INTERNET-DRAFT
<draft-ooms-v6ops-bgp-tunnel-04.txt>

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

October, 2004
Expires April, 2005

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.

Requirements

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].

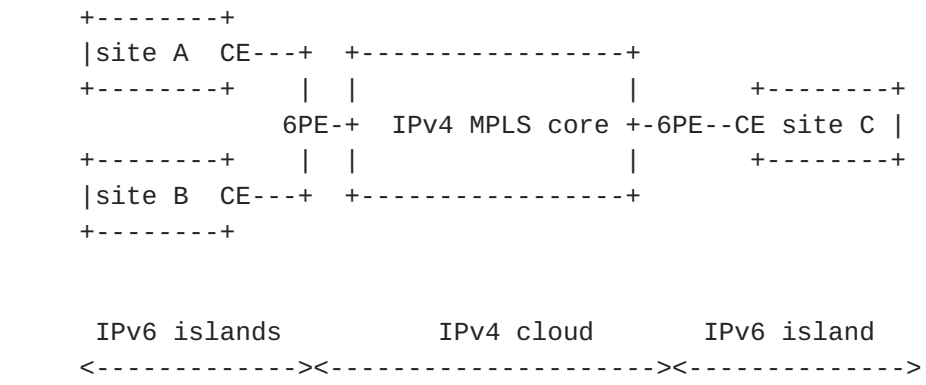
1. Introduction

There are several approaches for providing IPv6 connectivity over an MPLS core network [ISPSCEN] including (i) requiring that MPLS networks support setting up IPv6-signaled LSPs and set up IPv6 connectivity by using those, (ii) use only configured tunneling over IPv4-signaled LSPs, or (iii) use the IPv6 Provider Edge (6PE) 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 a network running native IPv6 as per [IPv6]. A typical example of an IPv6 island would be a customer's IPv6 site connected via its IPv6 Customer Edge (CE) router to one (or more) Dual Stack Provider Edge router(s) of a Service Provider. These IPv6 Provider Edge routers (6PE) are connected to an IPv4 MPLS core network.



The interconnection method described in this document typically applies to an ISP that has an IPv4 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 IPv6-over-IPv4 tunnelling. With the 6PE approach described here, the provider only has to upgrade some Provider Edge (PE) routers to Dual Stack operations so they behave as 6PE routers (and route reflectors if those are used for exchange of IPv6 reachability among 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.). Also with the 6PE approach, no tunnels need to be explicitly configured, and no IPv4 headers need to be inserted in front of the IPv6 packets.

The ISP obtains IPv6 connectivity to its peers and upstreams using means outside of the scope of this memo, and its 6PE routers readvertise it over the IPv4 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 6PE approach described 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].

Note that the 6PE approach specified in this document provides global

IPv6 reachability. Support of IPv6 VPNs is not within the scope of this document and is addressed in [V6VPN].

2. Protocol Overview

Each IPv6 site is connected to at least one 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 be dual stack IPv4 and IPv6. The 6PE router MUST be configurable with at least one IPv4 address on the IPv4 side and at least one IPv6 address on the IPv6 side. The configured IPv4 address needs to be routable in the IPv4 cloud, and there needs to be a label bound via an IPv4 label distribution protocol to this IPv4 route.

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 need to 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-v6]:

The 6PE routers MUST exchange the IPv6 prefixes over MP-BGP sessions as per [MP-BGP-v6] running over IPv4. The MP-BGP AFI used MUST be IPv6 (value 2). 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. In addition, the 6PE MUST bind a label to the IPv6 prefix as per [LABEL]. The SAFI used in MP-BGP MUST be the "label" SAFI (value 4) as defined in [LABEL]. Rationale for this and label allocation policies are discussed in section 3.

(2) Transport IPv6 packets from Ingress 6PE router to Egress 6PE router over IPv4-signaled LSPs:

The Ingress 6PE router MUST forward IPv6 data over the IPv4-signaled 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.

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

3. Transport over IPv4-signaled LSPs and IPv6 label binding

In this approach, the IPv4-mapped IPv6 addresses allow a 6PE router that has to forward an IPv6 packet to automatically determine the IPv4-signaled LSP to use for a particular IPv6 destination by looking at the MP-BGP routing information.

The IPv4-signaled 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 MUST correspond to the IPv4-signaled LSP starting on the ingress 6PE Router and ending on the egress 6PE Router.

While this approach could conceptually 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 IPv4 Label Switch Router (LSR) upstream of the egress 6PE router without any IPv6 capabilities/upgrade on the penultimate router; this is because it still transmits MPLS packets even after the PHP (instead of having to transmit IPv6 packets and encapsulate them appropriately).

Also, an existing IPv4-signaled 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]).

This is why a second label is always used with the 6PE approach.

The label bound by MP-BGP to the IPv6 prefix indicates to the Egress 6PE Router that the packet is an IPv6 packet. This label advertised by the Egress 6PE Router with MP-BGP MAY be an arbitrary label value which identifies 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

This section discusses the case where two IPv6 islands are connected to different Autonomous Systems.

Like in the case of multi-AS backbone operations for IPv4 VPNs described in section 10 of [VPN], three main approaches can be distinguished:

(a) EBGP redistribution of IPv6 routes from AS to neighboring AS

This approach is the equivalent for exchange of IPv6 routes to procedure (a) described in section 10 of [VPN] for the exchange of VPN-IPv4 routes.

In this approach, the 6PE routers use IBGP (according to [MP-BGP-v6] and [LABEL] and as described in this document for the single-AS situation) to redistribute labeled IPv6 routes either to an Autonomous System Border Router (ASBR) 6PE router, or to a route reflector of which an ASBR 6PE router is a client. The ASBR then uses EBGP to redistribute the (non-labeled) IPv6 routes to an ASBR in another AS, which in turn distributes them to the 6PE routers in that AS as described earlier in this specification, or perhaps to another ASBR which in turn distributes them etc.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. IPv6 needs to be activated on the inter-ASBR links and each ASBR 6PE router has at least one IPv6 address on the interface to that link.

No inter-AS LSPs are used. There is effectively a separate mesh of

LSPs across the 6PE routers within each AS.

In this approach, the ASBR exchanging IPv6 routes may peer over IPv6 or over IPv4. The exchange of IPv6 routes MUST be carried out as per [MP-BGP-v6].

Note that the peering ASBR in the neighboring AS to which the IPv6 routes were distributed with EBGP, should in its turn redistribute

these routes to the 6PEs in its AS using IBGP and encoding its own IPv4 address as the IPv4-mapped IPv6 BGP Next Hop.

(b) EBGP redistribution of labeled IPv6 routes from AS to neighboring AS

This approach is the equivalent for exchange of IPv6 routes to procedure (b) described in section 10 of [VPN] for the exchange of VPN-IPv4 routes.

In this approach, the 6PE routers use IBGP (as described earlier in this document for the single-AS situation) to redistribute labeled IPv6 routes either to an Autonomous System Border Router (ASBR) 6PE router, or to a route reflector of which an ASBR 6PE router is a client. The ASBR then uses EBGP to redistribute the labeled IPv6 routes to an ASBR in another AS, which in turn distributes them to the 6PE routers in that AS as described earlier in this specification, or perhaps to another ASBR which in turn distributes them etc.

There may be one, or multiple, ASBR interconnection(s) across any two ASes. IPv6 may or may not be activated on the inter-ASBR links.

This approach requires that there be label switched paths established across ASes. Hence the corresponding considerations described for procedure (b) in section 10 of [VPN] apply equally to this approach for IPv6.

In this approach, the ASBR exchanging IPv6 routes may peer over IPv4 or IPv6 (in which case, IPv6 obviously needs to be activated on the inter-ASBR link). When peering over IPv6, the exchange of labeled IPv6 routes MUST be carried out as per [MP-BGP-v6] and [LABEL]. When peering over IPv4, the exchange of labeled IPv6 routes MUST be carried out as per [MP-BGP-v6] and [LABEL] with encoding of the IPv4 address of the ASBR as an IPv4-mapped IPv6 address in the BGP Next Hop field.

(c) Multihop EBGP redistribution of labeled IPv6 routes between source and destination ASes, with EBGP redistribution of labeled IPv4 routes from AS to neighboring AS.

This approach is the equivalent for exchange of IPv6 routes to procedure (c) described in section 10 of [VPN] for exchange of VPN-IPv4 routes.

In this approach, IPv6 routes are neither maintained nor distributed by the ASBR routers. The ASBR routers need not be dual stack and may be IPv4/MPLS-only routers. An ASBR needs to maintain labeled IPv4 /32

routes to the 6PE routers within its AS. It uses EBGP to distribute these routes to other ASes. ASBRs in any transit ASes will also have to use EBGP to pass along the labled IPv4 /32 routes. This results in the creation of an IPv4 label switched path from the ingress 6PE router to the egress 6PE router. Now 6PE routers in different ASes can establish multi-hop EBGP connections to each other over IPv4, and can exchange labeled IPv6 routes (with an IPv4-mapped IPv6 BGP Next Hop) over those connections.

IPv6 need not be activated on the inter-ASBR links.

The considerations described for procedure (c) in section 10 of [VPN] 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 IPv4 /32 routes for the (6)PE routers are NOT made known to the P routers, apply equally to this approach for IPv6.

This approach requires that there be IPv4 label switched paths established across the ASes leading form a packet's ingress 6PE router to its egress 6PE router. Hence, the considerations described for procedure (c) in section 10 of [VPN] with respect to LSPs spanning multiple ASes apply equally to this approach for IPv6.

Note also that the exchange of IPv6 routes can only start after BGP has created IPv4 connectivity between the ASes.

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.

For the inter-AS distribution of IPv6 routes 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 [MP-BGP-v6].

For the inter-AS distribution of IPv6 routes according to case (b) and (c) of section 4 of this document, the procedures require that there be label switched paths established across the AS boundaries. Hence the appropriate trust relationships must exist between and among the set of ASes along the path. Care must be taken to avoid "label spoofing". To this end an ASBR 6PE SHOULD only accept labeled

packets from its peer ASBR 6PE if the topmost label is a label that it has explicitly signaled to that peer ASBR 6PE.

Note that for the inter-AS distribution of IPv6 routes according to case (c) of section 4 of this document, label spoofing may be more difficult to prevent. Indeed, the MPLS label distributed with the IPv6 routes via multi-hop EBGP is directly sent from the egress 6PE to ingress 6PEs in an other AS (or through route reflectors). This label is advertised transparently through the AS boundaries. When the egress 6PE that sent the labeled IPv6 routes receives a data packet that has this particular label on top of its stack, it may not be able to verify whether the label was pushed on the stack by an ingress 6PE that is allowed to do so. As such one AS may be vulnerable to label spoofing in a different AS. The same issue equally applies to the option (c) of section 10 of [VPN]. Just like it is the case for [VPN], addressing this particular security issue is for further study.

IANA Considerations

This document has no actions for IANA.

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.

[MP-BGP-v6] Marques P., et al., "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545.

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

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

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

[V6VPN] De Clercq J., Ooms D., Carugi M., Le Faucheur F., "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

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

Dirk Ooms
OneSparrow
Belegstraat 13, 2018 Antwerpen, Belgium
E-mail: dirk@onesparrow.com

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
03->04 clean-up of the requirements terminology
clarification of section 4

Intellectual Property Statement

Full Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the

Internet Society.