, L22 for SID-22 on PE2 and L31 for SID-31 on PE3.

The disadvantage of this method is that once the next-hop is changed, the old label needs to be released and new label needs to be allocated, the corresponding MPLS VPN label should be withdrawed and re-advertised with the new one. For example, in the FRR case, the primary next-hop may be down and the converged path contains the original backup next-hop, that may cause the oscillation of label allocation, especially when there're large numbers of SRv6 service SID on the PE.

This document proposes that the MPLS label can be allocated per RD on ASBR. And it is required that the RD of different VRF instance MUST be different.

The ILM table is built based on the mapping relationship between the allocated label and RD, and it leads to the query of the RD context tables. The RD context tables are built based on the received VPN routes.

For one packet, the ASBR would query two tables to get the outgoing action, this behavior is similar to that in inter-AS option A in RFC4364, but unlike option A, there's no VRF instance configuration on the ASBR.

The per-RD allocation method applies for MPLS VPN inter-AS option B as well, i.e., similar operation can be done on ASBR1. The difference is that the encapsulating IPv6 action in the RD Entry Table should be replace by pushing the MPLS VPN label.

In addition, per-RD allocation can also apply for SRv6 VPN SID on ASBR2 when it received VPN routes from ASBR1, and the endpoint behaviour can still be END.DT4/6.

3.1.1. Detailed Example for Service Interworking

This section provides a detailed example for service interworking between SRv6 and MPLS based on the per-RD label allocation method. The reference topology is shown in Figure 1.

Control Plane:

1)Egress PE2 advertises a BGP SRv6 VPN route: RD:RD1, prefix:prefix2, next hop: lo-PE2, service SID: SID22.

Egress PE3 advertises a BGP SRv6 VPN route: RD:RD1, prefix:prefix2, next hop: lo-PE3, service SID: SID31.

PE3 acts as the backup of PE2 for prefix2, and the route from PE3 has lower priority.

These routes propagates to ASBR2.

2)The reachability of lo-PE2 and lo-PE3 is advertised by IGP within the domain.

3)ASBR2 learns <RD1,prefix1> with next hop lo-PE2 and SID22. ASBR2 learns <RD1,prefix1> with next hop lo-PE3 and SID31.

ASBR2 allocates MPLS label label-21 based on RD1.

The ILM forwarding table and RD context Table is shown in Figure 2.

ASBR2 uses the destination address of the packet to match the corresponding entry and determine the outgoing action.

The outgoing action shown in RD1 context table only provides an example when the SRv6 service is provided with best-effort connectivity, the ASBR2 encapsulates the payload in an outer IPv6 header where the destination address is the SRv6 Service SID provided by the egress PE. If SRv6 service is provided in conjunction with an underlay SLA from the egress PE, ASBR2 may encapsulate the payload packet in an outer IPv6 header with the segment list of SR policy associated with the related SLA along with the SRv6 Service SID associated with the route using the Segment Routing Header (SRH) [RFC8754]. The detailed encapsulate method is specified in [I-D.ietf-bess-srv6-services] and this document doesn't change that procedure.

ILM table

In label	Action
label-21	pop, lookup table.RD1

RD1 context table

Prefix	Outgoing Action
prefix2	Primary: encap IPv6 with DA=SID22, fwd to PE2 Backup: encap IPv6 with DA=SID31, fwd to PE3

Figure 2: Forwarding State on ASBR2

4) ASBR2 advertises an MPLS VPN label to ASBR1: RD:RDa, prefix:prefix2, next hop: ASBR2, VPN label: label-21.

5) ASBR1 learns <RDa,prefix2> with next hop ASBR2 and label-21.

ASBR1 changes the next-hop to itself and allocates a new label label-11, then ASBR1 advertises a BGP MPLS VPN route: RD:RDb, prefix:prefix1, next hop: ASBR1, VPN label: label-11.

The route propagates via RR to PE1.

6) PE1 learns <RDa,prefix1> with next hop ASBR1 and label-11.

7) An LSP from PE1 to ASBR1 is build via LDP, the corresponding label of the tunnel is label-T.

Data Plane:

1) PE1 performs MPLS label stack encapsulation with VPN label and tunnel label.

The packet leaving PE1: <label-T,label-11;payload>.

2) ABSR1 pop the tunnel label and swap the VPN label.

The packet leaving ASBR1 towards ASBR2: <label-21;payload>.

3) Based on the received label-21, ASBR2 looks up table.RD1.

Because the destination address of the payload match prefix2, ASBR2 encapsulates the payload in an outer IPv6 header where the destination address is SID21 and forward the packet to PE2.

If PE2 fails, ASBR2 would choose the backup outgoing action and forwards the packet to PE3 with SID31.

3.2. Control of the Allocation Method

The per-RD label allocation method can be enabled by configuration on ASBRs.

If many configurations are required in a large-scale network, controlling the allocation by protocol extensions may be taken into consideration.

A new BGP Extended Communities Attribute[RFC4360] is defined for this purpose.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
|      0x03     |     TBD1      |         Allocation Type
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
|                          Reserved
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

Figure 3: Label Allocation Extended Community

The value of the high-order octet of the extended type field is 0x03, which indicates it is transitive.

The value of the low-order octet of the extended type field for this community is TBD1, which indicates it is the label allocation extended community.

Allocation Type field is used to indicate the label/SID allocation method, the value 0x01 indicates that the label/SID should be allocated based on the RD value in the NLRI. Other values are reserved for future definition.

The egress PE advertises VPN route with the label allocation extended community to indicate that the ASBR should allocate the incomming MPLS label or SID for the received route based on the allocation method identified in the extended community.

An implementation SHOULD ensure that the label/SID allocation on ASBR doesn't conflict with each other.

4. IANA Considerations

IANA is requested to assign one new values from the "BGP Opaque Extended Community" type Registry. It is from the transitive range. The new value is called "Label Allocation Extended Community" (0x03TBD1). This document is the reference for the assignments.

5. Security Considerations

This document does not change the underlying security issues inherent in [RFC4364] and [I-D.ietf-bess-srv6-services].

6. References

6.1. 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>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

6.2. Informative 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", Work in Progress, Internet-Draft, draft-ietf-bess-srv6-services-08, 10 November 2021, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-srv6-services-08>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

Authors' Addresses

Yao Liu
ZTE
Nanjing
China

Email: liu.yao71@zte.com.cn

Shaofu Peng
ZTE
Nanjing
China

Email: peng.shaofu@zte.com.cn