INTERNET DRAFT
<draft-ooms-v6ops-bgp-tunnel-03.txt>

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

April, 2004
Expires October, 2004

Connecting IPv6 Islands over IPv4 MPLS using IPv6 Provider Edge Routers (6PE)

Status of this Memo

Abstract

This document explains how to interconnect IPv6 islands over a Multi-Protocol Label Switching (MPLS)-enabled IPv4 cloud. This approach relies on IPv6 Provider Edge routers (6PE) which are Dual Stack in order to connect to IPv6 islands and to the MPLS core which is only required to run IPv4 MPLS. The 6PE routers exchange the IPv6 reachability information transparently over the core using the Multi-Protocol Border Gateway Protocol (MP-BGP) over IPv4. In doing so, the BGP Next Hop field is used to convey the IPv4 address of the 6PE router so that dynamically established IPv4-signaled MPLS Label Switched Paths (LSPs) can be used without explicit tunnel configuration.

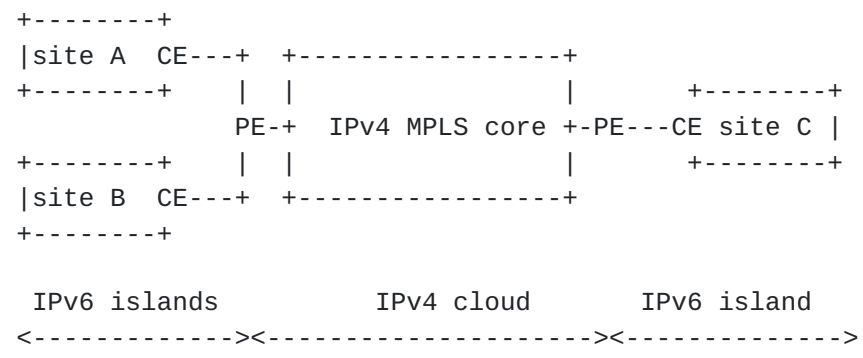
1. Introduction

There are several approaches for providing IPv6 connectivity over an MPLS core network [ISPSCEN]: (i) having the MPLS networks deploy native IPv6 support, (ii) require that MPLS networks support setting up IPv6 LSPs and set up IPv6 connectivity by using either these or configured tunneling, (iii) use only configured tunneling over IPv4 LSPs, or (iv) use the IPv6 Provider Edge (PE) approach.

This document specifies operations of the 6PE approach for interconnection of IPv6 islands over an IPv4 MPLS cloud. 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 approach uses MP-BGP over IPv4, relies on identification of the 6PE routers by their IPv4 address and uses IPv4-signaled MPLS LSPs that don't require any explicit tunnel configuration.

Throughout this document, the terminology of [IPV6] and [VPN] is used.

In this document an 'IPv6 island' is an IPv6-upgraded network. A typical example of an island would be a customer's 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. These Provider Edge routers are connected to an MPLS core network.



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 IPv6 nor use only configured IPv6-over-IPv4 tunnelling. With the 6PE approach described here, the provider only has to upgrade some Provider Edge (PE) routers to Dual Stack MP-BGP routers (6PE routers) while leaving the IPv4 MPLS core routers untouched. These 6PE routers provide

connectivity to IPv6 islands. They may also provide other services simultaneously (IPv4 connectivity, IPv4 L3VPN services, L2VPN services, etc.)

The ISP must obtain IPv6 connectivity to its peers and upstreams using means outside of the scope of this memo, and have its 6PE routers readvertise it over the MPLS core with MP-BGP.

The interface between the edge router of the IPv6 island (Customer Edge (CE) router) and the 6PE 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 6PE router for the distribution of IPv6 reachability information. Alternatively, static routes and/or a default route may be used on the 6PE router and the CE router 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.

The scenario is also described in [ISPSCEN].

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [KEYWRD].

2. Protocol Overview

Each IPv6 site is connected to at least one Dual Stack MP-BGP-speaking Provider Edge router that is located on the border of the IPv4 MPLS cloud. We call such a router a 6PE router. The 6PE router MUST have at least one IPv4 address on the IPv4 side and at least one IPv6 address on the IPv6 side. The IPv4 address MUST be routable in the IPv4 cloud.

The 6PE 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 6PE router. As a result of this, every considered 6PE router knows which MPLS label to use to send packets to any other 6PE 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 call the 6PE router receiving IPv6 packets from an IPv6 site an Ingress 6PE router (relative to these IPv6 packets). We call a 6PE router forwarding IPv6 packets to an IPv6 site an Egress 6PE router (relative to these IPv6 packets).

Interconnecting IPv6 islands over an IPv4 MPLS cloud takes place through the following steps:

(1) Exchange IPv6 reachability information among 6PE routers with MP-BGP [MP-BGP]:

> The 6PE routers MUST exchange the IPv6 prefixes over MP-BGP sessions running over IPv4. In doing so, the 6PE routers convey their IPv4 address as the BGP Next Hop for the advertised IPv6 prefixes. Since MP-BGP assumes that the BGP Next Hop is of the same address family as the NLRI, the 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" which precisely fits the above purpose. Therefore, the IPv4 address of the egress 6PE router MUST be encoded as an IPv4-mapped IPv6 address in the BGP Next Hop field.

(2) Tunneling IPv6 packets from Ingress 6PE router to Egress 6PE router via MPLS LSPs:

> The Ingress 6PE router MUST tunnel IPv6 data over the IPv4 LSP towards the Egress 6PE 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 used in step (1) above MUST be IPv6 (value 2). The MP-BGP SAFI is discussed below in the tunneling section.

As required by BGP specification, PE routers must form a full peering mesh unless Route Reflectors are used.

3. Tunneling over MPLS LSPs

In this approach, the IPv4-mapped IPv6 addresses allow a 6PE 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 tunnels are set up automatically.

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 6PE 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 6PE Router and ending on the egress 6PE 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 6PE 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]).

Therefore, 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). 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 6PE Router that the packet is an IPv6 packet. The bottom label advertised by the Egress 6PE 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 6PE Router MUST be able to accept any such advertised label.

4. Crossing Multiple IPv4 Autonomous Systems

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

1. The border routers between the IPv4 ASes are not 6PE routers, i.e they are IPv4-only BGP routers. The 6PE routers of the IPv6 islands from the different IPv4 ASes will be configured as multi-hop MP-EBGP peers for the exchange of IPv6 reachability. Alternatively, where the total number of such 6PE routers is high, IPv6 reachability across ASes can be achieved via MP-BGP connection of Route Reflectors in different ASes. Labeled IPv4 routes for the 6PE routers are exchanged across ASes so that direct inter-AS LSPs can be used to tunnel traffic across ASes from ingress 6PE router to egress 6PE router. Note that the exchange of IPv6 routes can only start after BGP has created IPv4 connectivity between the ASes.

2. The border routers between the IPv4 ASes are 6PE routers. Each of these border 6PE routers will peer with the 6PE routers in its AS and regular IPv6 routing will take place between the two ASes. No inter-AS LSPs are used. There is effectively a separate mesh of LSPs across the 6PE Routers of each AS.

5. Security Considerations

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

Acknowledgements

We wish to thank Gerard Gastaud and Eric Levy-Abegnoli who contributed to this document, and we wish to thank Tri T. Nguyen who initiated this document, but who unfortunately passed away much too soon. We also thank Pekka Savola for his valuable comments and suggestions.

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", RFC 2858.

[V6ADDR] Deering, S., and R. Hinden, "IP Version 6 Addressing Architecture", RFC 3513

Informative References

[ISPSCEN] Lind M., et al., "Scenarios and Analysis for Introducing IPv6 into ISP Networks", draft-ietf-v6ops-isp-scenarios-analysis, (work in progress)

[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

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

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

[V6VPN] Nguyen T., Gastaud G., De Clercq J., Ooms D.,"BGP-MPLS VPN extension for IPv6 VPN over an IPv4 infrastructure", draft-ietf-l3vpn-bgp-ipv6 (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-l3vpn-rfc2547bis (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

APPENDIX A

[RFC-editor note: remove before publication]

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.
02->03 adopted various comments

Intellectual Property Statement

Full Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.