Workgroup: BESS Working Group
Internet-Draft:
draft-mishra-idr-v4-islands-v6-core-4pe-06
Published: 22 October 2023
Intended Status: Informational
Expires: 24 April 2024
Authors: G. Mishra J. Tantsura M. Mishra
Verizon Inc. Microsoft, Inc. Cisco Systems
S. Madhavi A. Simpson
Juniper Networks, Inc. Nokia
S. Chen
Huawei Technologies

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

Abstract

As operators migrate from an IPv4 core to an IPv6 core for global table internet routing, 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 Core Underlay Network.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

"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 Multiprotocol Label Switching (MPLS) [RFC3031] LDPv6 [RFC5036] enabled IPv6-Only core, Segment Routing (SR) enabled SR-MPLS [RFC8660] IPv6-Only core or SRv6 [RFC8986] IPv6-Only core. The 4PE routers exchange the IPv4 reachability information transparently over the core using the Multiprotocol Border Gateway Protocol (MP-BGP) over IPv6. In doing so, the BGP Next Hop field egress PE FEC (Forwarding Equivalency Class) is used to convey the IPv6 address of the 4PE router learned dynamically via IGP so that the dynamically established IPv6-signaled MPLS Label Switched Paths (LSPs) or SRv6 Network Programming IPv6 forwarding path instantiation and can be utilized without any explicit tunnel configuration.

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 tunnel termination at the tunnel endpoints which creates added layer of complexity to the existing MPLS or Segment routing underlay transport layer. The 4PE design provides a solution for MPLS as well as Segment Routing environment, where all tunnels are established dynamically using existing Service Provider Network MPLS signalling or SRv6 Network Programming thereby addressing environments where the effort to configure and maintain explicitly configured tunnels is not acceptable.

Alternative designs exist in 6MAN and v6OPS Working groups related to 4to6 transition technologies referred to as "IPv4aas" IPv4 as-a-service solutions [RFC9313] such as 464XLAT, Dual--Stack Lite, MAP-E, MAP-T, however this document focuses on a BGP based solution "4PE' to connecting IPv4 islands over an IPv6 Core network.

4PE design specifies operations of the 4PE approach for interconnection of IPv4 islands over an MPLS LDP IPv6 core, Segment Routing SR-MPLS IPv6 core or SRv6 IPv6 core. The approach requires that the Provider Edge (PE) routers Provider Edge - Customer Edge (PE-CE) connections to Customer Edge (CE) IPv4 islands to be Dual Stack using Multiprotocol BGP (MP-BGP) routers [RFC4760], while the core is a [RFC5565] Softwire Mesh Framework single protocol Provider (P) Core routers, are required only to support IPv6-Only dataplane to transport IPv4 packets over an IPv6-Only Core supporting three core technologies, MPLS LDPv6, Segment Routing SR-MPLS and Segment Routing IPv6 (SRv6). The approach uses MP-BGP over IPv6, relies on identification of the 4PE routers by their IPv6 address, and uses an underlay transport label switched IPv6-signaled MPLS, SR-MPLS LSP's, underlay SRv6 SRv6-TE or SRv6 SRv6-BE Best Effort path instantiation without any requirements for complex explicit tunnel configurations.

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. These Dual Stacked or IPv4-Only Provider Edge routers (4PE) are connected to an IPv6 MPLS core network.

The interconnection method described in this document typically applies to an Internet Service Provider (ISP) or Enterprise that has an MPLS LDP IPv6 core, Segment Routing SR-MPLS IPv6 core or SRv6 IPv6 core, that is already offering IPv6 BGP/MPLS VPN services, that wants to continue support IPv4 services to its customers. These 4PE PE Edge routers provide connectivity to the Customer Edge (CE) IPv4 islands Edge routers. They may also provide IPv4 and IPv6 services simultaneously (IPv4 and IPv6 connectivity, L3VPN services, L2VPN services, etc.). With the 4PE approach, no tunnels need to be explicitly configured, and no IPv6 headers need to be inserted in front of the IPv4 packets between the customer and provider edge, PE-CE Demark.

The main use case for 4PE is where the operator needs to provide IPv4 island connectivity over an IPv6 Core network that uses MPLS, SR-MPLS, SRv6 for the underlay transport where Layer 3 IP/VPN overlay 4VPE or VPN-IPv4 AFI/SAFI 1/128 [RFC4364] is not utilized such as for internet service providers carrying the internet routing table in the global table and not in a Layer 3 IP/VPN separate VRF instance or any other similar style Layer 3 VPN service offering.

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. Static routing or a dynamic routing protocol Interior Gateway Protocol IGP, Open Shortest Path First (OSPF) or Intermediate System Intermediate System (ISIS) or Exterior Gatway Protocol such as BGP may run between the CE router and the 4PE router for the distribution of IPv4 Network Layer Reachability Information (NLRI).

The 4PE design described in this document can be used for customers that require both IPv4 and IPv6 service as well as for customers that require IPv4-Only connectivity thus providing global IPv4 reachability.

Deployment of the 4PE approach over an existing IPv6 MPLS or Segment Routing core uses existing mechanisms in the core underlay transport, using new standardized procedures and techinques for ingress and egress 4PE specification standardization defined in this document. Configuration and operations of the 4PE approach has similarities with the configuration and operations of an IPv4 VPN service [RFC4364] or IPv6 VPN service [RFC4659] over an IPv6 MPLS or Segment Routing core because they all use MP-BGP to distribute IPv4 Network Layer Reachability Information (NLRI) for transport over an IPv6 Core.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. 4PE Design Protocol Overview

Each IPv4 site is connected to at least one Provider Edge router that is located on the border of the MPLS [RFC3031] LDP IPv6, Segment Routing SR-MPLS [RFC8660] IPv6 or SRv6 [RFC8986] Core Network. The PE router providing IPv4 connectivity to the IPv4 Islands over an IPv6-Only Core 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. In the MPLS LDP IPv6 and SR-MPLS IPv6 Core, or SRv6 corescenario, the 4PE IPv6 address Loopback0 MUST to be routable within the IPv6 core. For the MPLS LDP IPv6 Core there MUST be an LDP IPv6 label binding, and for SR-MPLS an IPv6 Prefix / Node SID label binding and for SRv6 SRH processing of SRv6 SID list and SRv6 Network Programming [RFC8986] SRv6 End.DX4 (Cross Connect to Next Hop), End.DT4 (Table lookup), End.DT46, using SRv6 BGP Overlay Services [RFC9252] for the 4PE SRv6 service overlay.

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 an IPv6 MPLS LSP, SR-MPLS LSP or instantiate an SRv6 Best Effort (BE) or Segment Routing Traffic Engineering (SR-TE) [RFC9256]. 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 over an MPLS LDP IPv6 Core, Segment Routing SR-MPLS IPv6 core, or SRv6 IPv6 core. 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. The IPv6 address of the egress 4PE next hop router is encoded using [RFC8950] next hop encoding for the BGP Next Hop field with a length of 16 or 32 bytes. The next hop encoding [RFC8950] is constructed using MP-BGP for IPv6 [RFC2545] is a 16 byte IPv6 Global Unicast Address followed by the 16 byte IPv6 Link Local Address if the Next Hop is on a common subnet with peer. The ingress and egress 4PE router has the option to bind a label to the IPv4 prefix as per [RFC8277] using BGP Labeled Unicast herinafter called BGP-LU, AFI/SAFI Address Family (AFI) / Subsequent Address Family Identifier (SAFI) 2-tuple "1/4".

2. Transport IPv4 packets from the ingress 4PE router to the egress 4PE router over IPv6-signaled LSPs, SRv6 BE or SR-TE instantiated path over an IPv6-only core:

The Ingress 4PE router MAY forward IPv4 NLRI as Labeled prefixes using BGP-LU SAFI over the IPv6-signaled LSP towards the Egress 4PE router identified by the IPv4 address advertised in the 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 or other use cases such as in a BGP only Data Center [RFC7938] where Spine layer eBGP Route Servers are utilized as per BGP specification [RFC4271].

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 LSP to use for a particular IPv4 destination by using the MP-BGP IPv4 NLRI.

To ensure interoperability between routers that implement the 4PE design over MPLS [RFC3031] LDP IPv6 Core described in this document, ingress and egress 4PE SHOULD support building the underlay tunneling using IPv6-signaled MPLS LSPs established by LDP [RFC5036] or Resource Reservation Protocol (RSVP-TE) [RFC3209].

To ensure interoperability between routers that implement the 4PE design over SR-MPLS [RFC8660], SHOULD support building static or stateful PCE SID list for IPv6 signaled LSP to egress 4PE IPv6 Loopback endpoint, or SRv6 [RFC8986] SRH processing of SRv6 SID list [RFC8754] and SRv6 Network Programming [RFC8986] SRv6 End.DX4 (Cross Connect to Next Hop), End.DT4 (Table lookup), End.DT46 using SRv6 BGP Overlay Services [RFC9252] for the 4PE SRv6 service overlay, static or stateful PCE SID list to egress 4PE IPv6 loopback endpoint.

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 Xwithout 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 that it allows for Penultimate Hop Popping (PHP) on the IPv6 Label Switch Router (LSR), upstream of the egress 4PE router, after the topmost label has been popped, the Bototm of Stack (BOS) service label is now still present, so the PHP node still transmits the labeled packets, instead of having to transmit unlableled IPv4 packets and encapsulate them appropriately so they are not dropped.

Another reason for second level bottom of stack label is for the existing IPv6-signaled LSP that is using "IPv6 Explicit NULL label" over the last hop because that LSP is already being used to transport IPv6 traffic with the Pipe Diff-Serv Tunneling Model as defined in [RFC3270]), thus 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], so that 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 otherwise and 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 3 VPN's per prefix where all prefixes are labeld, Per-CE label allocation mode where all prefixes from a CE next hop are given the same label and a 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 applicable to the 4PE design. 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]. The 4PE design provides the same operator flxeiblity as BGP/MPLS VPN [RFC4798], 2 level label stack option using Per-CE label allocation mode where the next hop is label so all prefixes associated with CE get the same label. The 4PE design provides the same operator flxeiblity as BGP/MPLS VPN [RFC4798], 2 level label stack option using Per-VRF label allocation mode where all prefixes within a VRF get the same is label.

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, and 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.

5. 4PE SR-MPLS Support

Segment Routing (SR) [RFC8402] leverages the source-routing paradigm to steer packets from a source node through a controlled set of instructions, called segments, by prepending the packet with an SR header in the MPLS data plane SR-MPLS [RFC8660] through a label stack or IPv6 data plane using an Segment Routing Header (SRH) header via SRv6 [RFC8754] to construct an SR path. Segment Routing will be referred to hereinafter as "SR". SR uses instructions called segments which can be topological segments used for transport underlay traffic steering or service instructions for overlay services. SR's Source Routing Architecture provides a mechansim to steer a flow onto a topological path, while maintaining per flow state only on the ingress source nodes within the SR domain. SR-MPLS reuses the Interior Gateway Protocol (IGP) control plane as well as the MPLS forwarding plane functions as the SR segments are instantiated as MPLS labels and the Segment Routing SR-MPLS Header is instantiated as a stack of MPLS labels. SR-MPLS L2 VPN and L3 VPN services can be steered using Traffic Engineered paths using SR-TE Policy coloring for the path instantiation per [RFC9256] and [I-D.ietf-idr-segment-routing-te-policy].

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 MPLS data plane procedrues defined in [RFC3031]. The 4PE BGP overlay the design for SR-MPLS is identical to MPLS where the Ingress and Egress PE Label Stack on the 4PE router contains the Service label with Bottom of Stack "S" bit set and contains the IPv4 NLRI prefixes "labeled" using BGP-LU, IPv4 Address Family Identifier (AFI) IPv4 (value 1) Subsequent Address Family Identifier (SAFI)(value 4).

4PE design with SR-MPLS data plane MUST also use "IPv6 Explicit Null label" value 2 defined in [RFC4182] Pipe Diff-Serv Tunneling Model as defined in [RFC3270].

SR-MPLS can use Inter-AS options for 4PE procedures which is identical to MPLS as well as can use SR-TE Policy and Binding SID for candidate path per [RFC9256] and [I-D.ietf-idr-segment-routing-te-policy].

6. 4PE SRv6 Support

Segment Routing (SR) [RFC3031] SRv6 leverages the source-routing paradigm to steer packets from a source node through a controlled set of instructions, called segments, by prepending the packet with a new SR header over an IPv6 data plane called an IPv6 Routing Extension Header type 4 called a Segment Routing Header (SRH) header with IPv6 SRH encoding [RFC8754] to construct an SR steered path. SRv6 Network Programming framework provides the mechanism based on segment endpoint behaviors to encode a sequence of instructions called Segments into an IPv6 header. SRv6 defines a topological or service segment as an IPv6 address with is called hereinafter a SID or "Segment ID". Each SID is encoded into an SRH header per [RFC8754]on the SR domain source node in the SR domain to steer a flow onto a topological path. In SRv6 each SID is an IPv6 address with format LOC:FUNC:ARG where the LOCATOR field "LOC" is the L most significant bits of the SID, followd by F bits of FUNCTION field "FUNC" and A bits of ARGUMENT "ARG". Each node in the SRv6 domain has a "LOC" prefix assigned which is routable and it leads to the SRv6 node which instantiates the SID by performing the endpoint processing on the node. The SRv6 SID FUNCTION "FUNC" field is used to encode the BGP/MPLS L3 VPN [RFC4364] or BGP EVPN Service labels [RFC7432] as defined in SRv6 BGP Overlay Services [RFC9252]. Intermediate nodes within an SRv6 domain process the topolocial SID at each segment endpoint defined in the SRH header until the packet reaches the egress PE where decapsulation happens similar to BGP/ MPLS L3 VPN [RFC4364], where the service labels encoded in the FUNC field can be instantiated and processed for the corresponding Layer 2 VPN and Layer 3 VPN service specific endpoint functions. SRv6 based BGP services referes to Layer 2 VPN and Layer 3 VPN overlay services with BGP as a control plane and SRv6 as a Data Plane to provide Best Effort (BE) which means that an SRH is not present and is reffered to as SRv6-BE. SRv6 based BGP services referes to Layer 2 VPN and Layer 3 VPN overlay services with BGP as a control plane and SRv6 as a Data Plane to provide Traffic Engineered (TE) which means that an SRH is present is reffered to as SRv6-TE policy for SRH topological instruction encoding for SR-TE Policy coloring for path steering instantiation per [RFC9256] and [I-D.ietf-idr-segment-routing-te-policy]. SRv6 Service SID and refers to an SRv6 SID associated with one of the service-specific endpoint behaviors on the advertising PE router such as END.DT (Table Lookup in a VRF) or END.DX (Cross Connect to a Next Hop) behaviors for Layer 3 VPN services defined in SRv6 Network Programming [RFC8986] BGP Prefix SID Attribue is used to carry the SRv6 SIDs and their associaed BGP Address Families and defines a SRv6 L3 Service TLV which encodes the SRv6 Service SID Information for SRv6 based L3 Services. SRv6-BE providing "Best Effort" connectivity where an SRH is not present, the egress PE signals the SRv6 Service SID with the BGP overlay service route and encapsulates the payload in an outer IPv6 header where the destination address is the SRv6 Service SID enclosed in SRv6 Service TLV(s) provided by the Egress PE in which case the underlay need only support plan IPv6 forwarding. SRv6-TE provides connectivity over a "Traffic Engineered" (TE) path by encapsulating the payload packet in an outer IPv6 header with the segment list of the SR policy related to the SLA along with SRv6 Service SID enclosed in SRv6 Service TLV(s) assoicated with route using SRH segment list encoding [RFC8754] from ingress PE to egress PE, the egress PE colors the overlay service route with a Color Extended Community [I-D.ietf-idr-segment-routing-te-policy] to instantiate the steering of flows with per flow state only maintained on the SRv6 source node and all underlay nodes whos SRv6 SID are part of the SRH Segment List MUST support the SRv6 Data Plane forwarding.

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] where the next hop encoding [RFC8950] is constructed using MP-BGP for IPv6 [RFC2545] is a 16 byte IPv6 Global Unicast Address followed by the 16 byte IPv6 Link Local Address if the Next Hop. In the 4PE design the SRv6 L3 Service SID is encoded as part of the SRv6 L3 Service TLV for SRv6 Netowrk Programming [RFC8986] endpoint behavior End.DX4 BGP Prefix SID Attribute encoding of SRv6 Service SID, SRv6 L3 Service TLV encoding [RFC9252] advertised by egress PEs which supports SRv6 based Layer 3 Services along with Service SID enclosed in SRv6 Layer 3 Service TLV, Label field for an IPv4 prefix is encoded with 20-bit label value set as specified by BGP-LU [RFC8277] to the whole or portion of the "FUNCTION" part of the SRv6 SID when the transposition encoding scheme is used or otherwise set to NULL. The "FUNCTION" part of the SRv6 SID now carries the overlay 4PE BGP-LU IPv4 Labeled prefix identical to MPLS and SR-MPLS.

In the 4PE design over an SRv6 network using SRv6 Netowrk Programming [RFC8986] forwarding plane would use endpoint behavior "Endpoint with decapsulation and specific IPv4 table lookup" behavior ("End.DT4" for short) is a variant of the End.T behavior for Global Table IPv4 Routing over SRv6 Core, The End.DT4 SID MUST be the last segment in an SR Policy, and a SID instance is assocated with a IPv4 FIB Table T. and SRv6 BGP Overlay Services [RFC9252] where the next hop encoding [RFC8950] is constructed using MP-BGP for IPv6 [RFC2545] is a 16 byte IPv6 Global Unicast Address followed by the 16 byte IPv6 Link Local Address if the Next Hop. In the 4PE design the SRv6 L3 Service SID is encoded as part of the SRv6 L3 Service TLV for SRv6 Netowrk Programming [RFC8986] endpoint behavior End.DT4 BGP Prefix SID Attribute encoding of SRv6 Service SID, SRv6 L3 Service TLV encoding [RFC9252] advertised by egress PEs which supports SRv6 based Layer 3 Services along with Service SID enclosed in SRv6 Layer 3 Service TLV, Label field for an IPv4 prefix is encoded with 20-bit label value set as specified by BGP-LU [RFC8277] to the whole or portion of the "FUNCTION" part of the SRv6 SID when the transposition encoding scheme is used or otherwise set to NULL. The "FUNCTION" part of the SRv6 SID now carries the overlay 4PE BGP-LU IPv4 Labeled prefix identical to MPLS and SR-MPLS.

4PE design with SRv6 data plane MUST also use "IPv6 Explicit Null label" value 2 defined in [RFC4182] Pipe Diff-Serv Tunneling Model as defined in [RFC3270].

SRv6 can use Inter-AS options for 4PE procedures which is equivalent to MPLS using SRv6 Service SID enocded in BGP Prefix SID Attribute as well as can use SR-TE Policy and Binding SID for candidate path per [RFC9256] and [I-D.ietf-idr-segment-routing-te-policy].

7. 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 and in that 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.

*Arbitrary label

*Explicit Null Label for Diffserv PIPE Mode UHP signaling

*Implicit Null label for PHP signaling

In these use cases we dispaly how the IPv4 prefixes tunnled over the IPv6 LSP can be labed or not labeled

*Labeled IPv4 prefixes

*Unlabeled IPv4 prefixes

All deployment options are applicable to intra-as and inter-as options A, B, C, AB, with Data planes MPLS, SR-MPLS, SRv6.

7.1. 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

7.2. 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 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

7.3. 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

7.4. 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

7.5. 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

7.6. 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

7.7. 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

7.8. Implicit Null with PE to PE LSP

Implicit 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

7.9. 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

7.10. 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.

7.11. 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.

8. Crossing Multiple IPv6 Autonomous Systems

8.1. Inter-AS 4PE Overview

This section discusses the use case where two IPv4 islands are connected to different Core Autonomous Systems (ASes)and utilizes 4 PE to connect the two Core ASes together. 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.

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.

8.2. Advertisement of IPv4 prefixes using Inter-AS Style Procedure A Procedures for 4PE

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 design, the Source 4PE routers within the Source AS use IBGP MP-BGP [RFC4760] carrying IPv4 NLRI over an IPv6 Next Hop using IPv6 Next hop encoding [RFC8950] and BGP-LU [RFC8277] to advertise labeled IPv4 prefixes to a Route Reflector to which it is a client, which then advertises the labeled IPv4 prefixes to an Autonomous System Border Router (ASBR) 4PE router which is also a client of the route reflector, connecting eBGP to another Autonomous System Border Router (ASBR) 4PE router. The ASBR then uses eBGP to advertise the (non-labeled) IPv4 prefixes to an ASBR in another AS, which in turn advertises the IPv4 prefixes to a route reflector within that AS of which it is a client which then advertises the IPv4 prefixes to all the 4PE routers in that directly connected AS or as described earlier in this specification to another ASBR, which in turn repeats the Inter-AS Procedure (a) herinafter in a case where ASN's are linked togetther with multiple 4PE AS hops.

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].

8.3. Advertisement of labeled IPv4 prefixes Inter-AS Style Procedure B and C

8.3.1. Advertisement of labeled IPv4 prefixes Inter-AS Style Procedure B

This 4PE Inter-AS extension involves the advertisement of labeled IPv4 prefixes over a segmented LSP using Inter-AS Style procedure (b). In this 4PE extension of Inter-AS Style procedure (b) the 4PE IPv4 BGP-LU labeled Unicast RIB is maintained on the ASBR.

This design 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 the Inter-AS Style Procedure (b) the Control plane carrying the Service label prefixes is together in the label stack with the Data Plane forwarding over the Inter-AS ASBR to ASBR link.

In this design, the Source 4PE routers within the Source AS use IBGP MP-BGP [RFC4760] carrying IPv4 NLRI over an IPv6 Next Hop using IPv6 Next hop encoding [RFC8950] and BGP-LU [RFC8277] to advertise labeled IPv4 prefixes to a Route Reflector to which it is a client, which then advertises the labeled IPv4 prefixes to an Autonomous System Border Router (ASBR) 4PE router which is also a client of the route reflector, connecting eBGP to another Autonomous System Border Router (ASBR) 4PE router. The ASBR then uses eBGP to advertise the labeled IPv4 prefixes to an ASBR in another AS, which in turn advertises the IPv4 prefixes to a route reflector within that AS of which it is a client which then advertises the IPv4 prefixes to all the 4PE routers in that directly connected AS or as described earlier in this specification to another ASBR, which in turn repeats the Inter-AS Procedure (a) herinafter in a case where ASN's are linked togetther with multiple 4PE AS hops.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. The label stack on the ASBR to ASBR, PE to PE link is 2 labels deep, with the IPv6 tompost transport label IPv6 signaled LSP using BGP-LU IPv6 Labeled Unicast, IPv6 Address Family Identifier (AFI) IPv4 (value 2) Subsequent Address Family Identifier (SAFI) (value 4) and Bottom of Stack BGP-LU IPv4 labeled Unicast Service label, IPv4 Address Family Identifier (AFI) IPv4 (value 1)

Subsequent Address Family Identifier (SAFI)(value 4) Thus IPv4 is not required to be activated on the Inter-AS ASBR to ASBR PE to PE links as IPv4 is tunnled through the IPv6 signaled LSP.

This 4PE Inter-AS procedure (b) described in Section 10 of [RFC4364] requires that there be label switched paths established across ASes. Hence the corresponding considerations described for procedure (b) in Section 10 of [RFC4364] apply equally to this design regarding trust relationship between Service Providers in extending the Inter-AS LSP between ASBR's.

8.3.2. Multi-hop advertisement of labeled IPv4 prefixes Inter-AS Style Procedure C

This 4PE Inter-AS extension involves the Route Reflector to Route Reflector Control Plane Multi-hop eBGP advertisement of labeled IPv4 Unicast prefixes between source and destination ASes, with Inter-AS link transport underlay IPv6 signaled LSP eBGP advertisement of labeled Unicast IPv4 prefixes from AS to neighboring AS. In this 4PE extension of Inter-AS Style procedure (c), the 4PE IPv4 BGP-LU labeled Unicast RIB is not maintained on the ASBR.

This design 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 the Inter-AS Style Procedure (c) the Control plane carrying the Service label prefixes eBGP Multihop, Route Reflector to Route Reflector is separated from the data plane forwarding over the Inter-AS ASBR to ASBR link which caries the underlay PE loopbacks advertised using BGP-LU between the Source and Destination AS over the Inter-AS ASBR-ASBR link. The Core AS underlay /128 PE loopbacks must be advertised in IPv6 Address Family Identifier (AFI) IPv4 (value 2) Subsequent Address Family Identifier (SAFI)(value 4).

In this design, the Source 4PE routers within the Source AS use IBGP MP-BGP [RFC4760] carrying IPv4 NLRI over an IPv6 Next Hop using IPv6 Next hop encoding [RFC8950] and BGP-LU [RFC8277] to advertise the control plane labeled IPv4 prefixes to a Route Reflector to which it is a client, which then advertises the labeled IPv4 Unicast prefixes over an eBGP Multihop Inter-AS peering to the route reflector in the Destination AS. The ASBR in the Source AS over the Inter-AS ASBR to ASBR link then uses eBGP to advertise the core underlay Labeled Unicast IPv6 PE loopabcks prefixes in the underlay to an ASBR in Destination AS, which in turn advertises the IPv6 PE loopabcks prefixes to a route reflector within the Destination AS of which it is a client which then advertises the PE loopabcks IPv6 prefixes to all the PE routers within the AS to establish an end to end LSP from ingress PE in the Source AS to egress PE in the Destination AS.

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

The considerations described for procedure (c) in Section 10 of [RFC4364] with respect to possible use of multi-hop eBGP connections via route-reflectors in different ASes, as well as with respect to the use of a third label in case the IPv6 /128 prefixes for the PE routers are NOT made known to the P routers, apply equally to this design for IPv4 underlay transport.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. The label stack on the ASBR to ASBR, PE to PE link is 2 labels deep, with the IPv6 tompost transport label IPv6 signaled LSP using BGP-LU IPv6 Labeled Unicast, IPv6 Address Family Identifier (AFI) IPv4 (value 2) Subsequent Address Family Identifier (SAFI) (value 4) and the route reflector to route reflector Multihop eBGP Peering next-hop-unchanged forwarding plane from ingress PE to egress PE loopback with unchanged next-hop is forwarded over the Inter-AS ASBR to ASBR PE-PE link, Bottom of Stack BGP-LU IPv4 labeled Unicast Service label, IPv4 Address Family Identifier (AFI) IPv4 (value 1) Subsequent Address Family Identifier (SAFI)(value 4) Thus IPv4 is not required to be activated on the Inter-AS ASBR to ASBR PE to PE links as IPv4 is tunnled through the IPv6 signaled LSP.

This 4PE design for procedure (c) in Section 10 of [RFC4364] requires that there be IPv6 label switched paths established across the ASes leading from a packet's ingress 4PE router to its egress 4PE router. Hence the considerations described for procedure (c) in Section 10 of [RFC4364], with respect to LSPs spanning multiple ASes, apply equally to this design for IPv4.

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

9. RFC 8950 Applicability to 4PE

The new MP-BGP extensions defined in [RFC8950] is used to support IPV4 islands over an IPv6 MPLS LDPv6 or SRv6 backbone. In this scenario the PE routers would use BGP Labeled unicast address family (BGP-LU) to advertise BGP with label binding and receive Labeled IPv4 NLRI in the MP_REACH_NLRI along with an IPv6 Next Hop from the Route Reflector (RR).

MP-BGP Reach Pseudo code:

If ((Update AFI == IPv4)

and (Length of next hop == 16 Bytes || 32 Bytes))

{

This is an IPv4 route, but

with an IPv6 next hop;

}

The MP_REACH_NLRI is encoded with:

- *AFI = 1
- *SAFI = 4
- *Length of Next Hop Network Address = 16 (or 32)
- *Network Address of Next Hop = IPv6 address of Next Hop whose RD is set to zero
- *NLRI = IPv4-VPN prefixes

During BGP Capability Advertisement, the PE routers would include the following fields in the Capabilities Optional Parameter:

- *Capability Code set to "Extended Next Hop Encoding"
- *Capability Value containing <NLRI AFI=1, NLRI SAFI=1, Nexthop AFI=2>

10. Implementations

4PE has been implemented by the following vendors

10.1. Cisco 4PE Implementation

4PE Context

Topmost label signaled by egress PE is implicit null by default for PHP mode for IPv6 LSP

Topmost label signaled by egress PE can be configured for explicit null for IPv6 LSP so that EXP Bits Diffserv QOS Pipe mode model

IPv4 prefixes tunneled over IPv6 LSP can be labeled or unlabeled

10.2. Juniper 4PE Implementation

4PE Context

Topmost label signaled by egress PE is implicit null by default PHP mode for IPv6 LSP

Topmost label signaled by egress PE can be configured for explicit null for IPv6 LSP so that EXP Bits Diffserv QOS Pipe mode model

IPv4 prefixes tunneled over IPv6 LSP can be labeled or unlabeled

10.3. Nokia 4PE Implementation

4PE Context

Topmost label signaled by egress PE is arbitrary label by default for IPv6 LSP

Topmost label signaled by egress PE can be configued for implicit null PHP mode for IPv6 LSP

Topmost label signaled by egress PE can be configured for explicit null for IPv6 LSP so that EXP Bits Diffserv QOS Pipe mode model

IPv4 prefixes tunneled over IPv6 LSP can be labeled or unlabeled

10.4. Huawei 4PE Implementation

4PE Context

Topmost label signaled by egress PE is implicit null by default PHP mode for IPv6 LSP

Topmost label signaled by egress PE can be configured for explicit null for IPv6 LSP so that EXP Bits Diffserv QOS Pipe mode model

IPv4 prefixes tunneled over IPv6 LSP can be labeled or unlabeled

11. IANA Considerations

There are not any IANA considerations.

12. 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.

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].

13. Acknowledgments

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

14. References

14.1. 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-01, 26 September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-sr-segtypes-ext-01>.

[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-25, 26 September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-25>.

[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>.

14.2. Informative References

[I-D.ietf-idr-dynamic-cap] Chen, E. and S. R. Sangli, "Dynamic Capability for BGP-4", Work in Progress, Internet-Draft, draft-ietf-idr-dynamic-cap-16, 21 October 2021, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-dynamic-cap-16>.

[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