Workgroup: BESS Working Group
Internet-Draft:
draft-lz-bess-srv6-service-capability-03
Published: 14 June 2022
Intended Status: Standards Track
Expires: 16 December 2022
Authors: Y. Liu Z. Zheng E. Metz
ZTE ZTE KPN

SRv6-based BGP Service Capability

Abstract

This draft describes the problems that may be encountered during the deployment of SRv6-based BGP services and the possible solutions.

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 16 December 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction
2. Requirements Language
3. the Co-existence Scenario
4. SRv6-based BGP Service Capability
5. Implementation Considerations
6. IANA Considerations
7. Security Considerations
8. References
   8.1. Normative References
   8.2. Informative References

Authors' Addresses

1. Introduction

[I-D.ietf-bess-srv6-services] defines procedures and messages for SRv6-based services. When an egress PE is enabled for BGP Services over SRv6 data-plane, it signals one or more SRv6 Service SIDs enclosed in SRv6 Service TLV(s) within the BGP Prefix-SID Attribute[RFC8669] attached to MP-BGP NLRIs. In other words, instead of defining new AFI/SAFIs for SRv6-based services to separate the SRv6-based service and MPLS-based service routes completely, this proposal leverages the existing AFI/SAFIs of MPLS-based services .

There're two methods to encode SRv6 service SIDs in the advertisement.

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

The second method is referred to as the Transposition Scheme in [I-D.ietf-bess-srv6-services], 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 of the SID.

[RFC8669] specifies that unknown TLVs in the BGP Prefix attribute MUST be ignored and propagated unmodified. PEs that only support MPLS may discard SRv6 Services TLV in the BGP Prefix attribute and treat the label in the NLRI as VPN route label for MPLS VPN.

This draft describes the problems that may be encountered during the deployment of SRv6-based services and the possible solutions.

2. 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].

3. the Co-existence Scenario

In the progress of network upgrading, some of the legacy devices that only support MPLS/SR-MPLS will coexist with the new devices capable of SRv6 for a long time.

As shown in Figure 1, PE1 is a legacy device that only supports MPLS-based services. PE2 and PE3 support both MPLS-based and SRv6-based services. There may be route reflector in the network to reflect the service routes. S-RR is a service route reflector that supports both MPLS and SRv6.

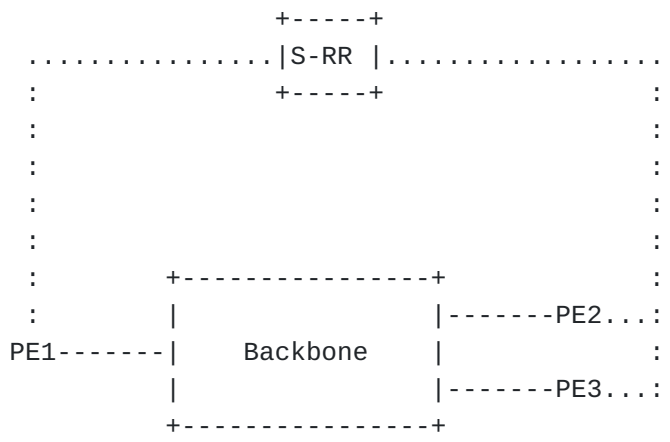


Figure 1: the Co-existence Scenario

On PE3, a SRv6 service SID sid-1 and a MPLS VPN route label label-1 are assigned for overlay service 1.

The SRv6 service SID and a MPLS VPN route label for the service 1 are advertised in separate UPDATE messages. ADD-PATH[RFC7911] is used to avoid path hiding. S-RR reflects both SRv6-VPN route and MPLS-VPN route to PE1. Since PE1 only supports MPLS, it may discard the SRv6 Service TLV(s) in the BGP Prefix attribute and treat the SRv6-based route as a MPLS-based route for service 1, then there're two MPLS-based routes for the same service 1 on PE1.

Depending on whether the Transposition Scheme is used, the following two scenarios are described separately.

Scenario 1:

If the Transposition Scheme is used, the function and/or argument part of sid-1 is encoded in the MPLS Label field of the NLRI of the SRv6-based service route.

PE1 may choose the route which is originally the SRv6-based route and use the label field in the NLRI of this route as MPLS VPN label for packet encapsulation.

Unless the allocation of SRv6 SIDs and MPLS labels on PE3 is aligned to ensure compatibility, the interpretation of the function and/or argument of the SRv6 SID (sid-1 in the example) will lead to incorrect forwarding of the traffic. In the example above, at PE3 packets may 1) be sent to the wrong service instance, in case the sid-1 function and/or argument value corresponds to an existing MPLS label, or 2) be dropped, in case the value of sid-1 does not correspond to an allocated MPLS label.

Scenario 2:

Sid-1 is encoded as a whole in the SRv6 Services TLV and the MPLS Label field of the corresponding NLRI is set to Implicit NULL.

If the SRv6 Services TLV in the UPDATE messages is discarded by PE1, from PE1's aspect, it has received a MPLS service route with an Implicit NULL label.

How to deal with the MPLS-based route with an Implicit NULL label is not standardized, different vendors may have different processing procedures which are unpredictable, e.g, set the route to invalid, send the packet to service 1 without the service route label or something else.

On PE2, only SRv6-based service is configured.

PE1 may receive SRv6 service routes from PE2 which supports SRv6 only, and discard the SRv6 Service TLV(s) in the BGP Prefix attribute and treat the function and/or argument part of SRv6 service SID as a MPLS VPN route label. PE1 may 1) not send packets to PE2 since there's no LSP between PE1 and PE2 2) send packets encapsulated in IPv6 to PE2 if there's route to PE2.

If the label field in the NLRI is Implicit NULL, how PE1 deals with it is unpredictable.

Overall, in the co-existence scenario, if the SRv6-based service routes are advertised to legacy devices, it may result in service failure and/or abnormal extra traffic flows in the network.

To avoid these problems, [I-D.ietf-bess-srv6-services] specifies that implementations MUST provide a mechanism to control

advertisement of SRv6-based BGP service routes on a per neighbor and per service basis.

This can be done by configuration. First the network operator must obtain whether the PEs in the network are capable of SRv6-based services. Then the operator should config on PEs or route reflectors based on each PE's capability, the configuration is per neighbor.

If there's a service route reflector, configurations on S-RR should ensure that the SRv6 service routes would not be reflected to legacy devices like PE1 that don't support SRv6.

If there's no route reflector in the network, which neighbors can the SRv6 service routes be advertised to should be specified when configuring SRv6 services on the PEs.

The above method may be feasible in small-scale networks, but are not applicable to large-scale networks.

The main reasons are:

a) The per neighbor configuration need to change with the device capability. When a PE is upgraded to support SRv6-based services or rolled back to an old version that only supports MPLS, the configuration on its neighbors or the RR should be changed to add this PE to or exclude it from the advertisement of SRv6-based BGP service routes.

Although this may be done automatically by the network management system, it is still not a easy job in a large-scale network and is not flexible enough.

b) The additional steps of device capability acquisition and capability based configuration increase the fault probability and troubleshooting difficulty. If the service from PE1 to PE3 fails, the operator needs to confirm the capability for SRv6-based service of the two devices, and then check the configuration on PE3 or RR to make sure that the SRv6-based service route is not advertised to PE1.

c) There is no standard solution for the network operator to obtain the PE's capability for SRv6-based services. If there are devices from multiple vendors in the network, there may be interconnection problems.

4. SRv6-based BGP Service Capability

If the BGP speaker can obtain the capability for SRv6-based services of its peers, the advertisement of SRv6-based BGP service routes can be controlled.

[RFC5492] defines the "Capabilities Optional Parameter". A BGP speaker can include a Capabilities Optional Parameter in a BGP OPEN message. This allows BGP speakers to communicate capabilities. 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 SRv6-based BGP service capability. If a BGP speaker has not sent the SRv6-based BGP service capability in its BGP OPEN message on a particular BGP session, or if it has not received the SRv6-based BGP service 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 SRv6 service TLVs. Like other capabilities, if the capability for SRv6-based services is enabled or removed, an established session needs to be reset to resend the OPEN message.

In this way, the advertisement of SRv6-based BGP service routes is controlled without per neighbor configuration, which makes it easier to implement and manage in the network.

In the co-existence scenario, the SRv6-based service routes would only be exchange between devices that support it based on this capability. There would be no UPDATE message that includes the SRv6 service TLV received by legacy devices.

Back to the scenario in Figure 1, since PE1 only supports MPLS and has not sent the SRv6-based BGP service capability in the OPEN message, the S-RR will not reflect the SRv6-based service routes of PE2 or PE3 to PE1, while the MPLS service routes from PE3 are reflected to PE1. So PE1 wouldn't receive any SRv6 SRv6-based service routes that may be misinterpretted, and the MPLS-based service between PE1 and PE3 is unaffected.

5. Implementation Considerations

PEs attached to the network, in general BGP speakers, indicate their ability to advertise and receive SRv6 based service routes through the SRv6 based BGP service capability.

If service route reflectors are deployed in the network, it's required that these RRs MUST support the SRv6-based BGP service capability.

In a co-existence/migration scenario it is RECOMMENDED that PEs that support SRv6 based service advertise both SRv6 and MPLS VPN routes. Advertisement the same VPN prefix as an SRV6 VPN or MPLS VPN route may be done through the BGP ADD-PATH capability or other method. Details of this is outside the scope of this document.

Use of the SRv6 based BGP service capability ensures that SRv6 based routes are distributed only among those nodes that support these. For example, consider a co-existence scenario with multihomed sites and PEs with different transport capabilities (no use of RRs) as shown in Figure 2. In this scenario CE1 is connected to PE2 and PE3, PE2 only supports MPLS VPN while PE3 supports SRv6 VPN, PE1 is the peer of PE2 and PE3, and supports both capabilities. If SRv6-based BGP service capability is introduced, PE1 will receive both MPLS VPN and SRv6 VPN routes, whereas PE2 will receive only MPLS VPN routes and PE3 will receive both.

In case a node receive both SRv6 VPN routes and MPLS VPN routes (for the same prefix), such as PE1, it may use one or both (e.g. ECMP) for forwarding depending on local policy.

The detailed deployment designs and implementation options are outside the scope of this document.

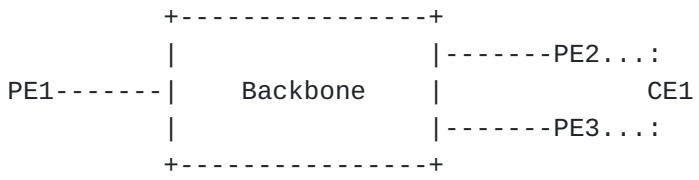


Figure 2: Multihomes Sites without RRs

6. IANA Considerations

This document defines a new Capability Codes option, named "SRv6 Service Capability" with an assigned value <TBD1> to indicate that a BGP speaker supports SRv6-based services. The length of this capability is 1.

7. Security Considerations

This extension to BGP does not change the underlying security issues inherent in [RFC5492] and [I-D.ietf-bess-srv6-services].

8. References

8.1. 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>.

[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>.

8.2. Informative References

[I-D.ietf-bess-srv6-services]
Dawra, G., 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-15, 22 March 2022, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-srv6-services-15>.

[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>.

Authors' Addresses

Yao Liu
ZTE
Nanjing
China

Email: liu.yao71@zte.com.cn

Zhang Zheng
ZTE
Nanjing
China

Email: zhang.zheng@zte.com.cn

Eduard Metz
KPN
Netherlands

Email: etmetz@gmail.com