BGP-MPLS VPN extension for IPv6 VPN

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months. Internet-Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet-Drafts as reference material or to cite them other than as a "working draft" or "work in progress."

To view the entire list of current Internet-Drafts, please check the "draft-abstracts.txt" listing contained in the Internet-Drafts Shadow Directories on ftp.is.co.za (Africa), ftp.nordu.net (Northern Europe), ftp.nis.garr.it (Southern Europe), munnari.oz.au (Pacific Rim), ftp.ietf.org (US East Coast), or ftp.isi.edu (US West Coast).

Distribution of this memo is unlimited.

Abstract

This document describes a method by which a Service Provider may use its packet switched backbone to provide Virtual Private Network services for its IPv6 customers. This method extends the "BGP/MPLS VPN" method [2547bis] for support of IPv6. In BGP/MPLS VPN, "Multiprotocol BGP" is used for distributing IPv4 VPN routes over the service provider backbone and MPLS is used to forward IPv4 VPN packets over the backbone. This document defines an IPv6 VPN address family and describes the corresponding route distribution in "Multiprotocol BGP". This document defines support of the IPv6 VPN service over both an IPv4 and an IPv6 backbone, and using various
tunnelling techniques over the core including MPLS, IPsec, IP-in-IP and GRE.

1. Introduction

This document adopts the definitions, acronyms and mechanisms described in [2547bis]. Unless otherwise stated, the mechanisms of [2547bis] apply and will not be re-described here.

A VPN is said to be an IPv6 VPN, when each site of this VPN is IPv6 capable and is natively connected over an IPv6 interface or sub-interface to the SP backbone via a Provider Edge device (PE).

A site may be both IPv4- and IPv6-capable. The logical interface on which packets arrive at the PE may determine the version, but alternatively the same logical interface MAY be used for both IPv4 and IPv6 in which case a per-packet header lookup determines the version. This document only concerns itself with handling of the IPv6 packets.

In a similar manner to how IPv4 VPN routes are distributed in [2547bis], BGP and its extensions are used to distribute routes from an IPv6 VPN site to all the other PE routers connected to a site of the same IPv6 VPN. PEs use VRFs to separately maintain the reachability information and forwarding information of each IPv6 VPN.

As it is done for IPv4 VPNs [2547bis], we allow each IPv6 VPN to have its own IPv6 address space, which means that a given address may denote different systems in different VPNs. This is achieved via a new address family, the VPN-IPv6 Address Family, in a fashion similar to the VPN-IPv4 address family definition given by [2547bis].

In addition to operation over MPLS Label Switched Paths (LSPs), the BGP/MPLS VPN solution is extended to allow operation over other tunnelling techniques including GRE tunnels, IP-in-IP tunnels [2547-GRE/IP] and IPsec tunnels [2547-IPsec]. In a similar manner, this document allows support of an IPv6 VPN service over MPLS LSPs as well as over other tunnelling techniques including GRE tunnels, IP-in-IP tunnels and IPsec tunnels.

This document allows support for an IPv6 VPN service over an IPv4 backbone as well as over an IPv6 backbone. The IPv6 VPN service supported is identical in both cases.

The IPv6 VPN solution defined in this document offers the following benefits:

- from both the Service Provider perspective and the customer
perspective, the VPN service that can be supported for IPv6 sites is identical to the one that can be supported for IPv4 sites;

- from the Service Provider perspective, operations of the IPv6 VPN service require the exact same skills, procedures and mechanisms as for the IPv4 VPN service;

- where both IPv4 VPNs and IPv6 VPN services are supported over an IPv4 core, the same single set of MP-BGP peering relationships and the same single PE-PE tunnel mesh MAY be used for both;

- independence of whether the core runs IPv4 or IPv6. So that the IPv6 VPN service supported before, and after a migration of the core from IPv4 to IPv6 is undistinguishable to the VPN customer.

2. The VPN Address Family

The BGP Multiprotocol Extensions [BGP-MP] allow BGP to carry routes from multiple "address families". We introduce the notion of the "VPN-IPv6 address family", that is similar to the VPN-IPv4 address family introduced in [2547bis].

2.1 The VPN-IPv6 Address Family

A VPN-IPv6 address is a 24-byte quantity, beginning with an 8-byte "Route Distinguisher" (RD) and ending with a 16-byte IPv6 address. If two VPNs use the same IPv6 address prefix (effectively denoting different physical systems), the PEs translate these into unique VPN-IPv6 address prefixes using different RDs. This ensures that if the same address is used in two different VPNs, it is possible to install two completely different routes to that address, one for each VPN.

The purpose of the RD is solely to allow one to create distinct routes to a common IPv6 address prefix, similarly to the purpose of the RD defined in [2547bis]. As it is possible per [2547bis], the RD can also be used to create multiple different routes to the very same system. This can be achieved by creating two different VPN-IPv6 routes that have the same IPv6 part, but different RDs. This allows BGP to install multiple different routes to the same system, and allows policy to be used to decide which packets use which route.

Note that VPN-IPv6 addresses and IPv6 addresses are always considered by BGP to be incomparable.

A VRF may have multiple equal-cost VPN-IPv6 routes for a single IPv6 address prefix. When a packet’s destination address is matched in a VRF against a VPN-IPv6 route, only the IPv6 part is actually matched.
When a site is IPv4- and IPv6-capable, the same RD MAY be used for the advertisement of IPv6 addresses and IPv4 addresses.

2.2. Encoding of Route Distinguishers

The RDs are encoded as per [2547bis]:

- TYPE field: 2 bytes
- VALUE field: 6 bytes

The interpretation of the VALUE field depends on the value of the TYPE field. As it is the case in [2547bis], 3 encodings can be used:

- TYPE field = 0 : the VALUE field consists of the following two subfields:
  * Administrator subfield: 2 bytes, it contains an Autonomous System Number
  * Assigned Number subfield: 4 bytes

- TYPE field = 1 : the VALUE field consists of the following two subfields:
  * Administrator subfield: 4 bytes, it contains a global IPv4 address
  * Assigned Number subfield: 2 bytes

- TYPE field = 2 : the VALUE field consists of the following two subfields:
  * Administrator subfield: 4 bytes, it contains a 4-byte Autonomous System Number
  * Assigned Number subfield: 2 bytes

3. VPN-IPv6 route distribution

3.1. Route Distribution Among PEs by BGP

As described in [2547bis], if two sites of a VPN attach to PEs which are in the same Autonomous System, the PEs can distribute VPN routes to each other by means of an (IPv4) iBGP connection between them. Alternatively, each PE can have an iBGP connection to a route reflector. Similarly, for IPv6 VPN route distribution, PEs can use an iBGP connection between them or use iBGP connections to a route reflector.

The PE routers:
(i) exchange, via MP-BGP [MP-BGP], reachability information for the IPv6 prefixes in the IPv6 VPNs.

(ii) announce themselves as the BGP Next Hop.

3.2 VPN IPv6 NLRI encoding

The advertising PE router MUST also assign and distribute MPLS labels with the IPv6 VPN routes. Essentially, PE routers do not distribute IPv6 VPN routes, but Labeled IPv6 VPN routes [MPLS-BGP]. When the advertising PE receives a packet that has this particular advertised label, the PE will pop the MPLS stack, and process the packet appropriately (i.e. forward it directly based on the label or perform a lookup in the corresponding IPv6-VPN context).

The BGP Multiprotocol Extensions [BGP-MP] are used to encode the MP_REACH NLRI. The AFI and SAFI fields MUST be set as follows:

- AFI: 2; for IPv6
- SAFI: 128; for MPLS labeled VPN-IPv6

The labeled VPN-IPv6 MP_REACH_NLRI itself is encoded as specified in [MPLS-BGP]. In the context of this extension, the prefix belongs to the VPN-IPv6 Address Family and thus consists of an 8-byte Route Distinguisher followed by an IPv6 prefix as specified in section 2 above.

3.2.1 BGP Next Hop encoding

3.2.1.1 BGP speakers with IPv6 connectivity

A BGP speaker SHALL advertise to its peers a Next Hop Network Address field containing a VPN-IPv6 address:

- whose RD is set to zero, and
- whose 16-byte IPv6 address is set to the global IPv6 address of the advertising PE.

potentially followed by another VPN-IPv6 address:

- whose RD is set to zero, and
- whose 16-byte IPv6 address is set to the link-local IPv6 address of the advertising PE.

The value of the Length of the Next Hop Network Address field in the
MP_REACH_NLRI attribute shall be set to 24 when only a global address is present, or 48 if a link-local address is also included in the Next Hop field.

The link-local address shall be included in the Next Hop field if and only if the advertising BGP speaker shares a common subnet with the peer the route is being advertised to [RFC2545].

In all other cases, a BGP speaker shall advertise to its peer in the Next Hop Network Address field only the global IPv6 address of the next hop.

As a consequence, a BGP speaker that advertises a route to an internal peer may modify the Network Address of Next Hop field by removing the link-local IPv6 address of the next hop.

An example scenario where both the global IPv6 address and the link-local IPv6 address shall be included in the BGP Next Hop address field is where the IPv6 VPN service is supported over a multi-AS backbone with redistribution of labeled VPN-IPv6 routes between ASBRs (of different AS) sharing a common IPv6 subnet: in that case, both the global IPv6 address and the link-local IPv6 address shall be advertised by the ASBRs.

3.2.1.2 BGP Speakers with IPv4 connectivity

A BGP speaker SHALL advertise to its peer a Next Hop Network Address field containing a VPN-IPv6 address:

- whose RD is set to zero, and

- whose 16-byte IPv6 address is encoded as an IPv4-mapped IPv6 address [V6ADDR] containing the IPv4 address of the advertising PE. This IPv4 address must be routable in the Service Provider’s backbone.

3.3 Route Target

The use of route target is specified in [2547bis] and applies to IPv6 VPNs. Encoding of the extended community attribute is defined in [BGP-EXTCOM].

3.4 BGP Capability Negotiation

In order for two PEs to exchange labelled IPv6 VPN NLRIs, they MUST use BGP Capabilities Negotiation to ensure that they both are capable of properly processing such NLRIs. This is done as specified in
[BGP-MP] and [BGP-CAP], by using capability code 1 (multiprotocol BGP), with AFI and SAFI values as specified above in section 3.2.

4. Determination of Tunnel Type

As mentioned earlier, this document allows support of IPv6 VPN service using various tunneling techniques in the core.

The tunneling type to use in the core for IPv6 VPN MAY be determined via configuration of PEs. Alternatively a mechanism to dynamically determine the tunneling type to use MAY be deployed (see for example [ENCAP-SIG] or [TUN-SAFI] and [SSA]).

5. Encapsulation

The ingress PE Router MUST tunnel IPv6 VPN data over the backbone towards the Egress PE router identified as the BGP Next Hop for the corresponding IPv6 VPN prefix.

When the 16-byte IPv6 address contained in the BGP Next Hop field is encoded as an IPv4-mapped IPv6 address (see section 3.2.1.2), the ingress PE MUST use IPv4 tunnelling.

When the 16-byte IPv6 address contained in the BGP Next Hop field is not encoded as an IPv4-mapped address (see section 3.2.1.2), the ingress PE MUST use IPv6 tunnelling.

Regardless of whether it is IPv4 or IPv6 tunnelling, the tunnelling type is determined as discussed above in section 4.

When a PE receives a packet from a CE, it looks up the packet’s IPv6 destination address in the VRF corresponding to that CE. This enables it to find a VPN-IPv6 route. The VPN-IPv6 route will have an associated MPLS label and an associated BGP Next Hop. First, this MPLS label is pushed on the packet. Then, this labelled packet is encapsulated into the tunnel for transport to the egress PE identified as the BGP Next Hop. Details of this encapsulation depend on the actual tunnelling technique as follows:

As with MPLS/BGP for IPv4 VPNs [2547-GRE/IP], when tunnelling is done using IPv4 tunnels or IPv6 tunnels (resp. IPv4 GRE tunnels or IPv6 GRE tunnels), encapsulation of the labelled IPv6 VPN packet results in an MPLS-in-IP (resp. MPLS-in-GRE) encapsulated packet as specified in [MPLS-in-IP/GRE].

As with MPLS/BGP for IPv4 VPNs, when tunnelling is done using an IPsec secured tunnel [2547-IPsec], encapsulation of the labelled IP6 VPN packet results in an MPLS-in-IP or MPLS-in-GRE encapsulated
packet [MPLS-in-IP/GRE]. The IPsec Transport Mode is used to secure this IPv4 or GRE tunnel from ingress PE to egress PE.

When tunnelling is done using IPv4 tunnels or IPv4 GRE tunnels (whether IPsec secured or not), the Ingress PE Router MUST use the IPv4 address which is encoded in the IPv4-mapped IPv6 address field of the BGP next hop field, as the destination address of the prepended IPv4 tunnelling header. It uses one of its IPv4 addresses as the source address of the prepended IPv4 tunneling header.

When tunnelling is done using IPv6 tunnels or IPv6 GRE tunnels (whether IPsec secured or not), the Ingress PE Router MUST use the IPv6 address which is contained in the IPv6 address field of the BGP next hop field, as the destination address of the prepended IPv6 tunnelling header. It uses one of its IPv6 addresses as the source address of the prepended IPv6 tunneling header.

When tunneling is done using MPLS LSPs, the LSPs can be established using any label distribution technique (LDP [LDP], RSVP-TE [RSVP-TE], ...). Nevertheless, to ensure interoperability among systems that implement this VPN architecture using MPLS LSPs as the tunneling technology, all such systems MUST support LDP [LDP].

When tunnelling is done using MPLS LSPs, the ingress PE Router directly pushes the LSP tunnel label on the label stack of the labelled IPv6 VPN packet (i.e. without prepending any IPv4 or IPv6 header). This pushed label corresponds to the LSP starting on the ingress PE Router and ending on the egress PE Router. The BGP Next Hop field is used to identify the egress PE router and as such the label to be pushed on the stack. In case the IPv6 address in the BGP Next Hop field is a IPv4-mapped IPv6 address, the embedded IPv4 address will determine the tunnel label to push on the label stack. In any other case, the IPv6 address in the BGP Next Hop field will determine the tunnel label to push on the label stack.

The bottom label is the label bound to the IPv6 VPN Prefix via BGP.

6. Address Scope

Since Link-local scope addresses are defined as uniquely identifying interfaces within (i.e., attached to) a single link only (see [SCOPE-ARCH]), those may be used on the PE-CE link but they are not supported for reachability across IPv6 VPN Sites and are never advertised via MP-BGP to remote PEs.

Global scope addresses are defined as uniquely identifying interfaces anywhere in the Internet. Global addresses are expected to be
commonly used within and across IPv6 VPN Sites. They are obviously supported by this IPv6 VPN solution for reachability across IPv6 VPN Sites and advertised via MP-BGP to remote PEs and processed without any specific considerations to their Global scope.

7. Multicast

Multicast operations is outside the scope of this document.

8. Inter-Provider Backbones

The same mechanisms described in [2547bis] can be used and have the same scalability properties.

9. Accessing the Internet from a VPN

The methods proposed by [2547bis] to access the global Internet can be used in the context of IPv6 VPNs and the global IPv6 Internet. Note however that if the IPv6 packets destined for the global IPv6 Internet need to traverse the SP backbone, and if this is an IPv4 only backbone, they must be tunnelled through that IPv4 backbone.

10. Management VPN

Where the IPv6 VPN service is supported over an IPv4 backbone, and where the Service Provider manages the CE, the Service Provider may elect to use IPv4 for communication between the management tool and the CE for such management purposes. In that case, regardless of whether a customer IPv4 site is actually connected to the CE or not, the CE is effectively part of an IPv4 VPN (i.e. it is attached to an IPv4 VRF) in addition to belonging to an IPv6 VPN. Considerations presented in [2547bis] on how to ensure that the management tool can communicate with such managed CEs from multiple VPNs without allowing undesired reachability across CEs of different VPNs, are applicable to the IPv4 VRF to which the CE attaches.

Where the IPv6 VPN service is supported over an IPv4 backbone, and where the Service Provider manages the CE, the Service Provider may elect to use IPv6 for communication between the management tool and the CE for such management purposes. Considerations presented in [2547bis] on how to ensure that the management tool can communicate with such managed CEs from multiple VPNs without allowing undesired reachability across CEs of different VPNs, are then applicable to the IPv6 VRF to which the CE attaches.

11. Security

The same security concerns as in [2547bis] are applicable.
12. Quality of Service

[2547bis] is applicable.

13. Acknowledgements

We would like to thank Gerard Gastaud who largely contributed to this document.

In Memoriam:

The authors would like to acknowledge the valuable contribution to this document from Tri T. Nguyen, who passed away in April 2002 after a sudden illness.

14. Normative References

[2547bis] Rosen et al., "BGP/MPLS VPNs", draft-ietf-l3vpn-rfc2547bis, work in progress

[BGP-EXTCOMM] Ramachandra, Tappan, Rekhter, "BGP Extended Communities Attribute", work in progress


[MPLS-BGP] Rekhter and Rosen, "Carrying Label Information in BGP4", RFC3107

[MPLS-ENCAPS] Rosen, Rekhter, Tappan, Farinacci, Fedorkow, Li, and Conta, "MPLS Label Stack Encoding", RFC3032


15. Informative References


[2547-GRE/IP] Rekhter and Rosen, "Use of PE-PE GRE or IP in RFC2547 VPNs", draft-ietf-l3vpn-gre-ip-2547, work in progress


16. Authors’ Addresses

Jeremy De Clercq
Alcatel
Fr. Wellesplein 1, 2018 Antwerpen, Belgium
E-mail: jeremy.de_clercq@alcatel.be

Dirk Ooms
Alcatel
Fr. Wellesplein 1, 2018 Antwerpen, Belgium
E-mail: dirk.ooms@alcatel.be

Marco Carugi
Nortel Networks S.A.
Parc d’activites de Magny-Les Jeunes Bois CHATEAUFORT
78928 YVELINES Cedex 9 - France
E-mail: marco.carugi@nortelnetworks.com

Francois Le Faucheur
Cisco Systems, Inc.
Village d’Entreprise Green Side - Batiment T3
400, Avenue de Roumanille
06410 Biot-Sophia Antipolis
France
E-mail: flefauch@cisco.com