Route Distinguisher Outbound Route Filter for BGP-4
draft-wang-idr-rd-orf-00

Abstract

   This draft defines a new Outbound Route Filter (ORF) type, called the
   Route Distinguisher ORF (RD-ORF). RD-ORF is applicable in BGP/MPLS
   IP Virtual Private Networks and Ethernet VPN, in which the Route
   Reflector (RR) and Autonomous System border router (ASBR) do not
   maintain the VPN routes.

Status of This Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at https://datatracker.ietf.org/drafts/current/.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time. It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."

   This Internet-Draft will expire on December 3, 2020.

Table of Contents

1. Introduction

   According to the description in [RFC4364], MPLS VPN can cross-domain
   networking in three ways, we usually called them Option A, B and C.
   In Option A, an Autonomous System border router (ASBR) is attached
   directly to an ASBR in another AS, they act as CE/PE of each other
   and exchange VPN route, so they SHOULD maintain VPN route information
   in Virtual Routing Forwarding (VRF). So, the route control mechanism
   based on Prefix Limit can be used in Option A scenario, to prevent
   the number of routes transmitted by one of them from exceeding the
   maximum number of another's routing tables. In Option B, a ASBR is
   connected to another via EBGP. ASBRs only perform VPN routing
   forwarding, the maintenance of VRF is done by PEs. In Option C, RRs
   are connected each other via MP-EBGP and are responsible for VPN
   routing forwarding, without maintaining VRFs. Due to ASBRs and RRs
   do not maintain VRFs in Option B and C, the route control mechanism
   based on Prefix Limit does not work.

   [RFC5291] defines a BGP-based mechanism called Outbound Route Filter
   (ORF), which allows a BGP speaker to send a set of ORFs to its BGP
   peers, the peers will constrain the routing information that they
   decide to transmit to the speaker.

   This draft defines a new ORF-type, called the Route Distinguisher ORF
   (RD-ORF). When the route table of a BGP speaker reaches its maximum,
   the BGP speaker will send a RD-ORF to its BGP peers that carry a RD.
   If a BGP speaker receives a RD-ORF from its BGP peer, it will filter
   the VPN routes it tends to send according to the RD-ORF entry.

   RD-ORF is applicable in BGP/MPLS IP Virtual Private Networks
   [RFC4364], and BGP/MPLS Ethernet VPN (EVPN) networks [RFC7432].

2. Conventions used in this document

   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 [RFC2119] .

3. Terminology

   The following terms are defined in this draft:

   o RD: Route Distinguisher, defined in [RFC4364]

   o ORF: Outbound Route Filter, defined in [RFC5291]

   o AFI: Address Family Identifier, defined in [RFC4760]

   o SAFI: Subsequent Address Family Identifier, defined in [RFC4760]

   o EVPN: BGP/MPLS Ethernet VPN, defined in [RFC7432]

   o RR: Router Reflector, provides a simple solution to the problem of
      IBGP full mesh connection in large-scale IBGP implementation.

   o VRF: Virtual Routing Forwarding, a virtual routing table based on
      VPN instance.

4. RD-ORF Encoding

   In this draft, we defined a new ORF type called Route Distinguisher
   Outbound Route Filter (RD-ORF). The ORF entries are carried in the
   BGP ROUTE-REFRESH message as defined in [RFC5291]. A BGP ROUTE-
   REFRESH message can carry one or more ORF entries. The ROUTE-REFRESH
   message which carries ORF entries contains the following fields:

   o AFI (2 octets)

   o SAFI (1 octet)

   o When-to-refresh (1 octet): the value is IMMEDIATE or DEFER

   o ORF Type (1 octet)

   o Length of ORF entries (2 octets)

   A RD-ORF entry contains a common part and type-specific part. The
   common part is encoded as follows:

   o Action (2 bits): the value is ADD, REMOVE or REMOVE-ALL

o Match (1 bit): the value is PERMIT or DENY

o Reserved (5 bits)

RD-ORF also contains type-specific part. The encoding of the type-
specific part is shown in Figure 1.

```
            +--------------------------------------+
            |                                      |
            |  Sequence (4 octets)                 |
            |                                      |
            +--------------------------------------+
            |                                      |
            |  Route Distinguisher (8 octets)      |
            |                                      |
            +--------------------------------------+
            |                                      |
            |  Reserved (4 octets)                 |
            |                                      |
            +--------------------------------------+
```
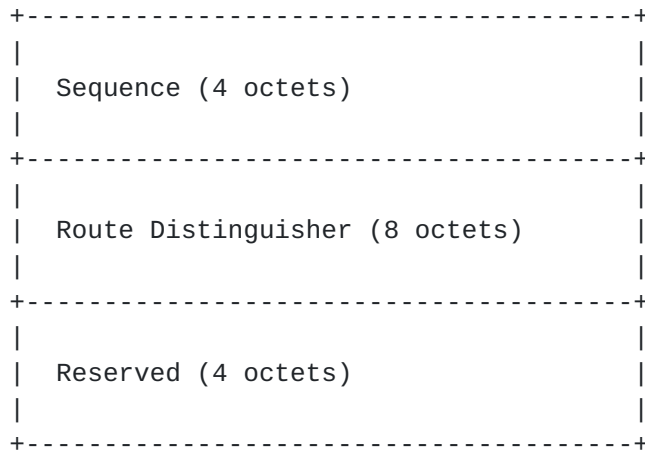
                Figure 1: RD-ORF type-specific encoding

o Sequence: identifying the order in which RD-ORF is generated

o Route Distinguisher: distinguish the different user routes. The
   RD-ORF filters the VPN routes it tends to send based on Route
   Distinguisher.

Note that if the Action component of an ORF entry specifies REMOVE-
ALL, the ORF entry does not include the type-specific part.

When the BGP ROUTE-REFRESH message carries RD-ORF entries, it must be
set as follows:

o The ORF-Type MUST be set to RD-ORF.

o The AFI MUST be set to IPv4, IPv6, or Layer 2 VPN (L2VPN).

o If the AFI is set to IPv4 or IPv6, the SAFI MUST be set to MPLS-
   labeled VPN address.

o If the AFI is set to L2VPN, the SAFI MUST be set to BGP EVPN.

o The Match field MUST be equal to DENY.

5. Operation

   In VPN/EVPN, RD is used to distinguish different user routes. When a
   BGP speaker's route table reaches its maximum, it will generate a BGP
   ROUTE-REFRESH message contains a RD-ORF entry, and send it to its BGP
   peers, reminding them not to send it routes whose Type field in the
   BGP header is set to UPDATE (BGP UPDATE message). In the RD-ORF
   entry, the RD field must be set to the Route Distinguisher of the BGP
   speaker.

   When a BGP speaker receives a ROUTE-REFRESH message that carries a
   RD-ORF entry, it will check whether the message violates the encoding
   rules or not. If the ROUTE-REFRESH message violates the rules, the
   BGP speaker MUST ignore it and log the event. Otherwise, the BGP
   speaker checks the Action field of the RD-ORF entry. If the Action
   field is set to ADD, the BGP speaker adds this entry to the Adj-RIB-
   out. The location is specified by the Sequence field. If the Action
   field is set to REMOVE, the BGP speaker removes the CP-ORF entry
   specified by the Sequence field from the Adj-RIB-out. If the Action
   field is set to REMOVE-ALL, the BGP speaker removes all entries in
   the Adj-RIB-out.

   For a BGP speaker, before sending BGP routes to its peers, it checks
   the Adj-RIB-out to find the related strategies. If it finds a RD-ORF
   associated with a BGP peer in the Adj-RIB-out, it will not send any
   BGP UPDATE message to that peer.

5.1. Application in Option B cross-domain scenario

   The Option B cross-domain scenario is shown in Figure 2:

```
 +------------+------------+                 +------------+------------+
 |      +----+----+        |                 |      +----+----+        |
 |      |         |        |                 |      |         |        |
 |  +----+   RR1  +----+   |                 |  +----+   RR2  +----+   |
 |  |    |        |    |   |                 |  |    |        |    |   |
 |  |    +--------+    |   |                 |  |    +--------+    |   |
 |  |                  |   |                 |  |                  |   |
 |  |IBGP         IBGP|   |                 |  |IBGP         IBGP|   |
 |  |                  |   |                 |  |                  |   |
 +-+--+----+      +----+--+-+         +-+--+----+      +----+--+-+
 |         |      |      | EBGP  |         |      |         |
 |   PE1   |      | ASBR1 +----------+ ASBR2 |      |   PE2   |
 |         |      |      |         |      |         |         |
 +-+-------+  AS1  +-------+-+         +-+-------+  AS2  +-------+-+
   +----------------------+             +----------------------+
```
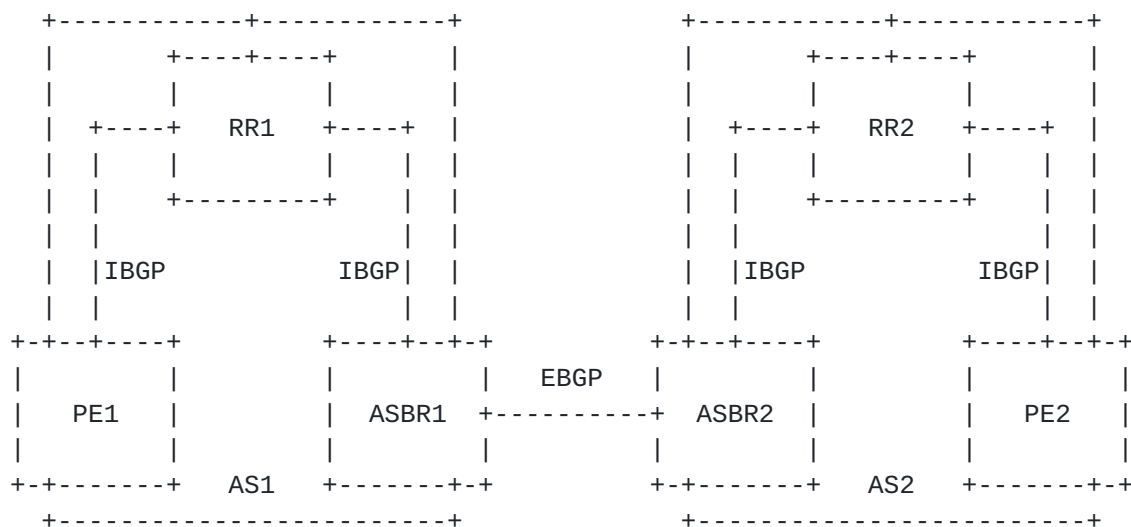
                  Figure 2: The Option B cross-domain scenario

In this scenario, PE1 and PE2 are responsible for maintaining VPN
routing of devices in AS1 and AS2, respectively. There is a direct
link between ASBR1 and ASBR2 via EBGP. In AS1, PE1 and ASBR1
establish IBGP session with RR1 to ensure the routes can be
transmitted in AS1, same in AS2.

Due to the maintenance of VPN routes is only done by PEs. ASBRs
cannot know whether the PEs' ability to handle VPN routes has reached
the upper limit or not, so it needs the RD-ORF to control the number
of routes.

Assume that PE1 and PE2 can transmit VPN routes through the network
architecture shown in Figure 2. When PE1 thinks it cannot handle
more VPN routes, it will send ROUTE-REFRESH message that carries a
RD-ORF entry to RR1. The entry consists of the following fields:

o AFI is set to IPv4 or IPv6

o SAFI is set to "MPLS-labeled VPN address"

o When-to-refresh is set to IMMEDIATE

o Action is set to ADD

o Match is set to DENY

o ORF Type is set to RD-ORF

o RD-ORF Sequence is set to 1

o RD-ORF Route Distinguisher is set to RD1

It noted that for a RD, the sequence of the first RD-ORF is equal to
1. When a PE needs to send a second RD-ORF entry associated with the
same RD, the RD-ORF sequence SHOULD increment.

When RR1 receives the ROUTE-REFRESH message, it checks the <AFI/SAFI,
ORF-Type> to determine whether it is interested in the entry. Then
RR1 will check the RD-ORF Route Distinguisher and sequence, find the
Route Distinguisher is equal to RD1 and the sequence is equal to 1,
indicating that it received the latest entry. If the above
conditions are not all met, RR1 will discard the entry; otherwise,
RR1 will add the RD-ORF entry into its Adj-RIB-out and stop sending
VPN routes containing that RD to PE1, and transmitting this RD-ORF
entry to ASBR1.

After receiving the RD-ORF entry, ASBR1 will repeat the above process
and send the entry to ASBR2. The RD-ORF entry will be transmitted
toward PE2, until PE2 receives it and add it into the Adj-RIB-out.

Before sending a VPN route (the RD is equal to RD1) toward PE1, PE2
will check its Adj-RIB-out and find the RD-ORF entry prevent it from
sending VPN route which carries RD1 to RR2. Then, PE2 will stop
sending that VPN route.

If PE1 can re-receive the route entries, it will send a ROUTE-REFRESH
message to RR1, carrying a RD-ORF entry consists of the following
fields:

o AFI is set to IPv4 or IPv6

o SAFI is set to "MPLS-labeled VPN address"

o When-to-refresh is set to IMMEDIATE

o Action is set to REMOVE

o Match is set to DENY

o ORF Type is set to RD-ORF

o RD-ORF Sequence is set to 2

o RD-ORF Route Distinguisher is set to RD1

When RR1 receives the ROUTE-REFRESH message, it checks the <AFI/SAFI,
ORF-Type> to determine whether it is interested in the entry. Then
RR1 will check the RD-ORF Route Distinguisher and sequence, find the
Route Distinguisher is equal to RD1 and the sequence is equal to 2,
indicating that it received the latest entry. If the above
conditions are not all met, RR1 will discard the entry; otherwise,
RR1 will remove the RD-ORF entry from its Adj-RIB-out and send this
entry to ASBR1.

After receiving the RD-ORF entry, ASBR1 will repeat the above process
and send the entry to ASBR2. The RD-ORF entry will be transmitted
toward PE2, until PE2 receives it and remove the entry from the Adj-
RIB-out.

Before sending a VPN route (the RD is equal to RD1) toward PE1, PE2
will check its Adj-RIB-out and find there is no filter associated
with RD1. Then, PE2 will send that VPN route.

5.2. Application in Option C cross-domain scenario

   The Option C cross-domain scenario is shown in Figure 3:

              Please view in a fixed-width font such as Courier.

                                MP-EBGP
               +---+
               | |
      +-----------+-----------+   +-----------+-----------+
      | +----+----+ | | +----+----+ |
      | | | | | | | |
      | +----+   RR1   +----+ | | +----+   RR2   +----+ |
      | | | | | | | | | | | |
      | | +---------+ | | | | +---------+ | |
      | | | | | | | |
      | |IBGP   IBGP| | | |IBGP   IBGP| |
      | | | | | | | |
     +-+--+----+   +----+--+-+   +-+--+----+   +----+--+-+
     | | | | | | | |
     | PE1 | | ASBR1 +   + ASBR2 | | PE2 |
     | | | | | | | |
     +-+-------+ AS1 +-------+-+   +-+-------+ AS2 +-------+-+
       +-----------------------+   +-----------------------+
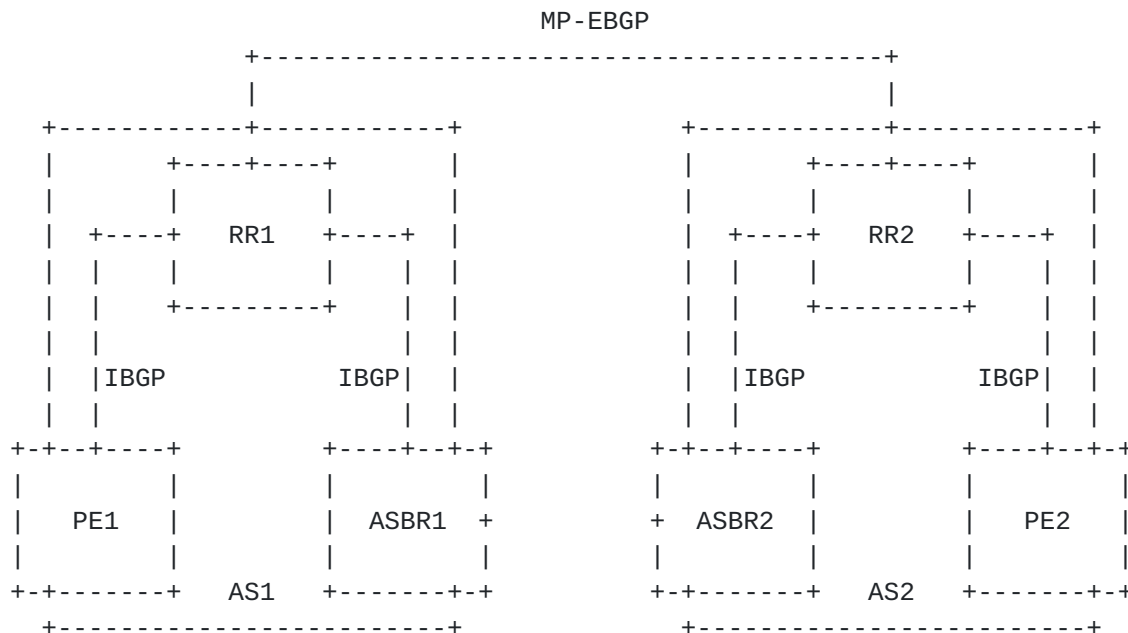

              Figure 3: The Option C cross-domain scenario

   In this scenario, PE1 and PE2 are responsible for maintaining VPN
   routing of devices in AS1 and AS2, respectively. In order to reduce
   the complexity that full-mesh brings to the network, RR1 and RR2
   establish MP-BGP session to transmit labeled routes. In AS1, PE1 and
   ASBR1 establish IBGP session with RR1 to ensure the routes can be
   transmitted in AS1, same in AS2.

   Due to the maintenance of VPN routes is only done by PEs. RRs cannot
   know whether the PEs' ability to handle VPN routes has reached the
   upper limit or not, so it needs the RD-ORF to control the number of
   routes.

   Assume that PE1 and PE2 can transmit VPN routes through the network
   architecture shown in Figure 3. When PE1 thinks it cannot handle
   more VPN routes, it will send ROUTE-REFRESH message that carries a
   RD-ORF entry to RR1. The entry consists of the following fields:

   o AFI is set to IPv4 or IPv6

   o SAFI is set to "BGP EVPN"

o When-to-refresh is set to IMMEDIATE

o Action is set to ADD

o Match is set to DENY

o ORF Type is set to RD-ORF

o RD-ORF Sequence is set to 1

o RD-ORF Route Distinguisher is set to RD1

When RR1 receives the ROUTE-REFRESH message, it checks the <AFI/SAFI,
ORF-Type> to determine whether it is interested in the entry. Then
RR1 will check the RD-ORF Route Distinguisher and sequence, find the
Route Distinguisher is equal to RD1 and the sequence is equal to 1,
indicating that it received the latest entry. If the above
conditions are not all met, RR1 will discard the entry; otherwise,
RR1 will remove the RD-ORF entry from its Adj-RIB-out and send this
entry to ASBR1 and RR2. After receiving the RD-ORF entry, RR2 and
ASBR1 will repeat the above process, then send the entry to ASBR2 and
PE2.

Before sending a VPN route (the RD is equal to RD1) toward PE1, PE2
will check its Adj-RIB-out and find the RD-ORF entry prevent it from
sending VPN route which carries RD1 to RR2. Then, PE2 will stop
sending that VPN route.

If PE1 can re-receive the VPN routes, it will send a ROUTE-REFRESH
message to RR1, carrying RD-ORF entries consist of the following
fields:

o AFI is set to IPv4 or IPv6

o SAFI is set to "BGP EVPN"

o When-to-refresh is set to IMMEDIATE

o Action is set to REMOVE

o Match is set to DENY

o ORF Type is set to RD-ORF

o RD-ORF Sequence is set to 2

o RD-ORF Route Distinguisher is set to RD1

When RR1 receives the ROUTE-REFRESH message, it checks the <AFI/SAFI,
ORF-Type> to determine whether it is interested in the entry. Then
RR1 will check the RD-ORF Route Distinguisher and sequence, find the
Route Distinguisher is equal to RD1 and the sequence is equal to 2,
indicating that it received the latest entry. If the above
conditions are not all met, RR1 will discard the entry; otherwise,
RR1 will remove the RD-ORF entry from its Adj-RIB-out and send this
entry to ASBR1 and RR2. After receiving the RD-ORF entry, RR2 and
ASBR1 will repeat the above process, then send the entry to ASBR2 and
PE2.

Before sending a VPN route (the RD is equal to RD1) toward PE1, PE2
will check its Adj-RIB-out and find there is no filter associated
with RD1. Then, PE2 will send that VPN route.

6. Security Considerations

A BGP speaker will maintain the RD-ORF entries in Adj-RIB-out, this
behavior consumes its memory and compute resources. To avoid the
excessive consumption of resources, [RFC5291] specifies that a BGP
speaker can only accept ORF entries transmitted by its interested
peers.

7. IANA Considerations

This document defines a new Outbound Route Filter type - Route
Distinguisher Outbound Route Filter. The code point is from the "BGP
Outbound Route Filtering (ORF) Types". It is recommended to set the
code point of RD-ORF to 66.

8. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
           Requirement Levels", BCP 14, RFC 2119,
           DOI 10.17487/RFC2119, March 1997,
           <https://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, <https://www.rfc-editor.org/info/rfc4364>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter,
           "Multiprotocol Extensions for BGP-4", RFC 4760,
           DOI 10.17487/RFC4760, January 2007,
           <https://www.rfc-editor.org/info/rfc4760>.

   [RFC5291] Chen, E. and Y. Rekhter, "Outbound Route Filtering
              Capability for BGP-4", RFC 5291, DOI 10.17487/RFC5291,
              August 2008, <https://www.rfc-editor.org/info/rfc5291>.

   [RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A.,
              Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based
              Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February
              2015, <https://www.rfc-editor.org/info/rfc7432>.

Authors' Addresses

   Wei Wang
   China Telecom
   Beiqijia Town, Changping District
   Beijing, Beijing 102209
   China

   Email: wangw36@chinatelecom.cn

   Aijun Wang
   China Telecom
   Beiqijia Town, Changping District
   Beijing, Beijing 102209
   China

   Email: wangaj3@chinatelecom.cn