Inter-AS Option D for BGP/MPLS IP VPN
draft-mapathak-interas-ab-01.txt

Abstract

   This document describes a new option known as an Inter-AS option D to
   the 'Multi-AS Backbones' section of [RFC4364]. This option combines
   VPN VRFs at the Autonomous System Border Router (ASBR) as described
   in 'Option A' with the distribution of labeled VPN-IP routes as
   described in 'Option B'. In addition, this option allows for a data
   plane consisting of two methods of traffic forwarding between
   attached ASBR pairs.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

MPLS VPN providers often need to inter-connect different ASes to
provide VPN services to customers. This requires the setting up of
Inter-AS connections at ASBRs. The inter-AS connections may or may
not be between different providers. The mechanisms to set up inter-

as connections are described in [RFC4364]. Of particular interest
for this draft are the ones documented in section 10 of [RFC4364].

For the option described in section 10, part (a) of [RFC4364],
commonly referred to as Option A, peering ASBRs are connected by
multiple sub-interfaces, with at least one interface for each VPN
that spans the two ASes. Each ASBR acts as a PE, and thinks that the
other ASBR is a CE. The ASBRs associate each sub-interface with a
VRF and a BGP session is established per sub-interface to signal IP
(unlabeled) prefixes. As a result, traffic within the VPN VRFs is
IP. The advantage of this option is that the VPNs are isolated from
each other and since the traffic is IP, QoS mechanisms that operate
on IP traffic can be applied to achieve customer SLAs. The drawback
of this option is that there needs to be one BGP session per sub-
interface (and at least one sub-interface per VPN), which can be a
potential scalability concern if there are a large number of VRFs.

For the option described in section 10, part (b) of [RFC4364],
commonly referred to as Option B, peering ASBRs are connected by one
or more sub-interfaces that are enabled to receive MPLS traffic. An
MP-BGP session is used to distribute the labeled VPN prefixes between
the ASBRs. Therefore, the traffic that flows between them is
labeled. The advantage of this option is that it's more scalable, as
there is no need to have one sub-interface and BGP session per VPN.
The drawback of this option is that QoS mechanisms that can only be
applied to IP traffic cannot be used as the traffic is MPLS. There
is also no isolation between the VRFs.

The solution described in this draft aims to address the scalability
concerns of Option A by using a single BGP session to signal VPN
prefixes. In this solution, the forwarding connections between the
ASBRs are maintained on a per-VRF basis, while the control plane
information is exchanged using a single MP-BGP session.

If the solution is used between any attached ASBR pairs belonging to
separate Autonomous Systems (AS), then VRF based route filtering
policies via RTs is achieved directly between ASBR pairs, independent
of PE based RT filtering within an AS.

1.1. Requirements Language

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 [RFC2119].

2. **Scope**

   The Inter-AS VPN option described in this draft is applicable to
   both, the IPv4 VPN services described in [RFC4364] and the IPv6 VPN
   services defined in [VPN-IPv6]. It is NOT applicable to MVPN IPv4
   and MVPN IPv6 services defined in [RFC6513]. Support of existing
   'Multi-AS' options, along with the new techniques are beyond the
   scope of this document.

3. **Inter-AS Option D Reference Model**

   Figure 1 shows an arbitrary Multi-AS VPN interconnectivity scenario
   between Customer Edge routers. CE1 and CE3, interconnected by
   Service Providers SP1 and SP2, belong to the same VPN, say Red. CE2
   and CE4 belong to a different VPN, say Green. This example shows 3
   interfaces ('red', 'white' and 'green') between ASBR1 (belonging to
   SP1) and ASBR2 (belonging to SP2).

   Interface 'red' is a VRF attachment circuit associated to VRF1 (on
   ASBR1 and ASBR2) for VPN Red and is used to trasport labeled or
   native IP VPN traffic between VRF pairs. Similarly, interface
   'green' is a VRF attachment circuit associated to VRF2 (on ASBR1 and
   ASBR2) for VPN Green and is used to transport labeled or native IP
   VPN traffic between VRF pairs. Interface 'white' is not associated
   with any VRF instances i.e. this interface is 'global' in nature (in
   the context of the connected ASes) and carries as a minimum all ASBR
   exported VPN-IP routing updates.

   We shall use the term "private interface forwarding" to describe the
   model where packets for the "Red" VPN are forwarded on the "red"
   interface, while packets belonging to "Green" VPN are forwarded on
   the "green" interface. There are no BGP sessions running on the
   "red" and "green" interfaces; rather the 'white' interface carries
   all ASBR VPN-IP routing updates exported from VRF pairs. We shall
   use the term "shared interface forwarding" to describe the model
   where the "white" interface will be used to forward all the traffic
   between the ASBRs. For shared interface forwarding outside of a VRF
   context, interfaces 'red' and 'green' are not required. In addition
   to carrying all ASBR VPN-IP routing updates, interface 'white'
   transports labeled IP VPN traffic or native IP traffic. IP VPN
   packets entering or leaving the ASBR via this interface may be
   forwarded using normal MPLS mechanisms (e.g. through use of the LFIB)
   or through a lookup within a VRF that has been identified via MPLS
   label values.

```
                    CE1----\    /--------RR1--------\
                          \  /                    \
                        PE1---SP1 MPLS Cloud---ASBR1
                        /                    /  |  \
                CE2-----/              red /   |   \ green
                                        /  white  \
                                        \    |   /
                CE3-----\                \   |  /
                      \                   \  | /
                        PE2--SP2 MPLS Cloud---ASBR2
                        / \                    /
                  CE4----/   \--------RR2--------/
```
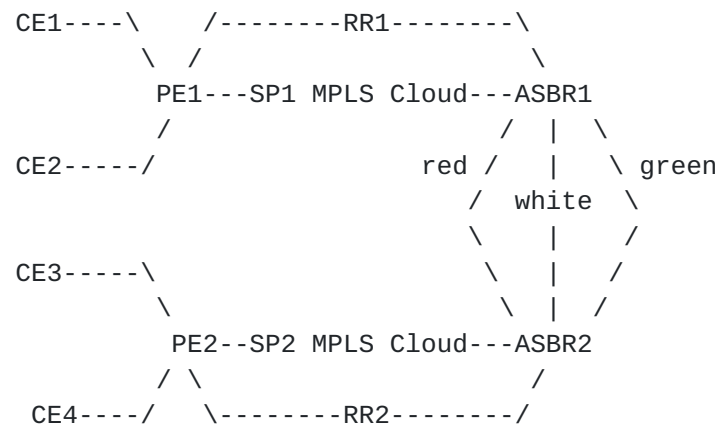
                   Figure 1

   In the diagram above:

   1. CE1 and CE3 belong to VPN Red.

   2. CE2 and CE4 belong to VPN Green.

   3. PE1 uses RDs RD-red1 and RD-green1 for VPN Red (VRF Red) and VPN
   Green (VRF Green) respectively.

   4. PE2 uses RDs RD-red2 and RD-green2 for VPN Red (VRF Red) and VPN
   Green (VRF Green) respectively.

   5. ASBR1 has VRFs Red and Green provisioned with RD-red3 and RD-
   green3 respectively.

   6. ASBR2 has VRFs Red and Green provisioned with RD-red4 and RD-
   green4 respectively.

   7. There are 3 interfaces between ASBR1 and ASBR2.

   8. On each ASBR, one interface is associated with VRF Red and one
   with VRF Green. These are the interfaces marked "red" and "green"
   respectively.

   9. There is a third interface over which there is an MP-BGP session
   between the ASBRs. This is the interface marked "white".

   10. VPN route importing is achieved by configuring the appropriate
   RTs.

   11. The PE and ASBR routers in each AS peer with a route-reflector
   in that AS.

The following sections describe in detail the different modes of
operation for Option D.

4. **Private Interface Operation without Carrier's Carrier (CSC)**

This section describes how route distribution and packet forwarding
takes place when using the private interface forwarding option
without the use of CSC, ie. the traffic sent between the private
interfaces is unencapsulated.

Route Distribution:

[The following description is for VPN Red, but Route Distribution for
VPN Green is exactly analogous to this]

1. CE1 advertises a prefix N to PE1.

2. PE1 advertises a VPN prefix RD-red1:N to RR1, which in turn
advertises it to ASBR1 via iBGP.

3. ASBR1 imports the prefix into VPN Red and creates a prefix RD-
red3:N.

4. ASBR1 advertises the imported prefix RD-red3:N to ASBR2. It sets
itself as the next-hop for this prefix and also allocates a local
label that is signaled with the prefix.

5. By default, ASBR1 does not advertise the source prefix RD-red1:N
to ASBR2. This advertisement is suppressed as the prefix is being
imported into an Option D VRF.

6. ASBR2 receives the prefix RD-red3:N and imports it into VPN Red
as RD-red4:N.

7. While installing the prefix into the VRF Red RIB table, ASBR2
sets the nexthop of RD-red4:N to ASBR1s interface address in VRF Red.
The routing context for this entry is also set to that of VRF Red.

8. While installing the MPLS forwarding entry for RD-red4:N, by
default, the label that was advertised by ASBR1 for the prefix is not
installed in the Forwarding Information Base. This enables the
traffic between the ASBRs to be IP.

9. ASBR2 advertises the imported prefix RD-Red4:N to RR2, which in
turn advertises it to PE2. It sets itself as the next-hop for this
prefix and also allocates a local label that is signaled as part of
the VPN-IP NLRI.

   10. By default, ASBR2 does not advertise the source prefix RD5:N to
   PE2. This advertisement is suppressed.

   11. PE2 imports the RD-red4:N into VRF Red as RD-red2:N.

   Packet Forwarding

   The packet forwarding would work just as it would in an Option A
   scenario:

   1. CE3 sends a packet destined for N to PE2.

   2. PE2 encapsulates the packet with the VPN label allocated by ASBR2
   and the IGP label (if any) needed to tunnel the packet to ASBR2.

   3. The packet arrives on ASBR2 with the VPN Label, ASBR2 pops the
   VPN Label and sends the packet as IP to ASBR1 on the "red" interface.

   4. The IP packet arrives at ASBR1 on the "red" interface. ASBR1
   then encapsulates the packet with the VPN Label allocated by PE1 and
   the IGP label needed to tunnel the packet to PE1.

   5. The packet arrives on PE1 with the VPN label; PE1 disposes the
   VPN label and forwards the IP packet to CE1.

5. Private Interface Forwarding with CSC

   Let's assume that VPN Red is used to provide VPN service to its
   customer carrier who in turn provides a VPN service to a customer.
   This implies that VPN RED is used to provide an LSP between the PE
   (PE3 and PE4) loopbacks of the baby carrier in the following
   topology:

                            /--------RR1--------\
                           /   \
       PE3---(CSC-)CE1---(CSC-)PE1---SP1 MPLS Cloud---ASBR1
                                              | |
                                              | |
                                        white| |red
                                              | |
                                              | |
       PE4---(CSC-)CE2---(CSC-)PE2---SP2 MPLS Cloud---ASBR2
                           \   /
                            \--------RR2----------/
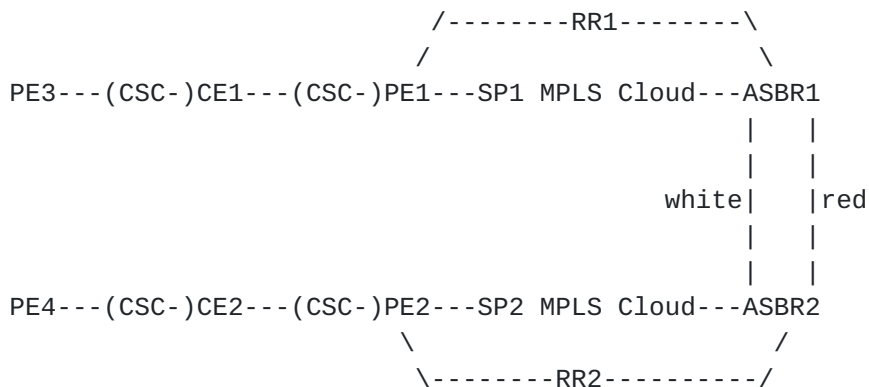

       Figure 2

Thus, let's assume that in the diagram above:

1.   CSC-PE1 uses RD RD-red1 for VPN Red (VRF Red).

2.   CSC-PE2 uses RD RD-red2 for VPN Red (VRF Red).

3.   ASBR1 has VRF Red provisioned with RD-red3.

4.   ASBR2 has VRF Red provisioned with RD-red4.

5.   There are 2 interfaces between ASBR1 and ASBR2.

6.   On each ASBR, one interface is associated with VRF Red. This is
the interface marked "red" in the Figure 2.

7.   There is a second interface over which there is an MP-BGP session
between the ASBRs. This interface is in the global context and is
marked "white" in the figure.

Route Distribution:

1.   CSC-CE1 advertises PE3s loopback N to PE1.

2.   CSC-PE1 advertises a VPN prefix RD-red1:N to RR1, which
advertises it to ASBR1 via MP-iBGP.

3.   ASBR1 imports the prefix into VPN Red and creates a prefix RD-
red3:N.

4.   ASBR1 advertises the imported prefix RD-red3:N to ASBR2. It sets
itself as the next-hop for this prefix and also allocates a local
label that is signaled as part of the VPN-IP NLRI.

5.   By default, ASBR1 does not advertise the source prefix RD-red1:N
to ASBR2. This advertisement is suppressed as the prefix is being
imported into an Option D VRF.

6.   ASBR2 receives the prefix RD-red3:N and imports it into VPN Red
as RD-red4:N.

7.   While installing the prefix into the VRF Red RIB table, ASBR2
sets the nexthop of RD-red4:N to ASBR1s interface address in VRF Red.
The nexthop routing context is also set to that of VRF Red.

8.   While installing the MPLS forwarding entry for RD-red4:N, the
outgoing label is installed in forwarding. This enables the traffic
between the ASBRs to be MPLS.

9. ASBR2 advertises the imported prefix RD-red4:N to RR2, which
advertises it to CSC-PE2. It sets itself as the next-hop for this
prefix and also allocates a local label that is signaled as part of
the VPN-IP NLRI.

10. By default, ASBR2 does not advertise the source prefix RD-red4:N
to PE2. This advertisement is suppressed.

11. PE2 imports the RD-red4:N into VRF Red as RD-red2:N.

Packet Forwarding:

1. PE4 sends a MPLS packet destined for N to CSC-CE2.

2. CSC-CE2 swaps the MPLS label and sends a packet destined for N to
CSC-PE2.

3. CSC-PE2 encapsulates the packet with the VPN label allocated by
ASBR2 and the IGP label needed (if any) to tunnel the packet to
ASBR2.

4. The packet arrives on ASBR2 with the VPN Label, ASBR2 swaps the
received VPN label with the outgoing label received from ASBR1 and
sends the MPLS packet on to the VRF Red interface.

5. The MPLS packet arrives at ASBR1 on the VRF red interface, ASBR1
then swaps the received MPLS label with a label stack consisting of
the VPN Label allocated by PE1 and the IGP label needed to tunnel the
packet to CSC-PE1.

6. The packet arrives on CSC-PE1 with the VPN label; PE1 disposes
the VPN label and forwards the MPLS packet to CSC-CE1.

7. CSC-CE1 in turn swaps the label and forwards the labeled packet
to PE3.

[6](#). Shared Interface Forwarding

This section describes how route distribution and packet forwarding
takes place when using the shared interface forwarding option. The
topology is the same as in Figure 1.

Route Distribution (VPN Red):

1. CE1 advertises a prefix N to PE1.

2. PE1 advertises a VPN prefix RD-red1:N to RR1, which advertises it
to ASBR1 via iBGP.

3. ASBR1 imports the prefix into VPN Red and creates a prefix RD-red3:N

4. ASBR1 advertises the imported prefix RD-red3:N to ASBR2. It sets itself as the next-hop for this prefix and also allocates a local label that is signaled with the prefix.

5. By default, ASBR1 does not advertise the source prefix RD-red1:N to ASBR2. This advertisement is suppressed as the prefix is being imported into an Option D VRF.

6. ASBR2 receives the prefix RD-red3:N and imports it into VPN Red as RD-red4:N

7. While installing the prefix into the VRF Red RIB table, ASBR2 retains the nexthop of RD-red4:N as received in the BGP update from ASBR1. This is the address of ASBR1's s shared interface address in the global table. The nexthop routing context is also left unchanged and corresponds to that of the global table.

8. While installing the MPLS forwarding entry for RD-red4:N, the outgoing label is installed in forwarding. This enables the traffic between the ASBRs to be MPLS.

9. ASBR2 advertises the imported prefix RD-red4:N to RR2, which advertises it to PE2. It sets itself as the next-hop for this prefix and also allocates a local label that is signaled as part of the VPN-IP NLRI.

10. By default, ASBR2 does not advertise the source prefix RD-red4:N to PE2. This advertisement is suppressed.

11. PE2 imports the RD-red4:N into VRF Red as RD-red2:N.

Packet Forwarding:

The packet forwarding would work just as it would in an Option B scenario:

1. CE3 sends a packet destined for N to PE2.

2. PE2 encapsulates the packet with the VPN label allocated by ASBR2 and the IGP label needed to tunnel the packet to ASBR2.

3. The packet arrives on ASBR2 with the VPN Label. ASBR2 swaps the received VPN label with the outgoing label received from ASBR1 and sends the MPLS packet on to the global shared link interface.

4. The MPLS packet arrives at ASBR1 on the global shared link
    interface. ASBR1 then swaps the received MPLS label with a label
    stack consisting of the VPN Label allocated by PE1 and the IGP label
    needed to tunnel the packet to PE1.

5. The packet arrives on PE1 with the VPN label; PE1 disposes the
    VPN label and forwards the IP packet to CE1.

7. Route Advertisement to External BGP Peers

//Keyur: Which figure.. how does this section differ from Section 8.
ASBR1 does route advertisement and VPN route processing using the
standard BGP-VPN rules. It processes the VRF Red RT extended
community attributes and learns the label bindings associated with
routes for VPN Red. VPN-IP routes are imported into VRF Red's Routing
Information Base (RIB) where their RT and RD attributes, assigned by
PE1 are removed.

ASBR1 VPN-IP routes are not advertised to RR1 as the original routes
applicable to VPN Red sourced by PE1 were received from an internal
BGP peer. Any installed routes that are not imported into VRF1 RIB
MAY be advertised to external BGP peers using the existing [RFC4364]
Multi-AS "Option B" techniques. Dependant on which packet forwarding
method is used, routes are then exported from VRFs and advertised
from ASBR1 to ASBR2 as described in the following sections.

7.1. Route Advertisement - Private interface forwarding

VPN-IP prefixes are advertised from ASBR1 to ASBR2 via a BGP session
that is in the global routing table context. This implies that the
advertised next-hop address is also reachable via the global routing
table context. In order to force traffic to be forwarded via an
interface 'red' that is in a VRF routing table context, VRF
forwarding entries need to be installed using a next-hop address that
is in VRF Red's (which resides on ASBR2) routing context. The
address of the next-hop could be the same as the global table address
of the remote ASBR (in this case ASBR1), although this would require
that the same IP address be used across all VRF attachment circuits
linking ASBR pairs.

Alternatively, if a Service Provider needs to number the VRF
interfaces differently from the global table VPN session, a
configuration method SHOULD be available so as to map the
corresponding global table VPNv4 neighbor address to an IP address
reachable in the given VRF.

ASBR1 exports routes associated to VPN Red from VRF Red's RIB to BGP
where RD and RT attributes, plus label bindings are attached to these

   routes. These labeled VPN-IP routes are advertised across interface
   'red' to ASBR2 via BGP, with a label value set to implicit-null and
   the 'S' bit set. The routes next-hop addresses is set either to
   ASBR1 (usually interface 'red') or an address reachable via interface
   'red'. ASBR2 imports the VRF Red's exported routes into VRF Red's
   RIB where the routes RT and RD attributes are removed. The next-hop
   of the imported routes is either set via a policy or left unchanged
   to an address in VRF Red's routing context. ASBR2 exports routes
   from VRF Red's RIB to BGP and attaches RT and RD attributes, as
   configured at VRF Red plus label bindings. Labeled VPN-IP routes are
   now advertised to PE2 via RR2 and so on.

7.2. Route Advertisement - Shared interface forwarding

   ASBR1 exports routes associated to VPN Red from VRF Red's RIB to BGP
   where RD and RT attributes, plus label bindings are attached to these
   routes. These labeled VPN-IP routes are advertised across interface
   'white' to ASBR2 via BGP, with a next-hop set to ASBR1. ASBR2
   imports the VRF Red exported routes into (its local) VRF Red RIB
   where the routes RT and RD attributes are removed. The imported
   routes next-hop is set to ASBR1, available via interface 'white'.
   ASBR2 exports routes from VRF Red's RIB to BGP and attaches RT and RD
   attributes, as configured at VRF Red plus label bindings. Labeled
   VPN-IP routes are now advertised to PE2 via RR2 and so on.

7.3. Route Advertisement to Internal BGP Peers

   All the received VPN-IP routes from an adjacent ASBR are imported
   into local VRFs on the receiving ASBR. The receiving ASBR needs to
   advertise these routes to its local IBGP sessions. The next-hop for
   these routes SHOULD be set to itself when the local ASRB advertises
   these routes to its IBGP sessions.

8. Option D Operation Requirements

8.1. Inter-AS IP VPN Route Distribution

   Routes received from internal or external peers that are imported
   into ASBR VRFs SHOULD NOT be readvertised to any BGP neighbors.
   Routes that are not imported into VRFs but are installed in the
   ASBR's global routing table MAY be readvertised using existing Option
   'B' techniques as described in the Multi-AS section of [RFC4364].
   The ASBR MUST be equipped with RT based filtering mechanisms that
   explicitly permit all or a subset of such RT values to be
   readvertised to its neighbors.

   VPN-IP routes that are converted by the ASBR MUST NOT be readvertised
   to the source peer of the route. It SHOULD be possible to remove/

   manipulate individual RT values when advertising routes on a per
   neighbor basis. This is useful where the SP wants to separate RT
   values advertised to EBGP peers from RT values advertised to IBGP
   peers.

8.2. Private Interface Forwarding Route Distribution

   For private interface forwarding, labeled VPN-IP routes advertised
   from ASBR to ASBR MUST have their next-hop set to an address within a
   VRF RIB. This address will usually be the VRF attachment circuit
   interface.

   If the Service Provider needs to number the VRF interfaces
   differently from the global table VPNv4 neighbor, a configuration
   method SHOULD be available so as to map the corresponding global
   table VPNv4 neighbor address to an IP address reachable in the given
   VRF. This route mapping policy SHOULD be configurable on both
   outbound and inbound peers.

8.3. Shared interface forwarding Route Distribution

   For shared interface forwarding outside of a VRF context, the next-
   hop must be a 'global' (non VRF RIB) address on an ASBR. This
   address will usually be the interface linking ASBR pairs.

9. Inter-AS Quality of Service for Option D

   It SHOULD be possible for the ASBR as a DS boundary node [DS-ARCH]
   operating traffic classification and conditioning functions to match
   on ingress and egress a combination of application (TCP, UDP port,
   RTP session, data pattern etc), IP Source Address, IP Destination
   Address or DS field per packet, per VRF or per VRF attachment circuit
   (in the case of private interface forwarding).

   Once matched, the packets Layer-2 header (if applicable), IP DSCP and
   MPLS EXP bits or combinations of the above should be capable of being
   re-marked, and optionally shaped per traffic stream, depending on the
   DS domain's Traffic Conditioning Agreement (TCA). This will assist
   where different DS domains have different TCA rules.

   For Private interface forwarding, the ASBR should be capable of
   forwarding explicit null labeled MPLS packets across VRF attachment
   circuits. This SHOULD assist with a pipe mode [DIFF-TUNNEL]
   operation of traffic conditioning behavior at the ASBR. MPLS based
   forwarding between attached ASBRs inherently should have these
   mechanisms built in.

10. Security Considerations

   This document doesnt not alter the underlying security properties of
   BGP based VPNs. In particular, the the private interface forwarding
   using a new Multi-AS option defined in this document has same
   security implications as Multi-AS option 'a'of [RFC4364]. The global
   interface forwarding using a new Multi-AS option defined in this
   document is outside the scope of this document.

   This document doesnt not alter the underlying security properties of
   BGP based VPNs for the shared interface forwaring using the new
   Multi-AS option. The security implications for this mechanism are
   same as Multi-AS option 'b' of [RFC4364].

11. Acknowledgements

   The authors wish to acknowledge the contributions of the authors of
   the original Option D draft: Marko Kulmala, Ville Hallivuori, Jyrki
   Soini, Jim Guichard, Robert Hanzl and Martin Halstead. The authors
   would like to thank Eric Rosen for his comments.

12. References

12.1. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private
              Networks (VPNs)", RFC 4364, February 2006.

12.2. Informative References

   [RFC2858] Bates, T., Rekhter, Y., Chandra, R., and D. Katz,
              "Multiprotocol Extensions for BGP-4", RFC 2858, June 2000.

Authors' Addresses

   Manu Pathak
   Affirmed Networks
   35 Nagog Park
   Acton, MA 01720
   USA

   Email: manu_pathak@affirmednetworks.com

Keyur Patel
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: keyupate@cisco.com

Arjun Sreekantiah
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: asreekan@cisco.com