                        Multicast VPN Using BIER
                         draft-ietf-bier-mvpn-04

Abstract

   The Multicast Virtual Private Network (MVPN) specifications require
   the use of multicast tunnels ("P-tunnels") that traverse a Service
   Provider's backbone network. The P-tunnels are used for carrying
   multicast traffic across the backbone. A variety of P-tunnel types
   are supported. Bit Index Explicit Replication (BIER) is a new
   architecture that provides optimal multicast forwarding through a
   "multicast domain", without requiring intermediate routers to
   maintain any per-flow state or to engage in an explicit tree-building
   protocol. This document specifies the protocol and procedures that
   allow MVPN to use BIER as the method of carrying multicast traffic
   over an SP backbone network.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   [RFC6513] and [RFC6514] specify the protocols and procedures that a
   Service Provider (SP) can use to provide Multicast Virtual Private
   Network (MVPN) service to its customers. Multicast tunnels are
   created through an SP's backbone network; these are known as
   "P-tunnels". The P-tunnels are used for carrying multicast traffic
   across the backbone. The MVPN specifications allow the use of
   several different kinds of P-tunnel technology.

Bit Index Explicit Replication (BIER) ([BIER_ARCH]) is an
architecture that provides optimal multicast forwarding through a
"multicast domain", without requiring intermediate routers to
maintain any per-flow state or to engage in an explicit tree-building
protocol. The purpose of the current document is to specify the
protocols and procedures needed in order to provide MVPN service
using BIER to transport the multicast traffic over the backbone.

Although BIER does not explicitly build and maintain multicast
tunnels, one can think of BIER as using a number of implicitly
created tunnels through a "BIER domain". In particular, one can
think of there as being one Point-to-Multipoint (P2MP) tunnel from
each "Bit Forwarding Ingress Router" (BFIR) to all the "Bit
Forwarding Egress Routers" (BFERs) in the BIER domain, where a BIER
domain is generally co-extensive with an IGP network. These
"tunnels" are not specific to any particular VPN. However, the MVPN
architecture provides protocols and procedures that allow the traffic
of multiple MVPNs to be aggregated on a single P-tunnel. In this
document, we specify how to use these multi-VPN aggregation
procedures to enable BIER to transport traffic from multiple MVPNs.

MVPN traffic must sometimes traverse more than one IGP domain,
whereas BIER only carries multicast traffic within a single IGP
domain. However, the MVPN specifications allow P-tunnels to be
"segmented", where the segmentation points may either be Autonomous
System Border Routers (ASBRs), as described in [RFC6514], or Area
Border Routers (ABRs), as described in [RFC7524]. As long as the
segmentation points are capable of acting as BFIRs and BFERs, BIER
can be used to provide some or all of the segments of a P-tunnel.

This revision of the document does not specify the procedures
necessary to support MVPN customers that are using BIDIR-PIM. Those
procedures will be added in a future revision.

This document uses the following terminology from [BIER_ARCH]:

o BFR: Bit-Forwarding Router.

o BFIR: Bit-Forwarding Ingress Router.

o BFER: Bit-Forwarding Egress Router.

This document uses the following terminology from [RFC6513]:

o MVPN: Multicast Virtual Private Network -- a VPN [RFC4364] in
   which multicast service is offered.

   o P-tunnel. A multicast tunnel through the network of one or more
      SPs. P-tunnels are used to transport MVPN multicast data

   o C-S: A multicast source address, identifying a multicast source
      located at a VPN customer site.

   o C-G: A multicast group address used by a VPN customer.

   o C-flow: A customer multicast flow. Each C-flow is identified by
      the ordered pair (source address, group address), where each
      address is in the customer's address space. The identifier of a
      particular C-flow is usually written as (C-S,C-G).
      Sets of C-flows can be identified by the use of the "C-*" wildcard
      (see [RFC6625]), e.g., (C-*,C-G).

   o I-PMSI A-D Route: Inclusive Provider Multicast Service Interface
      Auto-Discovery route. Carried in BGP Update messages, these
      routes are used to advertise the "default" P-tunnel for a
      particular MVPN.

   o S-PMSI A-D route: Selective Provider Multicast Service Interface
      Auto-Discovery route. Carried in BGP Update messages, these
      routes are used to advertise the fact that particular C-flows are
      bound to (i.e., are traveling through) particular P-tunnels.

   o PMSI Tunnel attribute (PTA). This BGP attribute carried is used
      to identify a particular P-tunnel. When C-flows of multiple VPNs
      is carried in a single P-tunnel, this attribute also carries the
      information needed to multiplex and demultiplex the C-flows.

   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. Use of the PMSI Tunnel Attribute

   As defined in [RFC6514], the PMSI Tunnel attribute is used to
   identify the particular P-tunnel to which one or more multicast flows
   are being assigned.

   The PMSI Tunnel attribute (PTA)contains the following fields:

   o "Tunnel Type". IANA is requested to assign a new tunnel type
      codepoint for "BIER". This codepoint will be used to indicate
      that the PMSI is instantiated by BIER.

   o "Tunnel Identifier". When the "tunnel type" field is "BIER", this
      field contains two subfields:

1. The first subfield is a single octet, containing the sub-
    domain-id of the sub-domain to which the BFIR will assign the
    packets that it transmits on the PMSI identified by the NLRI
    of the BGP I-PMSI or S-PMSI A-D route that contains this PTA.
    (How that sub-domain is chosen is outside the scope of this
    document.)

2. The second subfield is the BFR-Prefix (see [BIER ARCH]) of the
    originator of the route that is carrying this PTA. This will
    either be a /32 IPv4 address or a /128 IPv6 address. Whether
    the address is IPv4 or IPv6 can be inferred from the total
    length of the PMSI Tunnel attribute.

o "MPLS label". This field contains an upstream-assigned MPLS
   label. It is assigned by the router that originates the BGP route
   to which the PTA is attached. Constraints on the way in which the
   originating router selects this label are discussed below.

o "Flags". When the tunnel type is BIER, two of the flags in the
   PTA Flags field are meaningful. Details about the use of these
   flags can be found in Section 2.2.

   * "Leaf Info Required per Flow (LIR-pF)". This flag is
      introduced in [EXPLICIT_TRACKING]. A BFIR SHOULD NOT set this
      flag UNLESS it knows that all the BFERs in the BIER domain (or
      at least all the BFERs to which it needs to transmit) support
      this flag. (How this is known is outside the scope of this
      document.) Procedures for the use of this flag are given in
      Section 2.2.2

   * "Leaf Info Required Bit". See Section 2.2.1.

Note that if a PTA specifying "BIER" is attached to an I-PMSI or
S-PMSI A-D route, the route MUST NOT be distributed beyond the
boundaries of a BIER domain. That is, any routers that receive the
route must be in the same BIER domain as the originator of the route.
If the originator is in more than one BIER domain, the route must be
distributed only within the BIER domain in which the BFR-Prefix in
the PTA uniquely identifies the originator. As with all MVPN routes,
distribution of these routes is controlled by the provisioning of
Route Targets.

2.1. MPLS Label

Suppose an ingress PE originates two x-PMSI A-D routes, where we use
the term "x-PMSI" to mean "I-PMSI or S-PMSI". Suppose both routes
carry a PTA, and the PTA of each route specifies"BIER".

o If the two routes do not carry the same set of Route Targets
   (RTs), then their respective PTAs MUST contain different MPLS
   label values.

o If the ingress PE is supporting MVPN extranet ([RFC7900])
   functionality, and if the two routes originate from different
   VRFs, then the respective PTAs of the two routes MUST contain
   different MPLS label values.

o If the ingress PE is supporting the "Extranet Separation" feature
   of MVPN extranet (see Section 7.3 of [RFC7900], section), and if
   one of the routes carries the "Extranet Separation" extended
   community and the other does not, then the respective PTAs of the
   two routes MUST contain different MPLS label values.

o If segmented P-tunnels are being used, then the respective PTAs of
   the two routes MUST contain different MPLS label values, as long
   as the NLRIs are not identical. In this case, the MPLS label can
   be used by the BFER to identify the particular C-flow to which a
   data packet belongs, and this greatly simplifies the process of
   forwarding a received packet to its next P-tunnel segment. This
   is explained further below. See also Section 3.

When segmented P-tunnels are being used, an ABR or ASBR may receive,
from a BIER domain, an x-PMSI A-D route whose PTA specifies "BIER".
This means that BIER is being used for one segment of a segmented
P-tunnel. The ABR/ASBR may in turn need to originate an x-PMSI A-D
route whose PTA identifies the next segment of the P-tunnel. The
next segment may also be "BIER". Suppose an ASBR receives x-PMSI A-D
routes R1 and R2, and as a result originates x-PMSI A-D routes R3 and
R4 respectively, where the PTAs of each of the four routes specify
BIER. Then the PTAs of R3 and R4 MUST NOT specify the same MPLS
label, UNLESS both of the following conditions hold:

o R1 and R2 have the same "originating router" in their respective
   NLRIs.

o R1 and R2 specify the same MPLS label in their respective PTAs.

The ABR/ASBR MUST then program its dataplane such that a packet
arriving with the upstream-assigned label specified in route R1 is
transmitted with the upstream-assigned label specified in route R3,
and a packet arriving with the upstream-assigned label specified in
route R2 is transmitted with the label specified in route R4. Of
course, the data plane must also be programmed to encapsulate the
transmitted packets with an appropriate BIER header, whose BitString
is determined by the multicast flow overlay.

2.2. Explicit Tracking

   When using BIER to transport an MVPN data packet through a BIER
   domain, an ingress PE functions as a BFIR (see [BIER_ARCH]). The
   BFIR must determine the set of BFERs to which the packet needs to be
   delivered. This can be done in either of two ways:

   1. Using the explicit tracking mechanism based on the "Leaf Info
       Required" flag specified in [RFC6513] and [RFC6514]. This method
       is further described in Section 2.2.1.

   2. Using the explicit tracking mechanism based on the LIR-pF flag
       specified in [EXPLICIT_TRACKING]. This method, further described
       in Section 2.2.2, may be used if (and only if) segmented
       P-tunnels are not being used.

2.2.1. Using the LIR Flag

   To determine the set of BFERs to which the packets of a given C-flow
   must be sent, a BFIR MUST originate a (C-S,C-G) S-PMSI A-D route for
   the given C-flow. It MUST attach a PTA to that route, and MUST set
   the LIR flag in the PTA. Per [RFC6514], the BFERs that need to
   receive that C-flow will respond with (C-S,C-G) Leaf A-D routes. By
   matching the received Leaf A-D routes to the originated S-PMSI A-D
   routes, the originator of the S-PMSI A-D route determines the set of
   BFERs that need to receive the multicast data flow that is identified
   in the NLRI of S-PMSI A-D route.

   The PTA MAY specify a tunnel type ("BIER") and a non-zero MPLS label.
   (If it specifies one of these it MUST also specify the other.)
   Alternatively, the PTA MAY specify "no tunnel type" and a zero MPLS
   label. In this case, the tunnel type ("BIER") and non-zero MPLS
   label MUST be specified in an I-PMSI A-D route or in a wildcard
   S-PMSI A-D route that "matches" (according to the rules of [RFC6625])
   the C-flow in question.

2.2.2. Using the LIR-pF Flag

   If segmented P-tunnels are not being used, the BFIR can determine the
   set of BFERs that need to receive the packets of a given (C-S,C-G)
   C-flow as follows. The BFIR MUST originate a wildcard S-PMSI A-D
   route (either (C-*,C-*), (C-*,C-G), or (C-S,C-G) and the PTA of that
   route MUST the following settings:

   o The LIR-pF flag MUST be set;

   o The tunnel type MUST be set to "BIER;

o A non-zero MPLS label MUST be specified.

Per [EXPLICIT_TRACKING], a BFER that needs to receive (C-S,C-G)
traffic from the BFIR will respond with a Leaf A-D route.

A BFIR MUST NOT use this method of finding the set of BFERs needing
to receive a given C-flow unless it knows that all those BFERs
support the LIR-pF flag. How this is known is outside the scope of
this document.

This method greatly reduces the number of S-PMSI A-D routes that a
BFIR needs to originate; it can now originate as few as one such
route (a (C-*,C-*) S-PMSI A-D route), rather than one for each
C-flow. However, the method does not provide a way for the BFIR to
assign a distinct label to each C-flow. Therefore it cannot be used
when segmented P-tunnels are in use (see Section 3 for an
explanation).

Note: if a BFIR originates a (C-*,C-*) S-PMSI A-D route with the LIR-
pF flag set, but also originates a more specific wildcard route that
matches a particular (C-S,C-G), the BFERs will not originate Leaf A-D
routes for that (C-S,C-G) unless the LIR-pF flag is also set in the
more specific wildcard route. If the BFIR also originates a
(C-S,C-G) S-PMSI A-D route without the LIR flag set, the BFERs will
not originate Leaf A-D routes for that (C-S,C-G) unless the LIR flag
is also set in that route.

3. Data Plane

The MVPN application plays the role of the "multicast flow overlay"
as described in [BIER_ARCH].

3.1. Encapsulation and Transmission

To transmit an MVPN data packet, an ingress PE follows the rules of
[RFC6625] to find the S-PMSI A-D route or I-PMSI A-D route that is a
"match for transmission" for that packet. (In applying the rules of
[RFC6625], any S-PMSI A-D route with a PTA specifying "no tunnel
information" is ignored.) If the matching route has a PTA specifying
"BIER", the (upstream-assigned) MPLS label from that PTA is pushed on
the packet's label stack. Then the packet is encapsulated in a BIER
header and forwarded, according to the procedures of [BIER_ARCH] and
[BIER_ENCAPS]. (See especially Section 4, "Imposing and Processing
the BIER Encapsulation", of [BIER_ENCAPS].)

In order to create the proper BIER header for a given packet, the
BFIR must know all the BFERs that need to receive that packet. It
determines this by finding all the Leaf A-D routes that correspond to

the S-PMSI A-D route that is the packet's match for transmission.
There are two different cases to consider:

1. The S-PMSI A-D route that is the match for transmission carries a
    PTA that has the LIR flag set but does not have the LIR-pF flag
    set.

    In this case, the corresponding Leaf A-D routes are those whose
    "route key" field is identical to the NLRI of the S-PMSI A-D
    route.

2. The S-PMSI A-D route that is the match for transmission carries a
    PTA that has the LIR-pF flag.

    In this case, the corresponding Leaf A-D routes are those whose
    "route key" field is derived from the NLRI of the S-PMSI A-D
    route according to the procedures described in Section 5.2 of
    [EXPLICIT_TRACKING].

3.2. Disposition

When a BFER receives an MVPN multicast data packet that has been
BIER-encapsulated, the BIER layer passes the following information to
the multicast flow overlay:

o   The BFR-prefix corresponding to the sub-domain-id and BFIR-id in
    the BIER header.

o   The "payload", which is an MPLS packet whose top label is an
    upstream-assigned label. The BFR-prefix provides the "context" in
    which the upstream-assigned label is interpreted.

    Note that per [RFC5331], the context for an upstream-assigned
    label is the IP address of the label assigner, which in this case
    is the BFR-prefix of the BFIR.

By looking up the upstream-assigned label in the appropriate context,
the multicast flow overlay determines whether the BFER is an egress
PE for the packet.

Note that if segmented P-tunnels are in use, a BFER might be a
P-tunnel segmentation border router rather than an egress PE, or a
BFER might be both an egress PE and a P-tunnel segmentation border
router. Depending upon the role of the BFER for given packet, it may
need to follow the procedures of Section 3.2.1, the procedures of
Section 3.2.2, or both.

3.2.1. At a BFER that is an Egress PE

From looking up the packet's upstream-assigned label in the context
of the packet's BFIR-prefix, the egress PE determines the egress VRF
for the packet. From the IP header of the payload, the multicast
states of the VRF, the upstream-assigned label, and the BFR-prefix,
the egress PE can determine whether the packet needs to be forwarded
out one or more VRF interfaces.

3.2.2. At a BFER that is a P-tunnel Segmentation Boundary

When segmented P-tunnels are being used, a BFER that receives a BIER-
encapsulated MVPN multicast data packet may need to be forwarded on
its next P-tunnel segment. The choice of the next P-tunnel segment
for the packet depends upon the C-flow to which the packet belongs.
As long as the BFIR has assigned the MPLS label according to the
constraints specified in Section 2.1, the BFIR will have assigned
distinct upstream-assigned MPLS labels to distinct C-flows. The BFER
can thus select the proper "next P-tunnel segment" for a given packet
simply by looking up the upstream-assigned label that immediately
follows the BIER header.

4. Contributor Addresses

Below is a list of other contributing authors in alphabetical order:

IJsbrand Wijnands
Cisco Systems, Inc.
De Kleetlaan 6a
Diegem 1831
Belgium

Email: ice@cisco.com

5. Acknowledgments

The authors wish to thank Jeffrey Zhang for his ideas and
contributions to this work.

6. IANA Considerations

IANA is requested to assign a value for "BIER" from the "P-Multicast
Service Interface Tunnel (PMSI Tunnel) Tunnel Types" registry. The
reference should be this document.

7. Security Considerations

   The security considerations of [BIER_ARCH], [BIER_ENCAPS], [RFC6513]
   and [RFC6514] are applicable.

8. References

8.1. Normative References

   [BIER_ARCH]
              Wijnands, IJ., Rosen, E., Dolganow, A., Przygienda, T.,
              and S. Aldrin, "Multicast using Bit Index Explicit
              Replication", internet-draft draft-ietf-bier-architecture-
              03, January 2016.

   [BIER_ENCAPS]
              Wijnands, IJ., Rosen, E., Dolganow, A., Tantsura, J., and
              S. Aldrin, "Encapsulation for Bit Index Explicit
              Replication in MPLS Networks", internet-draft draft-ietf-
              bier-mpls-encapsulation-04.txt, April 2016.

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119,
              DOI 10.17487/RFC2119, March 1997,
              <http://www.rfc-editor.org/info/rfc2119>.

   [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private
              Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February
              2006, <http://www.rfc-editor.org/info/rfc4364>.

   [RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream
              Label Assignment and Context-Specific Label Space",
              RFC 5331, DOI 10.17487/RFC5331, August 2008,
              <http://www.rfc-editor.org/info/rfc5331>.

   [RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/
              BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February
              2012, <http://www.rfc-editor.org/info/rfc6513>.

   [RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP
              Encodings and Procedures for Multicast in MPLS/BGP IP
              VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012,
              <http://www.rfc-editor.org/info/rfc6514>.

   [RFC6625] Rosen, E., Ed., Rekhter, Y., Ed., Hendrickx, W., and R.
              Qiu, "Wildcards in Multicast VPN Auto-Discovery Routes",
              RFC 6625, DOI 10.17487/RFC6625, May 2012,
              <http://www.rfc-editor.org/info/rfc6625>.

8.2. **Informative References**

[EXPLICIT_TRACKING]
            Dolganow, A., Kotalwar, J., Rosen, E., and Z. Zhang,
            "Explicit Tracking with Wild Card Routes in Multicast
            VPN", internet-draft draft-ietf-bess-mvpn-expl-track-00,
            June 2016.

[RFC7524]   Rekhter, Y., Rosen, E., Aggarwal, R., Morin, T.,
            Grosclaude, I., Leymann, N., and S. Saad, "Inter-Area
            Point-to-Multipoint (P2MP) Segmented Label Switched Paths
            (LSPs)", RFC 7524, DOI 10.17487/RFC7524, May 2015,
            <http://www.rfc-editor.org/info/rfc7524>.

[RFC7900]   Rekhter, Y., Ed., Rosen, E., Ed., Aggarwal, R., Cai, Y.,
            and T. Morin, "Extranet Multicast in BGP/IP MPLS VPNs",
            RFC 7900, DOI 10.17487/RFC7900, June 2016,
            <http://www.rfc-editor.org/info/rfc7900>.

Authors' Addresses

   Eric C. Rosen (editor)
   Juniper Networks, Inc.
   10 Technology Park Drive
   Westford, Massachusetts 01886
   United States

   Email: erosen@juniper.net

   Mahesh Sivakumar
   Cisco Systems, Inc.
   510 McCarthy Blvd
   Milpitas, California 95035
   United States

   Email: masivaku@cisco.com

   Sam K Aldrin
   Google, Inc.
   1600 Amphitheatre Parkway
   Mountain View, California
   United States

   Email: aldrin.ietf@gmail.com

   Andrew Dolganow
   Nokia
   600 March Rd.
   Ottawa, Ontario K2K 2E6
   Canada

   Email: andrew.dolganow@nokia.com

   Tony Przygienda
   Juniper Networks, Inc.
   1137 Innovation Way
   San Jose, California 94089
   United States

   Email: prz@juniper.net