BESS Working Group
Internet-Draft
Intended status: Informational
Expires: 4 September 2025

G. Mishra
Verizon Inc.
J. Tantsura
Microsoft, Inc.
M. Mishra
Cisco Systems
S. Madhavi
Juniper Networks, Inc.
A. Simpson
Nokia
S. Chen
Huawei Technologies

3 March 2025

Connecting IPv4 Islands over IPv6 Core using IPv4 Provider Edge Routers (4PE)

draft-mishra-idr-v4-islands-v6-core-4pe-09

Abstract

As operators migrate from an IPv4 core to an IPv6 core for global table routing over the internet, the need arises to be able provide routing connectivity for customers IPv4 only networks. This document provides a solution called 4Provider Edge, "4PE" that connects IPv4 islands over an IPv6-Only network.

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 4 September 2025.

Copyright Notice

Copyright (c) 2025 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

The problem being solved here for operators is how to connect IPv4 Islands over and IPv6 core without using traditional explicit complex tunneling technologies or IPv6 transition technology tunneling mechanisms which can be overly complicated.

"6PE" [RFC4798] is the specification for connecting IPv6 Islands over IPv4 MPLS Core using IPv6 Provider Edge Routers (6PE). This document explains the "4PE" design procedures and how to interconnect IPv4 islands over a IPv6-Only network. The 4PE routers exchange the IPv4 reachability information transparently over the core using the Multiprotocol Border Gateway Protocol (MP-BGP) over IPv6. Each ingress and egress PE (4PE) router builds an IPv6-signaled path without any explicit tunnel configuration and no IPv6 headers need to be inserted in front of the IPv4 packet over the PE-CE edge. The 4PE design is an alternative to the use of standard overlay tunneling technologies such as GRE/IP or any other tunneling technologies which requries explicit tunneling where with 4PE the tunnels are established dynamically. Providing the IPv4 connectivity to customers over an IPv6 core network is a challenge and complicated without using MP-BGP as described in this document.

4PE design specifies operations of the 4PE approach for interconnection of IPv4 islands over an IPv6-Only network. The approach requires that the PE-CE IPv4 islands to be Dual Stack using Multiprotocol BGP (MP-BGP) routers [RFC4760], while the core remains an IPv6-Only network.

In this document an 'IPv4 island' is a network running native IPv4 as per [RFC1812]. A typical example of an IPv4 island would be a customer's IPv4 site connected via its IPv4 Customer Edge (CE) router to one (or more) Dual Stack Provider Edge router(s) of a Service Provider. The PE-CE interface between the edge router of the IPv4 island Customer Edge (CE) router and the 4PE router is a native IPv4 interface which can be multiple physical or logical.

The interconnection method described in this document typically applies to an operator that may already be offering IPv4 or IPv6 BGP/MPLS VPN services for private MPLS, that wants to continue support IPv4 services to its internet customers over the IPv6 global routing table. Configuration and operations of the 4PE overlay approach has similarities to IPv4 VPN overlay service [RFC4364] or IPv6 VPN overlay service [RFC4659] to distribute IPv4 Network Layer Reachability Information (NLRI) for transport over an IPv6-Only network.

2. Terminology

Terminolgoy used in defining the 4PE specification.

IPv6-Only Network: MPLS, SR-MPLS SRv6

PE: Provider Edge

CE: Customer Edge

PE-CE: Provider Edge - Customer Edge

Ingress 4PE Router: Dual Stack Router (customer side:IPv4-only, net:IPv6-only

Egress 4PE Router: Dual Stack Router (net:IPv6-only, IPv4-only customer side)

3. 4PE Design Protocol Overview

Each IPv4 site is connected to at least one Provider Edge router connected to the IPv6-Only network. The PE router providing IPv4 connectivity to the IPv4 Islands over an IPv6-Only network is called a 4PE router. The 4PE router MUST be IPv4 and IPv6 dual stack. The 4PE router MUST be configured with at least one IPv6 address on the IPv6 Core side interface and at least one IPv4 address on the IPv4 Customer side PE-CE interface. The 4PE IPv6 address Loopback0 MUST to be routable within the IPv6 core.

The source side 4PE router receiving IPv4 packets from the local Attachment Circuit (AC) PE-CE IPv4-Only or IPv4 and IPv6 Dual Stacked interface Source IPv4 Site is called the Ingress 4PE router relative to these IPv4 packets sent by the Source CE IPv4 Island. The destination side 4PE router forwarding IPv4 packets to the local Attachment Circuit (AC) PE-CE IPv4-Only or IPv4 and IPv6 Dual stacked interface from the Source IPv4 Site sending location is called the Egress 4PE router relative to these IPv4 packets received by the CE IPv4 Island. Every ingress 4PE router can signal a path to send to

any egress 4PE router without injecting any additional prefixes into the IPv6 core other then the IPv6 signaled next hop Loopback0 used to identify the Ingress and Egress 4PE router.

Interconnecting IPv4 islands takes place through the following steps:

1. Exchange IPv4 reachability information among 4PE Ingress and Egress PE routers using MP-BGP [RFC2545]:

The 4PE routers exchange IPv4 prefixes over MP-BGP sessions as per [RFC2545] running over IPv6, MP-BGP Address Family Identifier (AFI) IPv4=1. In doing so, the 4PE routers convey their IPv6 address FEC label binding as the BGP Next Hop for the advertised IPv4 prefixes [16 or 32 bytes].

2. Transport IPv4 packets from the ingress 4PE router to the egress 4PE router:

The ingress 4PE router MAY forward the IPV4 NLRI as labeled prefixes using BGP-LU [RFC8277] over an IPv6-signalled LSP towards the towards the Egress 4PE router with IPv6 next hop encoding per [RFC8950].

The 4PE design is fully applicable to both full mesh BGP peering between all Ingress and Egress PE’s as well as when Route Reflectors iBGP peering is used where the PEs are all Route Reflector Clients.

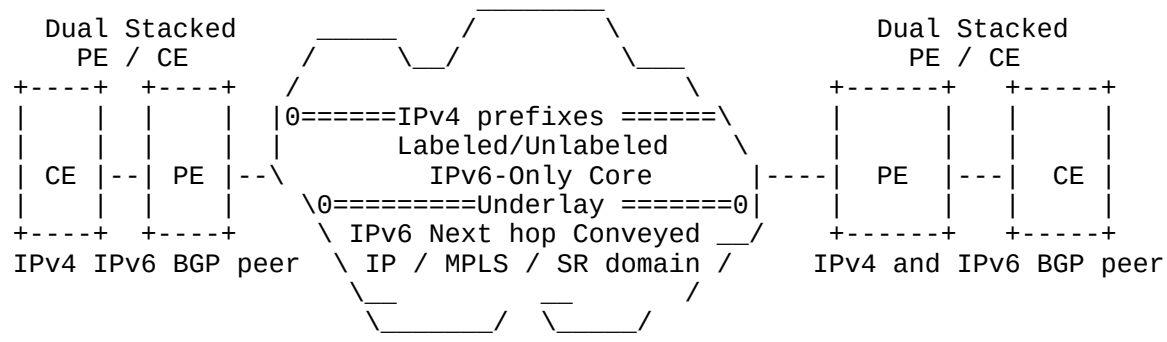

Figure 1: 4PE-Architecture

4. 4PE Design procedures

In this design, using IPv6 Next hop encoding defined in [RFC8950] allows a 4PE router that has to forward an IPv4 packets to automatically determine the IPv6-signaled path to use for a particular IPv4 destination by using the MP-BGP IPv4 NLRI.

When tunneling IPv4 packets over the IPv6 MPLS core, rather than successively prepend an IPv6 header and then perform label imposition based on the IPv6 header, the ingress 4PE Router has the option to directly perform label imposition of the IPv4 header without prepending any IPv6 header. The (outer) label imposed MUST correspond to the IPv6- signaled LSP starting on the ingress 4PE Router and ending on the egress 4PE Router.

While this design concept can operate in some situations using a single underlay topmost transport label, one option is to use a a second level of labels that are bound to the customer CE’s IPv4 prefixes via MP-BGP advertisements in accordance with [RFC8277].

The reason for labeling the IPv4 prefixes is as follows: 1. Allows for Penultimate Hop Popping (PHP) on the IPv6 Label Switch Router (LSR), upstream of the egress 4PE router. 2. After the topmost label has been popped, the Bototm of Stack (BOS) service label is now still present. 3. PHP node still transmits the labeled packets, instead of having to transmit unlableled IPv4 packets and now can encapsulate them appropriately so they are not dropped.

Another reason for second level bottom of stack label is for the existing IPv6-signaled LSP. This LSP is using "IPv6 Explicit NULL label" over the last hop. This is becauase the LSP is already being used to transport IPv6 traffic with the Pipe Diff-Serv Tunneling Model as defined in [RFC3270]). Thus the LSP could not be used to carry IPv4 with a single label since the "IPv6 Explicit NULL label" cannot be used to carry native IPv4 traffic [RFC3032]. While it could be used to carry Labeled IPv4 traffic [RFC4182]. [RFC3032] section 2.2 states that the LSR that pops the last label off the label stack must be able to identify the packets network layer protocol in this case IPv4. However, the label stack does not contain any field that explicitly carries the network layer protocol. Thus the network layer protocol must be inferrable from the value of the label which is popped from the bottom of the label stack along with subsequent headers. It is up to the network designer as to labeling the IPv4 prefixes or not based on the use case and desired and requirements. There maybe cases where it is not desirable to label the IPv4 prefixes and instead use a per CE label table LSP to carry the per CE unlabled IPv4 prefixes in a separate IPv4 routing context.

The label bound by MP-BGP to the IPv4 prefix indicates to the egress 4PE Router that the packet is an IPv4 packet. The label advertised by the egress 4PE Router with MP-BGP MAY be an explicit Null label Pipe mode Diff-Serv Tunneling Model use case as defined in [RFC3270]. In this case the topmost label can be preserved Ultimate Hop POP (UHP) to the egress PE. With the Default implicit-null Penultimate Hop (PHP) mode, the egress LSR P node would POP the topmost label revealing the native IPv4 packet which would be subsequently dropped as the Core underlay is an IPv6-Only core. There maybe cases where implicit null value 3 is not signaled by the egress PE either by default. In such case the implicit null is not signaled to the PHP node and thus is disabled. In this particular case explicit null label and Pipe mode Diff-Serv Tunneling Model is not necessary as the topmost label remains intact and preserved to the egress PE using any "arbitrary label".

BGP/MPLS VPN [RFC4364] defines 3 label allocation modes for Layer L2 and 3 VPN’s as follows: 1. Per Prefix label allocation mode where all prefixes are labeled. 2. Per-CE label allocation mode where all prefixes from a CE next hop are given the same label. 3. Per-VRF label allocation mode where all prefixes that belong to a VRF are given the same label. These options are available for L3 VPN for scalability and are also applicable to the 4PE. The two level label stack using a per prefix label allcoation mode is what is used in 6PE [RFC4798] with a requirement to label all the IPv6 prefixes using BGP-LU [RFC8277]. 4PE provides the same operator flxeiblity as BGP/ MPLS VPN [RFC4798], 2 level label stack option using Per-CE label allocation mode similar to MPLS VPN Per-VRF label allocation where the Per CE next hop is labeled so all prefixes associated within the CE get the same label.

5. 4PE MTU caveats

Every link in the IPv4 Internet must have an MTU of 576 octets or larger per [RFC1122]. Therefore, on MPLS links that are used for transport of IPv4, as per the 4PE approach, and that do not support link-specific fragmentation and reassembly, the MTU must be configured to at least 1280 octets plus the MPLS label stack encapsulation overhead bytes.

Some IPv4 hosts might be sending packets larger than the MTU available in the IPv6 MPLS core and rely on Path MTU discovery to learn about those links. To simplify MTU discovery operations, one option is for the network administrator to engineer the MTU on the core facing interfaces of the ingress 4PE consistent with the core MTU. ICMP ’ Destination Unreachable’ messages can then be sent back by the ingress 4PE without the corresponding packets ever entering the MPLS core. Otherwise, routers in the IPv6 MPLS network have the

option to generate an ICMP "Destination Unreachable" Fragmentation Required Type 3 Code 4 message using mechanisms as described in Section 2.3.2, "Tunneling Private Addresses through a Public Backbone" of [RFC3032].

Note that in the above case, should a core router with an outgoing link with an MTU smaller than 1280 receive an encapsulated IPv4 packet larger than 576, then the mechanisms of [RFC3032] may result in the "Unreachable" message never reaching the sender. This is because, according to [RFC4443], the underlay LSR (LSP or RSVP-TE tunnel) will build an ICMP "Unreachable " message filled with the invoking packet up to 1280 bytes. The LSR when forwarding downstream towards the egress PE as per [RFC3032], the MTU of the outgoing link will cause the packet to be dropped. This may cause significant operational problems. The originator of the packets will notice that his data is not getting through, without knowing why and where they are discarded. This issue would only occur if the above recommendation to configure MTU on MPLS links of at least 1280 octets plus encapsulation overhead is not used.

6. 4PE SR-MPLS Support

The 4PE design suports the Segment Routing SR-MPLS architecture [RFC8660], as SR-MPLS reuses the MPLS data plane with a new forwarding context using topological SIDs. The 4PE underlay signalling going from MPLS to SR-MPLS remains the same as the IPv6 LSP is still signalled as before from ingress PE to egress PE. The 4PE BGP overlay the design for SR-MPLS is identical to MPLS where the Ingress and Egress PE and the IPv4 NLIR can be optionally labeled.

All else remains the same as far as 4PE and Inter-AS options.

7. 4PE SRv6 Support

In the 4PE design over an SRv6 network using SRv6 Netowrk Programming [RFC8986] forwarding plane would use endpoint behavior "Endpoint with decapsulation and IPv4 cross-connect" behavior ("End.DX4" for short) is a variant of the End.X behavior for Global Table IPv4 Routing over SRv6 Core.

The End.DX4 SID MUST be the last segment in an SR Policy, and it is associated with one or more L3 IPv4 adjacencies and and SRv6 BGP Overlay Services [RFC9252] with the next hop encoding [RFC8950].

8. 4PE Deployment Options

In this section we display all the possible use cases and highlight the flexiblity of 6PE capabilities and use of 3 different topmost labaels that can be signaled

[RFC3032] does not require Penultimate Hop POP (PHP) to be enabled by default. When PHP is not signaled by the egress PE to the PHP node using implicit null value 3, an arbitrary label can be utilized for the topmost label. So in this case as PHP is not signaled by the egress PE node, PHP is not activated and thus the topmost label is presereved and not popped. Using an arbitarry label eliminates the need for explicit null value 1 for IPv4 and value 2 for IPv6 to be imposed as the means to preserve the topmost label for DiffServ PIPE mode.

In these use cases below we dispaly how the IPv4 prefixes tunnled over the IPv6 LSP can be either labed or not labeled depending on the customers design requirements

* Labeled IPv4 prefixes

* Unlabeled IPv4 prefixes

In this section we will describe three deployment modes below:

* Deployment Mode 1: Arbitrary label

* Deployment Mode 2: Explicit Null Label for Diffserv PIPE Mode UHP signaling

* Deployment Mode 3: Implicit Null label for PHP signaling

* Within each deployment mode we have the following three options:

* Option-1:Customer Prefix is labeled

* Option-2: Topmost PE-PE LSP is only labeled

* Option-2: Topmost with Per CE label table is only labeled

All deployment mode permutations are applicable to intra-as with Data planes MPLS, SR-MPLS, SRv6. They are applicable equally because the BGP overlay is data plane agnostic.

All deployment mode permutations are applicable to inter-as options A, B, C, AB, with Data planes MPLS, SR-MPLS, SRv6. They are applicable equally because the BGP overlay is data plane agnostic and inter-as options agnostic.

8.1. Deployment Mode-1 Arbitrary Labels

8.1.1. Mode-1 Arbitrary topmost with all customer prefixes labeled

Arbitrary topmost label where LERs signal IPv6 topmost LSP with 2 level label stack BOS set [RFC8277] 1/4 service label labeling all IPv4 customer prefixes

In this scenario all the attached CE prefixes in the global table are labled and this is similar to IP-VPN per perfix label allocation

Due to the per prefix label allocation in this scenario it is not as scalable and convergence maybe slower

8.1.2. Arbitrary topmost with PE to PE LSP

Arbitrary topmost label where LERs signal IPv6 topmost LSP with 2 level label stack, BOS set [RFC8277] 1/4 service label using ingress to egress PE loopback to loopback LSP single BOS label with all global table customer prefixes unlabeled. In this optimized scenario a single ingress 4PE to 4PE LSP is created to carry all the CE prefixes

This sceanario is most optimized from a label allocation perspective from all other scenarios in that only a single service label is allocated signaled by the service LSP which now is able to carry all of the global table prefixes populated by the attached CE’s as unlabeled IPv4 customer prefixes. This scenario is similar to IP-VPN Per-VRF Label allocation

This scenario provides per VRF prefix independent BGP PIC Edge like convergence with Per VRF prefix independence as when the PE LSP is withdrawn, all attached CE’s and related unlabled prefixes are as well withdrawn further optimizing the convergence and creating per VRF independence convergence

MPLS label allocation has a 20 bit label name space and thus allows for a maximum of 1 Millon labels. This is an MPLS protocol limit that is hardware and software independent. This scenario provides tremendous scale to the global internet table carried in the default VRF table now only allocating a single label for all 1 Million prefixes in the default VRF

8.1.3. Arbitrary topmost with per CE label table

Arbitrary topmost label where LERs signal IPv6 topmost LSP with 2 level label stack BOS set [RFC8277] 1/4 service label using per CE label table routing context LSP ingress to egress CE PE-CE interface PE side interface LSP single BOS label with per CE label table customer prefixes unlabeled.

This scenario is further optimized by creating a per CE next hop label table context similar to IP-VPN Per-CE or Per-Next-Hop label allocation mode where a single label is allocated per CE

In this scenario a single service label is allocated signaled by the CE interface IP between the ingress 4PE and egreess 4PE creating the per CE label context service LSP which we are now able to provide per CE next hop granularity label table context containing the per CE unlabled customer IPv4 prefixes.

This scenario provides further granularity and per CE independent BGP PIC Edge like convergence with per CE prefix independence as when the per CE LSP is withdrawn all the per CE related prefixes are as well withdrawn further optimizing the convergence and creating per CE independence granularity with the convergence

8.2. Deployment Mode-2 Explicit Null Label

8.2.1. Explicit Null topmost with all customer prefixes labeled

Explicit Null topmost label where LERs signal IPv6 topmost LSP with 2 level label stack BOS set [RFC8277] 1/4 service label labeling all IPv4 customer prefixes

In this scenario all the attached CE prefixes in the global table are labled and this is similar to IP-VPN per perfix label allocation

Due to the per prefix label allocation in this scenario it is not as scalable and convergence maybe slower

8.2.2. Explicit Null topmost with PE to PE LSP

Explicit Null topmost label where LERs signal IPv6 topmost LSP with 2 level label stack, BOS set [RFC8277] 1/4 service label using ingress to egress PE loopback to loopback LSP single BOS label with all global table customer prefixes unlabeled.

In this optimized scenario a single ingrees 4PE to 4PE LSP is created to carry all the CE prefixes

This sceanario is most optimized from a label allocation perspective from all other scenarios in that only a single service label is allocated signaled by the service LSP which now is able to carry all of the global table prefixes populated by the attached CE’s as unlabeled IPv4 customer prefixes. This scenario is similar to IP-VPN Per-VRF Label allocation

This scenario provides per VRF prefix independent BGP PIC Edge like convergence with Per VRF prefix independence as when the PE LSP is withdrawn, all attached CE’s and related unlabled prefixes are as well withdrawn further optimizing the convergence and creating per VRF independence convergence

MPLS label allocation has a 20 bit label name space and thus allows for a maximum of 1 Millon labels. This is an MPLS protocol limit that is hardware and software independent. This scenario provides tremendous scale to the global internet table carried in the default VRF table now only allocating a single label for all 1 Million prefixes in the default VRF

8.2.3. Explicit Null topmost with per CE label table

Explicit Null topmost label where LERs signal IPv6 topmost LSP with 2 level label stack BOS set [RFC8277] 1/4 service label using per CE label table routing context LSP ingress to egress CE PE-CE interface PE side interface LSP single BOS label with per CE label table customer prefixes unlabeled.

This scenario is further optimized by creating a per CE next hop label table context similar to IP-VPN Per-CE or Per-Next-Hop label allocation mode where a single label is allocated per CE

In this scenario a single service label is allocated signaled by the CE interface IP between the ingress 4PE and egreess 4PE creating the per CE label context service LSP which we are now able to provide per CE next hop granularity label table context containing the per CE unlabled customer IPv4 prefixes.

This scenario provides further granularity and per CE independent BGP PIC Edge like convergence with per CE prefix independence as when the per CE LSP is withdrawn all the per CE related prefixes are as well withdrawn further optimizing the convergence and creating per CE independence granularity with the convergence

8.3. Deployment Mode-3 Implicit Null Label

8.3.1. Implicit Null with all customer prefixes labeled

Implicit Null topmost label where LERs signal IPv6 topmost LSP with 2 level label stack BOS set [RFC8277] 1/4 service label labeling all IPv4 customer prefixes

In this scenario all the attached CE prefixes in the global table are labled and this is similar to IP-VPN per perfix label allocation

Due to the per prefix label allocation in this scenario it is not as scalable and convergence maybe slower

8.3.2. Implicit Null with PE to PE LSP

Implict Null topmost label where LERs signal IPv6 topmost LSP with 2 level label stack, BOS set [RFC8277] 1/4 service label using ingress to egress PE loopback to loopback LSP single BOS label with all global table customer prefixes unlabeled.

In this optimized scenario a single ingrees 4PE to 4PE LSP is created to carry all the CE prefixes

This sceanario is most optimized from a label allocation perspective from all other scenarios in that only a single service label is allocated signaled by the service LSP which now is able to carry all of the global table prefixes populated by the attached CE’s as unlabeled IPv4 customer prefixes. This scenario is similar to IP-VPN Per-VRF Label allocation

This scenario provides per VRF prefix independent BGP PIC Edge like convergence with Per VRF prefix independence as when the PE LSP is withdrawn, all attached CE’s and related unlabled prefixes are as well withdrawn further optimizing the convergence and creating per VRF independence convergence

MPLS label allocation has a 20 bit label name space and thus allows for a maximum of 1 Millon labels. This is an MPLS protocol limit that is hardware and software independent. This scenario provides tremendous scale to the global internet table carried in the default VRF table now only allocating a single label for all 1 Million prefixes in the default VRF

8.3.3. Implicit Null with per CE label table

Implicit Null topmost label where LERs signal IPv6 topmost LSP with 2 level label stack BOS set [RFC8277] 1/4 service label using per CE label table routing context LSP ingress to egress CE PE-CE interface PE side interface LSP single BOS label with per CE label table customer prefixes unlabeled.

This scenario is further optimized by creating a per CE next hop label table context similar to IP-VPN Per-CE or Per-Next-Hop label allocation mode where a single label is allocated per CE

In this scenario a single service label is allocated signaled by the CE interface IP between the ingress 4PE and egreess 4PE creating the per CE label context service LSP which we are now able to provide per CE next hop granularity label table context containing the per CE unlabled customer IPv4 prefixes.

This scenario provides further granularity and per CE independent BGP PIC Edge like convergence with per CE prefix independence as when the per CE LSP is withdrawn all the per CE related prefixes are as well withdrawn further optimizing the convergence and creating per CE independence granularity with the convergence

8.4. Arbitrary topmost with customer prefixes unlabeled

Arbitrary topmost IPv6 LSP BOS set single level label stack with all global table customer prefixes 1/1 unlabeled.

This scenario may require some deeper look into the packet Deep Packet Inspection (DPI) to determine next header inspection for protocol type so that the packets are not dropped.

8.5. Explicit Null topmost with customer prefixes unlabeled

Explicit null value 2 topmost IPv6 LSP BOS set single level label stack with all global table customer prefixes 1/1 unlabeled.

This scenario may require some deeper look into the packet Deep Packet Inspection (DPI) to determine next header inspection for protocol type so that the packets are not dropped.

9. Crossing Multiple IPv6 Autonomous Systems

Inter-AS 4PE Overview

This section describes the 4PE procedures for Inter-AS options [RFC4364] Section 10.

Like in the case of multi-AS backbone operations for IPv6 VPNs described in Section 10 of [RFC4364], there are three inter-as design options and a fourth option defined in [I-D.mapathak-interas-ab] that are described below.

* Inter-AS Option-A Back to Back VRF

* Inter-AS Option-B Segmented LSP

* Inter-AS Option-C End to End LSP with ASBR VRF offload

* Inter-AS Option-AB Combination of Option-A and Option-B

The Inter-AS connectivity is established by connecting the PE from one AS to the PE of another AS, whereby the PE providing global table routing reachability between ASes, as a 4PE router, is acting as an Autonomous System Boundary Router (ASBR) to provide the Inter-AS ASBR to ASBR, PE to PE connectivity between ASN’s. In the 4PE design the Inter-AS link extends the underlay transport LSP so it is now extended between the ASes. Bottom of Stack S bit is set and using BGP-LU IPv4 BGP Labeled Unicast all the IPv4 prefixes can now be advertised between the ASes.

9.1. Advertisement of IPv4 prefixes Inter-AS Procedure A

This 4PE Inter-AS extension involves the advertisement of IPv4 prefixes (non-Labeled) using Inter-AS Style procedure (a).

This design is the equivalent for exchange of IPv4 prefixes to Inter-AS Style procedure (a) Back to Back CE (no-labeled) Inter-AS path where each PE acts like a CE (No MPLS) as described in Section 10 of [RFC4364] for the exchange of VPN-IPv4 prefixes. In the Inter-AS Style Procedure (a) the Control plane carrying the (non-labeled) prefixes is together per VRF subinterfaces with the Data Plane forwarding over the Inter-AS ASBR to ASBR link.

In this scenario, the 4PE router uses iBGP to redistributes labeled IPv4 prefixes to a Route Reflector or Autonomous System Border Router (ASBR)4PE router to which an ASBR 4PE router it is a client. The ASBR then uses eBGP to advertise the (non labeled) IPv4 prefixes to an ASBR in another AS, which then distributes the IPv4 prefixes to 4PE routers in that AS or further redistributes to subsequent ASBRs and so on.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. IPv4 MUST to be activated on the Inter-AS ASBR to ASBR (non-labeled) links and each ASBR 4PE router MUST have at least one IPv4 address on the interface connected to the Inter-AS ASBR to ASBR, PE to PE link.

No inter-AS LSPs are used are used in this Inter-AS Procedure (a) as described in Section 10 of [RFC4364]. There is effectively a separate mesh of LSPs across the 4PE routers within each AS for which the (non-labeled) IPv4 prefixes are advertised within the AS as BGP-LU IPv4 labled prefixes carried in the IPv6 signaled transport LSP mesh.

In this design, the ASBR exchanging IPv4 prefixes MUST peer over IPv4. The exchange of IPv4 prefixes MUST be carried out as per [RFC4760].

9.2. Advertisement of labeled IPv4 prefixes Inter-AS Procedure B/C

9.2.1. Advertisement of labeled IPv4 prefixes Inter-AS Procedure B

This scenario involves the eBGP redistribution of overlay labeled IPv4 prefixes between source and destination ASs, along with underlay eBGP redistribution of labeled unicast IPv6 routes between source and destination ASs.

This scenario is the equivalent for exchange of IPv4 prefixes to Inter-AS procedure (b) described in Section 10 of [RFC4364] for the exchange of VPN-IPv4 prefixes.

In this scenario, the 4PE router uses iBGP to redistributes labeled IPv4 prefixes to a Route Reflector or Autonomous System Border Router (ASBR)4PE router to which an ASBR 4PE router it is a client. The ASBR then uses eBGP to advertise the labeled IPv4 prefixes to an ASBR in another AS, which then distributes the IPv4 prefixes to 4PE routers in that AS or further redistributes to subsequent ASBRs and so on.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. Thus IPv4 may or may not to be activated on the Inter-AS link

9.2.2. Advertisement of labeled IPv4 prefixes Inter-AS Procedure C

This scenario involves the eBGP multihop redistribution of overlay labeled IPv4 prefixes between source and destination ASs, along with underlay eBGP redistribution of labeled unicast IPv6 routes between source and destination ASs.

This scenario is the equivalent for exchange of IPv4 prefixes to Inter-AS procedure (c) described in Section 10 of [RFC4364] for exchange of VPN-IPv4 prefixes.

In this scenario the ASBRs need not be dual stacked as IPv4 prefixes redistributed between ASNs are tunneled over IPv6 and thus the IPv4 routes are not maintained or distributed on the 4PE ASBR routers. The 4PE ASBR only needs to maintain /128 IPv6 routes to all 4PE routers in its AS so it can redistribute these underlay routes to other ASs for inter-as reachability. The 4PE ASBRs and any transit ASBRs will use eBGP to pass along the /128 IPv6 routes to other ASs in order to create an end to end IPv6 LSP from source AS ingress PE rouer to destination AS egress PE router. Once the end to end IPv6 LSP is established, the 4PE routers in different ASs can now establish their eBGP multihop peering over IPv6 and now can exchange their IPv4 labeled unicast routes over the connection.

IPv4 need not be activated on the Inter-AS ASBR to ASBR, PE to PE links.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. IPv4 may or may not be activated on the Inter-AS link.

Note that the 4PE Inter-AS extension for procedure (c) in Section 10 of [RFC4364] that the exchange of IPv4 prefixes can only start after BGP has established IPv6 connectivity between the ASes.

10. IANA Considerations

There are not any IANA considerations.

11. Security Considerations

No new extensions are defined in this document. As such, no new security issues are raised beyond those that already exist in BGP-4 and use of MP-BGP for IPv6.

The security features of BGP and corresponding security policy defined in the ISP domain are applicable. It is recommended to provide use edge filtering and the domain boundaries as appropriate to secure the domain global table and limit access to meet the desired customer requiremennts.

For the inter-AS distribution of IPv6 prefixes according to case (a) of Section 4 of this document, no new security issues are raised beyond those that already exist in the use of eBGP for IPv6 [RFC2545].

12. Acknowledgments

Many thanks to Ketan Talaulikar, Robert Raszuk, Igor Malyushkin, Linda Dunbar, Huaimo Chen, Dikshit Saumya for your thoughtful reviews and comments.

13. References

13.1. Normative References

[I-D.ietf-idr-bgp-sr-segtypes-ext]
Talaulikar, K., Filsfils, C., Previdi, S., Mattes, P., and D. Jain, "Segment Routing Segment Types Extensions for BGP SR Policy", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-sr-segtypes-ext-08, 20 February 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-sr-segtypes-ext-08>.

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-26, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-26>.

[RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, DOI 10.17487/RFC1122, October 1989, <https://www.rfc-editor.org/info/rfc1122>.

[RFC1812] Baker, F., Ed., "Requirements for IP Version 4 Routers", RFC 1812, DOI 10.17487/RFC1812, June 1995, <https://www.rfc-editor.org/info/rfc1812>.

[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>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <https://www.rfc-editor.org/info/rfc2460>.

[RFC2545] Marques, P. and F. Dupont, "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545, DOI 10.17487/RFC2545, March 1999, <https://www.rfc-editor.org/info/rfc2545>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, DOI 10.17487/RFC3032, January 2001, <https://www.rfc-editor.org/info/rfc3032>.

[RFC3036] Andersson, L., Doolan, P., Feldman, N., Fredette, A., and B. Thomas, "LDP Specification", RFC 3036, DOI 10.17487/RFC3036, January 2001, <https://www.rfc-editor.org/info/rfc3036>.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, DOI 10.17487/RFC3107, May 2001, <https://www.rfc-editor.org/info/rfc3107>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[RFC3270] Le Faucheur, F., Ed., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, DOI 10.17487/RFC3270, May 2002, <https://www.rfc-editor.org/info/rfc3270>.

[RFC4029] Lind, M., Ksinant, V., Park, S., Baudot, A., and P. Savola, "Scenarios and Analysis for Introducing IPv6 into ISP Networks", RFC 4029, DOI 10.17487/RFC4029, March 2005, <https://www.rfc-editor.org/info/rfc4029>.

[RFC4182] Rosen, E., "Removing a Restriction on the use of MPLS Explicit NULL", RFC 4182, DOI 10.17487/RFC4182, September 2005, <https://www.rfc-editor.org/info/rfc4182>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[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>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <https://www.rfc-editor.org/info/rfc4760>.

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, DOI 10.17487/RFC5036, October 2007, <https://www.rfc-editor.org/info/rfc5036>.

[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>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7938] Lapukhov, P., Premji, A., and J. Mitchell, Ed., "Use of BGP for Routing in Large-Scale Data Centers", RFC 7938, DOI 10.17487/RFC7938, August 2016, <https://www.rfc-editor.org/info/rfc7938>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[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>.

[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>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

[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>.

[RFC8950] Litkowski, S., Agrawal, S., Ananthamurthy, K., and K. Patel, "Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop", RFC 8950, DOI 10.17487/RFC8950, November 2020, <https://www.rfc-editor.org/info/rfc8950>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

[RFC9252] Dawra, G., Ed., Talaulikar, K., Ed., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)", RFC 9252, DOI 10.17487/RFC9252, July 2022, <https://www.rfc-editor.org/info/rfc9252>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[RFC9313] Lencse, G., Palet Martinez, J., Howard, L., Patterson, R., and I. Farrer, "Pros and Cons of IPv6 Transition Technologies for IPv4-as-a-Service (IPv4aaS)", RFC 9313, DOI 10.17487/RFC9313, October 2022, <https://www.rfc-editor.org/info/rfc9313>.

13.2. Informative References

[I-D.mapathak-interas-ab]
Pathak, M., Patel, K., and A. Sreekantiah, "Inter-AS Option D for BGP/MPLS IP VPN", Work in Progress, Internet-Draft, draft-mapathak-interas-ab-02, 28 May 2015, <https://datatracker.ietf.org/doc/html/draft-mapathak-interas-ab-02>.

[RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, DOI 10.17487/RFC4659, September 2006, <https://www.rfc-editor.org/info/rfc4659>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <https://www.rfc-editor.org/info/rfc4684>.

[RFC4798] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", RFC 4798, DOI 10.17487/RFC4798, February 2007, <https://www.rfc-editor.org/info/rfc4798>.

[RFC4925] Li, X., Ed., Dawkins, S., Ed., Ward, D., Ed., and A. Durand, Ed., "Softwire Problem Statement", RFC 4925, DOI 10.17487/RFC4925, July 2007, <https://www.rfc-editor.org/info/rfc4925>.

[RFC5549] Le Faucheur, F. and E. Rosen, "Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop", RFC 5549, DOI 10.17487/RFC5549, May 2009, <https://www.rfc-editor.org/info/rfc5549>.

[RFC5565] Wu, J., Cui, Y., Metz, C., and E. Rosen, "Softwire Mesh Framework", RFC 5565, DOI 10.17487/RFC5565, June 2009, <https://www.rfc-editor.org/info/rfc5565>.

[RFC6074] Rosen, E., Davie, B., Radoaca, V., and W. Luo, "Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)", RFC 6074, DOI 10.17487/RFC6074, January 2011, <https://www.rfc-editor.org/info/rfc6074>.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

Authors' Addresses

Gyan Mishra
Verizon Inc.
Email: gyan.s.mishra@verizon.com

Jeff Tantsura
Microsoft, Inc.
Email: jefftant.ietf@gmail.com

Mankamana Mishra
Cisco Systems
821 Alder Drive,
MILPITAS
Email: mankamis@cisco.com

Sudha Madhavi
Juniper Networks, Inc.
Email: smadhavi@juniper.net

Adam Simpson
Nokia
Email: adam.1.simpson@nokia.com

Shuanglong Chen
Huawei Technologies
Email: chenshuanglong@huawei.com