INTERNET DRAFT
<draft-ooms-v6ops-bgp-tunnel-02.txt>

J. De Clercq, D. Ooms
Alcatel
S. Prevost
BTexact
F. Le Faucheur
Cisco

March, 2004
Expires September, 2004

Connecting IPv6 Islands across IPv4 MPLS Clouds with BGP

Status of this Memo

Abstract

This document explains how to interconnect IPv6 islands over an MPLS-enabled IPv4 cloud, including the exchange of IPv6 reachability information using BGP. The approach requires the edge routers that are connected to IPv6 islands to be Dual Stack MP-BGP-speaking routers while the core routers are only required to run IPv4 MPLS. The hosts in the IPv6 islands can use native IPv6 addresses. The approach uses MP-BGP over IPv4, relies on identification of the MP-BGP-speaking edge routers by their IPv4 address and uses IPv4-signaled MPLS LSPs that don't require any explicit tunnel configuration.

Table of Contents

Changes
ngtrans history (draft-ietf-ngtrans-bgp-tunnel-0x.txt)
00->01: editorial changes
extended section 4
01->02: editorial changes
added tunnel-specific considerations
added case of multiple IPv4 domains between IPv6 islands
added discussion on v6[v4]addresses in appendix A
02->03: complete rewrite: it turned out that two interpretations of the previous drafts existed, the two different interpretations are described explicitly in this version
03->04: renaming of the two approaches
editorial changes
clearly indicate which part requires standards track
04->05: added 5.1.3 to clarify how DS-BGP routers agree on tunnel type

v6ops history (draft-ooms-v6ops-bgp-tunnel-0x.txt)
05->00 individual submission: no changes. The document passed ngtrans last call early 2002, but the transfer to the IESG was postponed because of the reorg and closing down of ngtrans.
00->01 no changes
01->02 according to v6ops mailing list discussion, the scope of the document was restricted to the "MP-BGP over IPv4 using LSPs" approach.

1. Introduction

This document explains how to interconnect IPv6 islands over an IPv4 cloud, including the exchange of IPv6 reachability information using BGP. The approach requires the edge routers that are connected to IPv6 islands to be Dual Stack MP-BGP-speaking routers while the core routers are only required to run IPv4 MPLS. The hosts in the IPv6 islands can use native IPv6 addresses. The approach uses MP-BGP over IPv4, relies on identification of the MP-BGP-speaking edge routers by their IPv4 address and uses IPv4-signaled MPLS LSPs that don't require any explicit tunnel configuration.

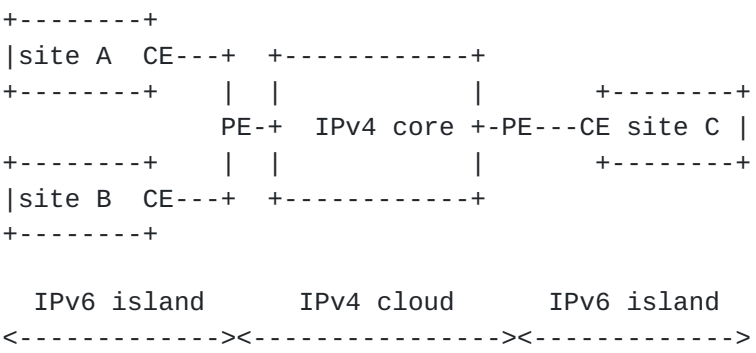
The use of the keywords "MUST", "MAY", etc. is in accordance with

[KEYWRD].

2. Terminology

The terminology of [IPV6] and [TRANS] applies to this document. We also use some of the terminology of [VPN].

In this document an 'IPv6 island' is an IPv6-upgraded network (which can be cross-AS). A typical example of an island would be a Customer IPv6 site connected via its IPv6 Customer Edge (CE) router to one (or more) Dual Stack Provider Edge (PE) router(s) of a Service Provider.



3. Applicability

The interconnection method described in this document typically applies to an ISP that has an MPLS network and is familiar with BGP (possibly already offering BGP/MPLS VPN services) and that wants to offer IPv6 services to some of its customers. However, the ISP may not (yet) want to upgrade its network core to native IPv6. With the mechanisms described here, the provider only has to upgrade some Provider Edge (PE) routers in some POPs to Dual Stack MP-BGP routers. These Dual Stack MP-BGP routers provide connectivity to IPv6 islands. They may also provide other services simultaneously (IPv4 connectivity, IPv4 L3VPN services, L2VPN services, etc.)

The ISP may also have access to the global IPv6 Internet. The ISP provides global IPv6 connectivity through its peering relationship with an upstream ISP, or by peering relationships with other IPv6 ISPs in the default free routing zone (DFZ).

A Dual Stack MP-BGP router in the provider's network is connected to an upstream IPv6 ISP or forms part of the IPv6 backbone network, such as the 6bone. The ISP advertises IPv6 reachability of its IPv6 allocated prefix using MP-BGP to its IPv6 upstream provider or into the IPv6 DFZ. The IPv6 prefixes received from the upstream provider or from the DFZ can be redistributed within the ISP using MP-BGP.

The interface between the edge router of the IPv6 island (Customer Edge router or CE) and the PE router is a native IPv6 interface which can be physical or logical. A routing protocol (IGP or EGP) may run between the CE router and the PE router for the distribution of IPv6 reachability information. Alternatively, static routes and/or a default route may be used on PE and CE to control reachability. An IPv6 island may connect to the provider network over more than one interface.

The methods in this document can be used for customers that already have an IPv4 service from the network provider and additionally require an IPv6 service, as well as for customers that require only IPv6 connectivity.

4. Description

Each IPv6 site is connected to at least one Dual Stack MP-BGP-speaking edge router that is located on the border with the IPv4 cloud. We refer to such a router as a DS-BGP router. The DS-BGP router MUST have at least one IPv4 address on the IPv4 side and one IPv6 address on the IPv6 side. The IPv4 address MUST be routable in the IPv4 cloud.

The PE routers that are attached to IPv6 islands need to insert a route (normally a /32 IPv4 address prefix) providing reachability to themselves (i.e. to their "IPv4 address") into the IGP routing tables of the IPv4 backbone. This enables MPLS, at each node in the backbone network, to assign an MPLS label corresponding to the route to each PE router. As a result of this, every considered PE router knows which MPLS label to use to send packets to any other PE router. Note that an MPLS network offering BGP/MPLS IP VPN services already fulfills these requirements.

No extra routes will be injected in the IPv4 cloud.

We refer to the DS-BGP router receiving IPv6 packets from an IPv6 site as an Ingress DS-BGP router (relative to these IPv6 packets). We refer to a DS-BGP router sending IPv6 packets to an IPv6 site as an Egress DS-BGP router (relative to these IPv6 packets).

Interconnecting IPv6 islands over an IPv4 cloud requires following steps:

(1) Exchange IPv6 reachability information among DS-BGP Routers:

(1.a) The DS-BGP routers exchange, via MP-BGP [MP-BGP], IPv6 reachability information over the IPv4 cloud with their peers.

(1.b) In doing so, the Egress DS-BGP routers announce themselves as the BGP Next Hop. The Egress DS-BGP router conveys to its peer its IPv4 address as the BGP Next Hop.

(2) Tunnel IPv6 packets from Ingress DS-BGP Router to Egress DS-BGP Router: the Ingress DS-BGP router tunnels an IPv6 packet over the MPLS IPv4 cloud towards the Egress DS-BGP router identified as the BGP Next Hop in step (1.b) for the packet's destination IPv6 address.

With this approach, the DS-BGP routers MUST run MP-BGP over an IPv4 stack (MP-BGP/TCP/IPv4).

Since MP-BGP assumes that the BGP Next Hop is of the same address family as the NLRI, this IPv4 address needs to be embedded in an IPv6 format. The IPv4-mapped IPv6 address is defined in [V6ADDR] as an "address type used to represent the addresses of IPv4 nodes as IPv6 addresses", thus this precisely fits for the above purpose. Encoding the routable IPv4 address into a IPv4-mapped IPv6 address allows the remote DS-BGP router to automatically tunnel data over the IPv4 cloud to the destination IPv6 island. Indeed, the IPv4 address contained in the IPv4-mapped IPv6 BGP Next Hop identifies an MPLS LSP that leads from the ingress PE router to the egress PE router.

The IPv4 address of the MP-BGP next hop MUST be encoded as an IPv4-mapped IPv6 address.

The ingress DS-BGP Router MUST tunnel IPv6 data over the IPv4 LSP towards the Egress DS-BGP router identified by the IPv4 address advertised in the IPv4-mapped IPv6 address of the BGP Next Hop for the corresponding IPv6 prefix.

The MP-BGP AFI MUST be IPv6 (value 2). The MP-BGP SAFI is discussed below in the tunneling section.

When the number of PEs is not too high, PEs MAY peer in a meshed fashion. Otherwise Route Reflectors MAY be used.

The hosts in the IPv6 island MAY have native IPv6 addresses. This is different from e.g. 6to4 [6TO4], which requires that special addresses (6to4 addresses) are allocated to the IPv6 hosts.

5. Tunneling over MPLS LSPs

In this approach, the IPv4-mapped IPv6 addresses allow a DS-BGP router that has to forward a packet to automatically determine the IPv4 endpoint of the tunnel by looking at the MP-BGP routing information.

Note that even when the number of peers is high, the number of tunnels is not a scalability concern from an operational viewpoint since those are automatic tunnels and thus require no configuration.

Considerations on 'common tunneling techniques' in [TRANS] are valid for this approach.

The IPv4 MPLS LSPs can be established using any existing technique (LDP, RSVP-TE, ...).

When tunneling IPv6 packets over the IPv4 MPLS backbone, rather than successively prepend an IPv4 header and then perform label imposition based on the IPv4 header, the ingress DS-BGP Router MUST directly perform label imposition of the IPv6 header without prepending any IPv4 header. The (outer) label imposed corresponds to the IPv4 LSP starting on the ingress DS-BGP Router and ending on the egress DS-BGP Router.

While this approach could operate in some situations using a single level of labels, there are significant advantages in using a second level of labels which are bound to IPv6 prefixes via MP-BGP advertisements in accordance with [LABEL]. For instance, use of a second level label allows Penultimate Hop Popping (PHP) on the Label Switch Router (LSR) upstream of the egress DS-BGP router without any IPv6 capabilities/upgrade on the penultimate router even when the IPv6 packet is directly encapsulated in MPLS (without an IPv4 header); since it still transmits MPLS packets even after the PHP (instead of having to transmit IPv6 packets and encapsulate them appropriately). Also, an existing IPv4 LSP which is using "IPv4 Explicit NULL label" over the last hop (say because that LSP is already used to transport IPv4 traffic with the Pipe Diff-Serv Tunneling Model as defined in [MPLS-DS]) could not be used to carry IPv6 with a single label since the "IPv4 Explicit NULL label" can not be used to carry native IPv6 traffic (see [MPLS-STACK]), while it could be used to carry labeled IPv6 traffic (see [EXP-NULL]). Thus, this approach MUST be used with a second label, advertised with BGP in accordance with [LABEL].

The SAFI used in MP-BGP MUST be the "label" SAFI (4) or the "VPN" SAFI (128) depending on the procedures for allocating these labels. The 'bottom label' (i.e. the second label when no PHP is used, or the only remaining label when PHP is used) indicates to the Egress DS-BGP Router that the packet is an IPv6 packet. The bottom label advertised by the Egress DS-BGP Router with MP-BGP MAY be an arbitrary label value and MAY identify an IPv6 routing context or outgoing interface to send the packet to, or MAY be the IPv6 Explicit Null Label. An Ingress DS-BGP Router MUST be able to accept any such advertised label.

6. Crossing multiple IPv4 domains

When the IPv6 islands are separated by multiple IPv4 domains, two cases can be distinguished:

1. The border routers between the IPv4 domains are not DS-BGP routers, i.e they are IPv4-only BGP routers. The DS-BGP routers of the IPv6 islands from the different IPv4 domains will be configured as multi-hop MP-EBGP peers for the exchange of IPv6 reachability. Alternatively, where the total number of such DS-BGP routers is high, IPv6 reachability across domains can be achieved via MP-BGP connection of Route Reflectors in different domains. One direct inter-domain LSP per pair of such DS-BGP routers will effectively be created. Note that the exchange of IPv6 routes can only start after BGP has created IPv4 connectivity between the domains.

2. The border routers between the IPv4 domains are DS-BGP routers. Each of these border DS-BGP routers will peer with the DS-BGP routers in its domain and regular IPv6 routing will take place between the two domains. No inter-domain LSPs are used. There is effectively a separate mesh of LSPs across the DS-BGP Routers of each domain.

7. Security considerations

The extensions defined in this document allow BGP to propagate reachability information about IPv6 routes over an IPv4 core. 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.

Acknowledgement

We like to thank G. Gastaud who contributed to this document, and we like to thank Tri T. Nguyen, who was the first to come up with the idea described in this document, but who unfortunately passed away much too soon.

Normative References

[IPV6] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC2460.

[KEYWRD]S. Bradner, Key words for use in RFCs to Indicate Requirement Levels, RFC2119, March 1997.

[LABEL] Rekhter Y., E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[MP-BGP]T. Bates, R. Chandra, D. Katz, Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC2858.

[V6ADDR]Deering, S., and R. Hinden, "IP Version 6 Addressing Architecture", draft-ietf-ipngwg-addr-arch-v3-07.txt (work in progress).

Informative References

[EXP-NULL] Rosen, E., et al., "Removing a Restriction on the use of MPLS Explicit NULL", draft-rosen-mpls-explicit-null-01.txt, work in progress

[6TO4] B. Carpenter, K. Moore, "Connection of IPv6 domains via IPv4 Clouds", RFC3056, February 2001.

[MPLS-DS] Le Faucheur et al., "MPLS Support for DiffServ", RFC 3270

[MPLS-STACK]Rosen, E., et al., "MPLS Label Stack Encoding", RFC 3032

[ISATAP] F. Templin, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP), draft-ietf-ngtrans-isatap-02.txt (work in progress).

[TRANS] R. Gilligan & E. Nordmark, "Transition Mechanisms for IPv6 Hosts and Routers", RFC2893.

[V6VPN] Nguyen T., Gastaud G., De Clercq J., Ooms D.,"BGP-MPLS VPN extension for IPv6 VPN over an IPv4 infrastructure", draft-ietf-ppvpn-bgp-ipv6-vpn-01.txt> (work in progress).

[VPN] Rosen E., Rekhter Y., Brannon S., Chase C., De Clercq J., Hitchin P., Marshall , Srinivasan V., "BGP/MPLS VPNs", draft-ietf-ppvpn-rfc2547bis-00.txt (work in progress).

Authors' Addresses

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

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

Stuart Prevost
BTexact Technologies
Room 136 Polaris House, Adastral Park,
Martlesham Heath, Ipswich, Suffolk IP5 3RE, England
E-mail: stuart.prevost@bt.com

Francois Le Faucheur
Cisco Systems
Domaine Green Side, 400, Avenue de Roumanille, Batiment T3
06 410 BIOT, SOPHIA ANTIPOLIS, FRANCE
E-mail: flefauch@cisco.com