IDR Working Group
Internet-Draft
Intended status: Experimental
Expires: 13 April 2026

W. Wang
A. Wang
China Telecom
H. Wang
Huawei Technologies
G. Mishra
Verizon Inc.
J. Dong
Huawei Technologies

10 October 2025

VPN Prefix Outbound Route Filter (VPN Prefix ORF) for BGP-4
draft-ietf-idr-vpn-prefix-orf-23

Abstract

This draft defines a new type of Outbound Route Filter (ORF), known as the Virtual Private Network (VPN) Prefix ORF. The VPN Prefix ORF mechanism is applicable when VPN routes from different Virtual Routing and Forwardings (VRFs) are exchanged through a single shared Border Gateway Protocol (BGP) session.

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 13 April 2026.

Copyright Notice

Table of Contents

1. Introduction

BGP Maximum Prefix feature [RFC4486] is often used at the network boundary to control the number of prefixes to be injected into the network. But for some scenarios when the VPN routes from several VRFs are advertised via one shared BGP session, there is lack of appropriate methods to control the flooding of VPN routes within one VRF to avoid overwhelming the processing of VPN routes in other VRFs, which consequently affects the route processing performance of other normal VRFs (such as route dropping, processing delays, and abnormal customer services). That is to say, the excessive VPN routes

advertisement SHOULD be controlled individually for each VRF in such shared BGP session.

There are several solutions that can be used to alleviate this problem:

* Route Target Constraint (RTC) as defined in [RFC4684]

* Address Prefix ORF as defined in [RFC5292]

* Covering Prefixes Outbound Route Filter (CP-ORF) mechanism as defined in [RFC7543]

* Provider Edge (PE) - Customer Edge (CE) edge peer Maximum Prefix

* Configuring the Maximum Prefix for each VRF on edge nodes

However, there are limitations to existing solutions:

1) Route Target Constraint

RTC can only filter the VPN routes from any uninterested VRFs, if the "offending routes (prefixes)" come from an interested VRF, the RTC mechanism can’t filter them.

2) Address Prefix ORF

Using Address Prefix ORF to filter VPN routes requires a pre-configuration, but it is impossible to know in advance which prefix MAY exceed the predefined threshold.

3) CP-ORF Mechanism

[RFC7543] defines the Covering Prefixes ORF (CP-ORF). A BGP speaker sends a CP-ORF to a peer in order to pull routes that cover a specified host address. A prefix covers a host address if it can be used to forward traffic towards that host address.

CP-ORF is applicable in Virtual Hub-and-Spoke[RFC7024] VPN and also BGP/MPLS Ethernet VPN (EVPN)[RFC7432] networks, but its primary function is to retrieve interested VPN prefixes and it cannot be used to filter overwhelmed VPN prefixes dynamically.

4) PE-CE edge peer Maximum Prefix

The BGP Maximum-Prefix feature is used to control how many prefixes can be received from a neighbor. By default, this feature allows a router to bring down a peer when the number of received prefixes from that peer exceeds the configured Maximum-Prefix limit. This feature is commonly used for external BGP peers. If it is applied to internal BGP peers, for example the VPN scenarios, all the VPN routes from different VRFs will share the common fate. If the number of VPN routes of a certain VPN exceeds the configured Maximum-Prefix limit, the BGP session will be shut down, which will affect the operation of other VPN routes transmitted via this BGP session.

5) Configuring the Maximum Prefix for each VRF on edge nodes

When a VRF overflows, it stops the import of routes. Any additional VPN routes are held into its Routing Information Base (RIB). However, PEs still need to parse the incoming BGP messages. This will cost CPU cycles and further burden the overflowed PE.

This draft defines a new type of Outbound Route Filter (ORF), called the VPN Prefix ORF. This ORF mechanism is event-driven and does not require pre-configuration. When the number of VPN routes in a VRF exceeds the prefix limit, the router will identify the VPN prefix (Route Distinguisher (RD), Route Target (RT), source PE, etc.) of the offending VPN routes in this VRF and send a VPN Prefix ORF message to its BGP peer, who announced these offending routes. Upon receiving a VPN Prefix ORF entry from its BGP peer, the BGP speaker will filter and withdraw any offending VPN routes that was announced to its peer.

The purpose of this mechanism is to control the outage within the minimum range and avoid route churn effects when a VRF on a device in the network overflows.

VPN Prefix ORF is applicable when the VPN routes from different VRFs are exchanged via one shared BGP session.

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 used in this draft:

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

* ASBR: Autonomous System Border Router.

* BGP: Border Gateway Protocol, defined in [RFC4760]

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

* MPLS: Multi-Protocol Label Switching.

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

* RD: Route Distinguisher, defined in [RFC4364]

* RIB: Routing Information Base.

* RR: Route Reflector, provides a simple solution to the problem of IBGP full mesh connection in large-scale IBGP implementation [RFC4456]

* RT: Route Target, defined in [RFC4364]

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

* VPN: Virtual Private Networks, defined in [RFC4364]

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

4. The general procedures of VPN Prefix ORF mechanism

The operation of VPN Prefix ORF mechanism on each device is independent, each of them makes a local judgment to determine whether it needs to send a VPN Prefix ORF message to its upstream peer. Operators can configure the algorithms in the devices according to their own circumstances.

This section describes the procedures for the receiving BGP peer to receive VPN route information from the sending BGP peer. The VPN information includes updated VPN routes and their corresponding VPN instance identification information. Based on the VPN instance identification information, the receiving BGP peer determines the newly added VPN routes. It then checks whether the number of newly added VPN routes has caused the total number of VPN routes to exceed the maximum route limit for the associated VPN instance.

If the route limit of the VPN instance, which is identified by the VPN instance identification information, is reached or exceeded, the receiving BGP peer will send a VPN Prefix ORF message to the sending BGP peer, indicating that it should stop sending the corresponding VPN routes which are identified by the VPN instance identification information.

Before originating a VPN Prefix ORF message, the device SHOULD compare the list of RTs carried by VPN routes with those imported by other VRFs on the device. If the route's RT is included in the import rules of other VRFs, the VPN Prefix ORF message MUST NOT be originated.

Before sending a VPN Prefix ORF entry, a sender SHOULD send a "default" entry to the VPN Prefix ORF receiver, to allow other allowed VPN prefixes to pass the filter. The "default" entry should be installed in advance in the VPN Prefixes ORF table, with the offending VPN routes process method set to 0, sequence set to 0xFFFFFFFF, length set to 8, and Route Distinguisher set to 0.

The receiving BGP peer and the sending BGP peer are iBGP peers within the same Autonomous System (AS). The VPN instance identification information is RD and the instruction information is sent using ORF in the ROUTE-REFRESH message.

The instruction information sent from the receiving BGP peer includes the following information:

* The ORF entries that are included in the ROUTE-REFRESH message.

* The Action field in the ORF entries is set to a value that instructs the sending BGP peer to add the corresponding filter condition to its outbound route filter.

* The Match field in the ORF entries is set to a value that instructs the sending BGP peer to deny VPN routes updates that match the corresponding ORF entries.

* The RD value that identifies the above mentioned VPN instance is added to the type-specific part of the ORF entries.

When multiple VRFs on a PE are receiving VPN routes with a specific RD, if one VRF exceeds its limit upon receiving routes with that RD, then the PE sends a VPN Prefix ORF message, which will prevent other VRFs that have not exceeded their limits from receiving VPN routes containing that RD, thereby avoiding any communication disruptions between these VRFs and the rejected VPN routes. In order to more finely control VPN routing, when not all VRFs on a PE that are interested in VPN routes with a specific RD exceed the limit, the PE MUST NOT send a VPN Prefix ORF entry.

When the VPN Prefix ORF mechanism is triggered, the device SHOULD send an alarm information to network operators. The detailed procedures for different scenarios are described below:

4.1. Intra-domain Scenarios and Solutions

For intra-AS VPN deployment, there are three scenarios:

* RD is allocated per VPN per PE, each VRF only import one RT (see Section 4.1.1).

* RD is allocated per VPN per PE. Multiple RTs are associated with such VPN routes, and are imported into different VRFs in other devices(see Section 4.1.2).

* RD is allocated per VPN, each VRF imports one/multiple RTs (see Section 4.1.3).

The following sections will describe solutions to the above scenarios in detail.

4.1.1. Scenario-1 and Solution (Unique RD, One RT)

In this scenario, PE1-PE4 and RR are iBGP peers. RD is allocated per VPN per PE. The offending VPN routes only carry one RT. We assume that the network topology is shown in Figure 1.

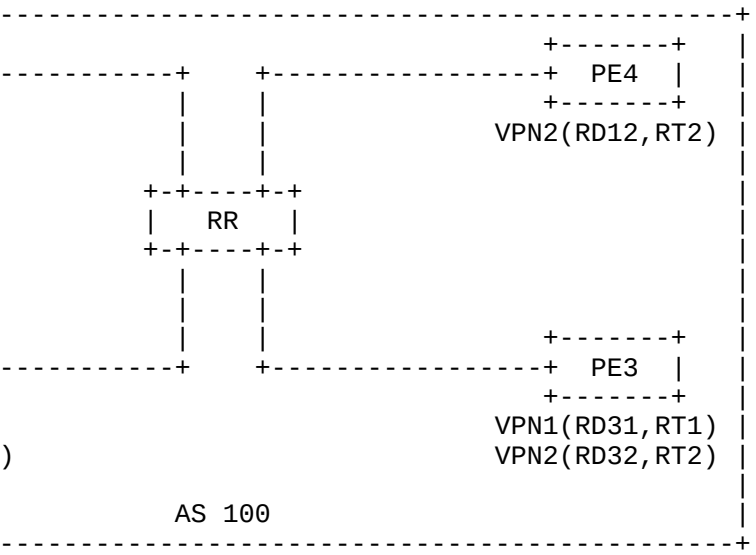

Figure 1 Network Topology of Scenario-1

When PE3 sends an excessive number of VPN routes with RT1, and both PE1 and PE2 import VPN routes with RT1, the process of offending VPN routes will influence performance of VRFs on PEs. PEs and RR SHOULD have appropriate mechanisms to identify and control the advertising of offending VPN routes.

a) PE1

If quota value is not set on PE1, and each VRF has a prefix limit on PE1. When the PE1 receives VPN routes from its BGP peer, it does the following:

S01. If (the prefix limit for VPN1 VRF is exceeded){
S02. PE1 sends a VPN Prefix ORF message to the RR and a warning message to the operator. The VPN Prefix ORF message will carry the RD is set to RD31, the RT value is set to RT1, the source PE is PE3. RR handles the offending VPN routes and controls the number of VPN routes according to the value of "Offending VPN routes process method".
S03. } else {
S04. PE1 SHOULD NOT trigger the VPN Prefix ORF mechanism, and only performs VPN route filtering for the target VRF.
S05. }

NOTE: When the prefix limit for VPN1 VRF is exceeded, there are no other VRFs on PE1 that need the VPN routes with RT1. PE1 sends a VPN Prefix ORF message to the RR and a warning message to the operator.

If each <RD31, source PE3> tuple imported into a VRF has a quota, and each VRF has a prefix limit. When the PE1 receives VPN routes from its BGP peer, it does the following:

```
S01. If (VPN routes associated with <RD31, PE3>
     tuple exceed the quota) {
S02.     If (the prefix limit of VPN1 VRF is not
         exceeded) {
S03.         PE1 sends a warning message to the
             operator, and the VPN Prefix ORF
             mechanism SHOULD NOT be triggered.
S04.     } else {
S05.         PE1 generates a BGP ROUTE-REFRESH
             message containing a VPN Prefix ORF
             entry with (RD31, source PE is PE3,
             RT is RT1), and send the entry to RR.
             RR handles the offending VPN routes
             according to the value of "Offending
             VPN routes process method".
S06.     }
S07. }
```

b) PE2

If quota value is not set on PE2, and each VRF has a prefix limit on PE2. When the PE2 receives VPN routes from its BGP peer, it does the following:

```
S01. If (the prefix limit for VPN1 VRF is exceeded) {
S02.     If (the prefix limit for VPN2 VRF is exceeded) {
S03.         PE2 sends a VPN Prefix ORF message to the RR and a
             warning message to the operator. The VPN Prefix ORF
             message will indicate the RD set to RD31, the RT
             value set to RT1. RR handles the offending VPN routes
             and controls the number of VPN routes according to
             the value of "Offending VPN routes process method".
S04.     } else {
S05.         PE2 SHOULD NOT trigger the VPN Prefix ORF mechanism,
             and only performs VPN route filtering for the target
             VRF.
S06.     }
S07. }
```

NOTE: PE2 cannot directly trigger the VPN Prefix ORF mechanism when the prefix limit of VPN1 VRF is exceeded, because VPN2 VRF requires the VPN routes with RT1. PE2 triggers the mechanism only when the prefix limits for both the VPN1 and VPN2 VRFs have been exceeded.

If each <RD31, source PE3> tuple imported into a VRF has a quota, and each VRF has a prefix limit. When the PE2 receives VPN routes from its BGP peer, it does the following:

```
S01. If (VPN routes associated with <RD31, PE3> tuple exceed the
     quota) {
S02.     If (the prefix limit of VPN1 VRF is not exceeded) {
S03.         PE2 sends a warning message to the operator, and the
             VPN Prefix ORF mechanism SHOULD NOT be triggered.
S04.     } else {
S05.         If (the prefix limit of VPN2 VRF is not exceeded) {
S06.             PE2 SHOULD NOT trigger the VPN Prefix ORF
                 mechanism, and only performs VPN route filtering
                 for the target VPN1 VRF, stopping the import of
                 VPN routes with <RD31, PE3>.
S07.         } else {
S08.             PE2 generates a BGP ROUTE-REFRESH message
                 containing a VPN Prefix ORF entry with (RD31,
                 source PE is PE3, RTs are RT1 and RT2), and send
                 the entry to RR. RR handles the offending VPN
                 routes according to the value of "Offending VPN
                 routes process method".
S09.         }
S10.     }
S11. }
```

4.1.2. Scenario-2 and Solution (Unique RD, Multiple RTs)

In this scenario, PE1-PE4 and RR are iBGP peers. RDs are allocated per VPN per PE. Multiple RTs are associated with the offending VPN routes and are imported into different VRFs on other devices. We assume the network topology is depicted in Figure 2.

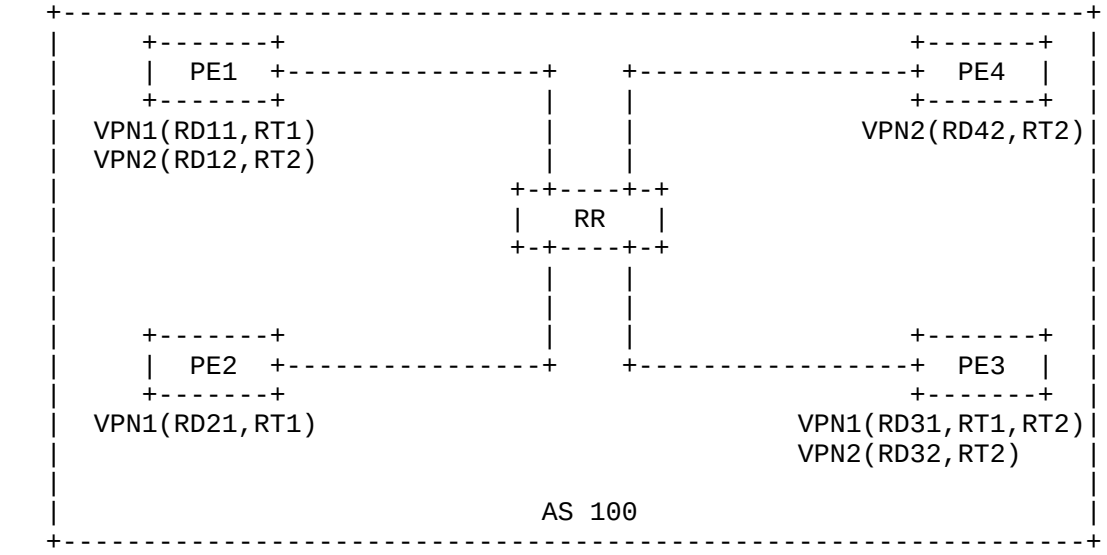

Figure 2 Network Topology of Scenario-2

When PE3 sends an excessive number of VPN routes with RT1 and RT2, while both PE1 and PE2 import VPN routes with RT1, and PE1 also imports VPN routes with RT2.

a) PE1

If quota value is not set on PE1, and each VRF has a prefix limit on PE1. Since VPN2 VRF requires the VPN routes with RT2, PE1 cannot directly trigger VPN Prefix ORF mechanism when the prefix limit of VPN1 VRF is exceeded. This case is similar to PE2 without quota in Section 4.1.1, which is modified as follows:

S03. PE1 sends a VPN Prefix ORF message to the RR and a warning message to the operator. The VPN Prefix ORF message will indicate the RD set to RD31, the RT value set to RT1 and RT2, source PE identified as PE3. RR handles the offending VPN routes and controls the number of VPN routes according to the value of "Offending VPN routes process method".

If each <RD31, source PE3> tuple imported into a VRF has a quota, and each VRF has a prefix limit. This case is similar to PE2 with quota in Section 4.1.1, which is modified as follows:

S08. PE1 generates a BGP ROUTE-REFRESH message containing a VPN Prefix ORF entry with (RD31, source PE is PE3, RTs are RT1 and RT2), and send the entry to RR. RR handles the offending VPN routes according to the value of "Offending VPN routes process method".

b) PE2

If quota value is not set on PE2, and each VRF has a prefix limit on PE2. Since only VPN1 VRF needs to import VPN routes with RT1, this case is similar to PE1 without quota in Section 4.1.1, which is modified as follows:

S02. PE2 sends a VPN Prefix ORF message to the RR and a warning message to the operator. The VPN Prefix ORF message will indicate the RD set to RD31, the RT value set to RT1 and RT2,source PE identified as PE3. RR handles the offending VPN routes and controls the number of VPN routes according to the value of "Offending VPN routes process method".

If each <RD31, source PE3> tuple imported into a VRF has a quota, and each VRF has a prefix limit. This case is similar to PE1 with quota in Section 4.1.1, which is modified as follows:

S05. PE2 generates a BGP ROUTE-REFRESH message containing a VPN Prefix ORF entry with (RD31, source PE is PE3, RTs are RT1 and RT2), and send the entry to RR. RR handles the offending VPN routes according to the value of "Offending VPN routes process method".

4.1.3. Scenario-3 and Solution (Universal RD)

In this scenario, PE1-PE4 and RR are iBGP peers. RD is allocated per VPN. One/Multiple RTs are associated with the offending VPN routes and are imported into different VRFs on other devices. We assume the network topology is shown in Figure 3.

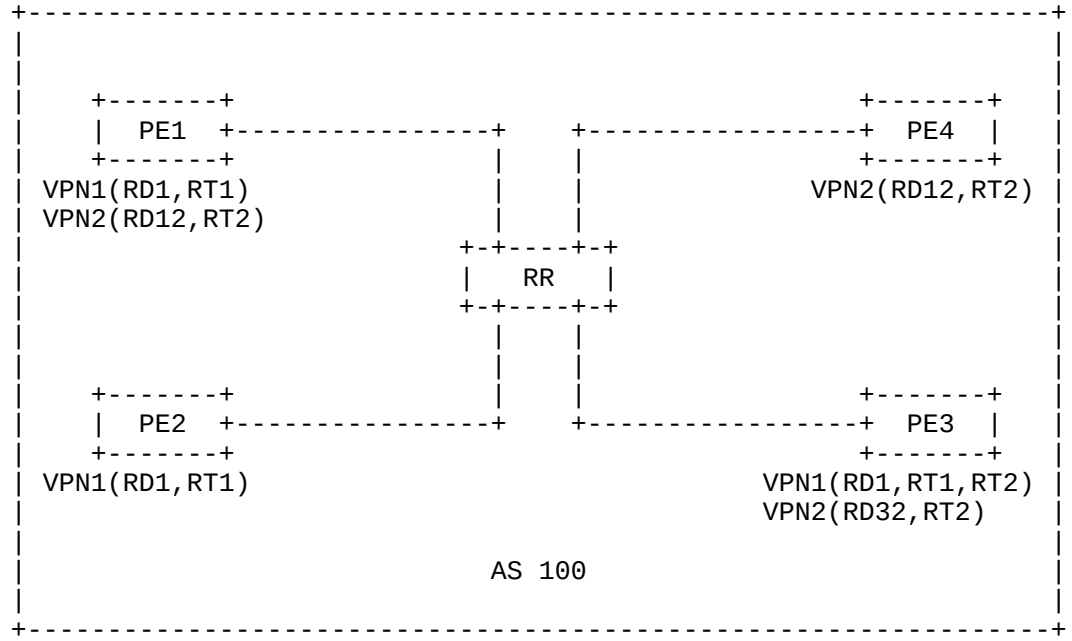

Figure 3 Network Topology of Scenario-3

When PE3 sends an excessive number of VPN routes associated with RD1, RT1 and RT2, and both PE1 and PE2 import VPN routes with RT1, the process of offending VPN routes can affect the performance of the VRFs on PEs.

a) PE1

If quota value is not set on PE1, and each VRF has a prefix limit on PE1. Since VPN2 VRF requires the VPN routes with RT2, PE1 cannot trigger VPN Prefix ORF mechanism directly when the prefix limit of VPN1 VRF is exceeded. This case is similar to PE2 without quota in Section 4.1.1, which is modified as follows:

S03. PE1 sends a VPN Prefix ORF message to the RR and a warning message to the operator. The VPN Prefix ORF message will indicate the RD set to RD1, the RT value set to RT1 and RT2, source PE identified as PE3. RR handles the offending VPN routes and controls the number of VPN routes according to the value of "Offending VPN routes process method".

If each <RD1, source PE3> tuple imported into a VRF has a quota, and each VRF has a prefix limit. This case is similar to PE2 with quota in Section 4.1.1, which is modified as follows:

S08. PE1 generates a BGP ROUTE-REFRESH message containing a VPN Prefix ORF entry with (RD1, source PE is PE3, RTs are RT1 and RT2), and send the entry to RR. RR handles the offending VPN routes according to the value of "Offending VPN routes process method".

b) PE2

If quota value is not set on PE2, and each VRF has a prefix limit on PE2. Since only VPN1 VRF needs to import VPN routes with RT1, this case is similar to PE1 without quota in Section 4.1.1, which is modified as follows:

S02. PE2 sends a VPN Prefix ORF message to the RR and a warning message to the operator. The VPN Prefix ORF message will indicate the RD set to RD1, the RT value set to RT1 and RT2, source PE identified as PE3. RR handles the offending VPN routes and controls the number of VPN routes according to the value of "Offending VPN routes process method".

If each <RD31, source PE3> tuple imported into a VRF has a quota, and each VRF has a prefix limit. This case is similar to PE1 with quota in Section 4.1.1, which is modified as follows:

S05. PE2 generates a BGP ROUTE-REFRESH message containing a VPN Prefix ORF entry with (RD1, source PE is PE3, RTs are RT1 and RT2), and send the entry to RR. RR handles the offending VPN routes according to the value of "Offending VPN routes process method".

5. Source PE Extended Community

We usually use next hop to identify the source, but it MAY NOT be useful in the following scenarios:

* a PE MAY have multiple addresses so that its BGP peer MAY receive several different next hop addresses from the same source.

* In Option B inter-domain scenario, the ASBR will change the next hop.

ORIGINATOR_ID is a non-transitive attribute generated by RR to identify the source, but ORIGINATOR_ID cannot be advertised outside the local AS. To address the above scenarios, we have defined a new Extended Community: Source PE Extended Community (SPE EC), which is designed to transmit the identifier of source. The value of SPE EC can be set by source PE, RR or Autonomous System Boundary Router (ASBR). Once set and attached to the BGP UPDATE message, its value SHOULD NOT be altered along the advertisement path.

The AS number of source PE can be conveyed by Source AS Extended Community, as defined in [RFC6514]

The format of SPE EC is shown as Figure 4.

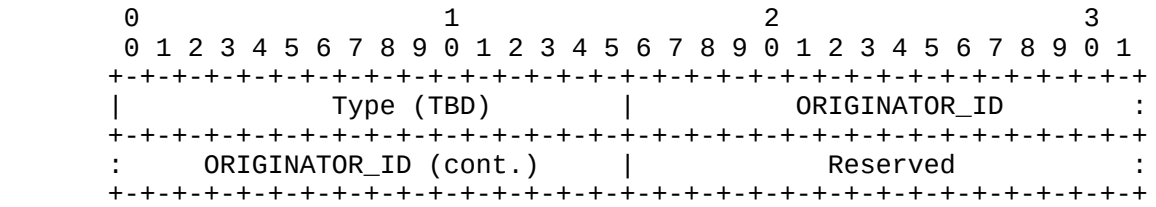

Figure 4 The format of SPE EC

Where:

* Type: specifies the type value assigned by IANA, now it is TBD.

* ORIGINATOR_ID: specifies the identifier of source.

* Reserved: MUST be zero on transmit.

For the RR/ASBR, it SHOULD perform the following:

* Check the existence of the SPE EC. If it exists, does not change it.

* If SPE EC does not exist, check the existence of ORIGINATOR_ID. If it exists, put it into SPE EC.

* If ORIGINATOR_ID does not exist, put the router-id of source PE into SPE EC.

The SPE EC SHOULD be attached by source PE, or else the RR SHOULD attach it, with the value set as the router-id of source PE. When none of them attach the SPE EC, the ASBR SHOULD attach it when the packet leaves the source AS, with the value set as the ORIGINATOR_ID.

This section updates route reflection procedures, which means [RFC4456] needs to be updated.

6. VPN Prefix ORF Encoding

In this section, we defined a new ORF type called VPN Prefix Outbound Route Filter (VPN Prefix 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:

* AFI (2 octets)

* SAFI (1 octet)

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

* ORF Type (1 octet): The type of VPN Prefix ORF is 66.

* Length of ORF entries (2 octets)

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

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

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

* Offending VPN routes process method (1 bit): if the value is set to 0, it means all offending VPN routes on the sender of VPN Prefix ORF message SHOULD be withdrawn; if the value is set to 1, it means the sender of VPN Prefix ORF message refuse to receive new offending VPN routes. The default value is 0.

* Reserved (4 bits)

VPN Prefix ORF also contains type-specific part. The encoding of the type-specific part is shown in Figure 5.

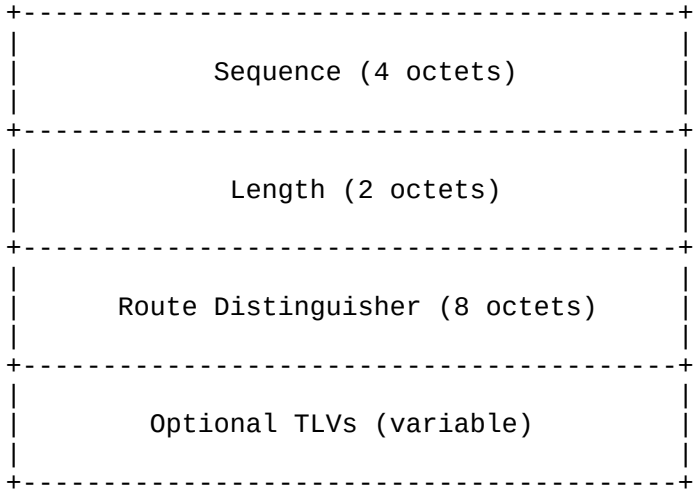

Figure 5: VPN Prefix ORF type-specific encoding

* Sequence: identifying the order in which VPN Prefix ORF is generated and evaluated. It can uniquely identify a VPN Prefix ORF entry together with AFI/SAFI, ORF-Type, and Route Distinguisher. The sequence numbers SHOULD be discontinuous to facilitate the insertion of new rules at a later stage.

* Length: identifying the length of this VPN Prefix ORF entry.

* Route Distinguisher: distinguish the different user routes. The VPN Prefix ORF filters the VPN routes it tends to send based on Route Distinguisher. If RD is equal to 0, it means all VPN prefixes.

* Optional TLVs: carry the potential additional information to give the extensibility of the VPN Prefix ORF mechanism. Its format is shown in Figure 6. If one or more TLV(s) are unrecognized, the whole VPN Prefix ORF entry SHOULD be removed.

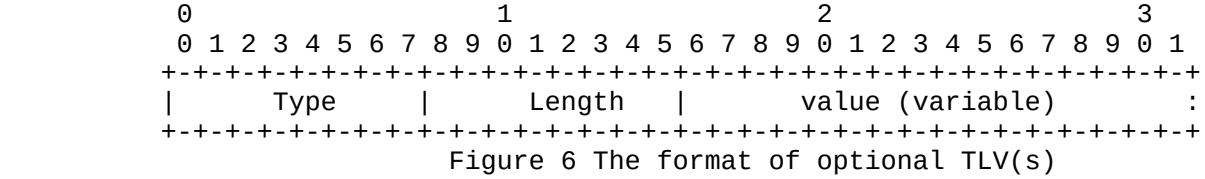

Figure 6 The format of optional TLV(s)

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 VPN Prefix ORF entries, it MUST be set as follows:

* The ORF-Type MUST be set to 66 (VPN Prefix ORF).

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

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

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

* The purpose of VPN Prefix ORF is to block unwanted VPN prefixes, then the "action" of one valid entry SHOULD be set to "DENY". In order to allow other allowed VPN prefixes pass the filter, one default, last resort entry SHOULD be installed in advance in the VPN Prefixes ORF table, with the RD is set to 0 and the corresponding Sequence are set to 0xFFFFFFFF.

According to [RFC5291], if any of the fields of a VPN Prefix ORF entry in the message contains an unrecognized value, the whole specified ORF previously received is removed.

A BGP speaker that is willing to receive ORF entries from its peer, or a BGP speaker that would like to send ORF entries to its peer, advertises this capability to the peer by using the Outbound Route Filtering Capability defined in [RFC5291].

6.1. Source PE TLV

Source PE TLV is defined to identify the source of the VPN routes. For the sender of VPN Prefix ORF, it will check the existence of SPE EC. If it exists, the sender will put it into Source PE TLV. Otherwise, the value of Source PE TLV SHOULD be set to next hop address.

The Source PE TLV SHOULD only appear once within an individual ORF entry. If one ORF entry contains multiple Source PE TLVs, it SHOULD be ignored.

The source PE TLV contains the following types:

* IPv4 Source PE TLV: Type = 1 (suggested), Length = 4 octets, value = next hop address in IPv4 format.

* IPv6 Source PE TLV: Type = 2 (suggested), Length = 16 octets, value = next hop address in IPv6 format.

* Source PE identifier TLV: Type = 3 (suggested), Length = 4 octets, value = the value of ORIGINATOR_ID in Source PE Extended Community.

6.2. Source AS TLV

Source AS TLV is defined to identify the source AS number of source PE.

The Source AS TLV SHOULD only appear once within an individual ORF entry. If one ORF entry contains multiple Source AS TLVs, it SHOULD be ignored.

The encoding of Source AS TLV is as follows:

Type = 4 (suggested), Length = 4 octets, value = the value of Source AS in Source AS Extended Community as defined in [RFC6514].

6.3. Route Target TLV

Route Target TLV is defined to identify the RT of the offending VPN routes. RT and RD can be used together to filter VPN routes when the source VRF contains multiple RTs, and the VPN routes with different RTs MAY be assigned to different VRFs on the receiver. The Route Target TLV contains the following types:

Type = 5 (suggested), Length = 8*n (n is the number of RTs that the offending VPN routes attached) octets, value = the RT of the offending VPN routes. If multiple RTs are included, there MUST be an exact match.

7. Operation process of VPN Prefix ORF mechanism on receiver

The VPN Prefix ORF is used mainly to block the unwanted BGP updates. When the receiver receives VPN Prefix ORF entry, it SHOULD check first whether the "Match" bit is "DENY" or not.

If the "Match" bit is "PERMIT", and is the "default" entry (the offending VPN routes process method equal to 0, sequence equal to 0xFFFFFFFF, length is equal to 8, and Route Distinguisher is equal to 0), the entry SHOULD be installed. Otherwise, if the "Match" bit is "PERMIT", the entry SHOULD be discarded and a warning SHOULD be sent to the operator.

The following procedures will only be evaluated when the "Match" bit is "DENY".

The receiver of VPN Prefix ORF entries, which MAY be a RR, ASBR or PE, when receives VPN Prefix ORF entry from its BGP peer, it does the following:

```
S01. The receiver checks the combination of <AFI/SAFI, ORF-Type,
     Sequence, Route Distinguisher> of the received VPN Prefix
     ORF entry.
S02. If (the combination does not already exist in the ORF-Policy
     table) {
S03.     The receiver adds the VPN Prefix ORF entry to the
         ORF-Policy table.
S04. } else {
S05.     If (Action is ADD) {
S06.         Overwrite the old VPN Prefix ORF entry with the new
             one.
S07.     else {
               Remove the corresponding VPN Prefix ORF entry.
S08. }
```

The filtering conditions for the stored VPN Prefix ORF entries contain the RD and RT of the source PE.

If the SPE EC is not attached to the BGP Update message of the VPN prefixes, the receiver SHOULD use NEXT_HOP or ORIGINATOR_ID as the originator of VPN Prefix to match against the VPN Prefix ORF entry.

After installing the filter entries for the outbound VPN prefixes, the RR or ASBR does the following before sending VPN routes:

```
S01. RR or ASBR check if there are matching filtering conditions
     in the ORF-Policy table for the VPN routes.
S02. If (matching filtering conditions does not exist) {
S03.     The RR/ASBR sends the VPN routes.
S04. } else {
S05.     If (the "Offending VPN routes process method" bit is set
         to 0) {
S06.         The RR/ASBR withdraws all the VPN routes identified
             by RD, RT and any relevant information in the optional
             TLVs within the entry, and stop sending the
             corresponding VPN routes to the sender of the VPN
             Prefix ORF entry.
S07.     } else {
S08.         The receiver withdraws the extra VPN routes according
             to the value of RD, RT and any relevant information
             in optional TLVs within the entry, and stop sending
             the corresponding VPN routes to the sender of the
             VPN Prefix ORF entry.
S09. }
```

8. Withdraw of VPN Prefix ORF entries

When the VPN Prefix ORF mechanism is triggered, a warning message will be generated and sent to the network operators. Operators SHOULD manually configure the network to resume normal operation. Since devices can record the VPN Prefix ORF entries sent by each VRF, operators can identify the entries that need to be withdrawn and manually trigger the withdraw process.

The withdrawal of the VPN Prefix ORF mechanism is manually triggered, and its activation requires two conditions:

1. Network operation and maintenance personnel have confirmed through device alarms that the issue of "offending routes", which originally caused the VRF route count to exceed the limit—has been resolved;

2. Operation and maintenance personnel have located the target ORF entry to be withdrawn. Devices record the VPN Prefix ORF entries sent by each VRF, providing a basis for personnel to locate the target of the withdrawal.

Operation and maintenance personnel manually configure withdrawal commands on the device that triggered the ORF (typically the original ORF sender, such as a PE with an exceeded route limit). The commands MUST include the unique identification information of the target ORF entry, and set the "Action" field of the ORF entry to "REMOVE" (for removing a single entry) or "REMOVE-ALL" (for removing all entries of the same type).

The withdrawal of ORF entries relies on manual intervention from a management entity (e.g., NMS), and there is no automatic withdrawal mechanism. This is to prevent route disruptions caused by misoperations.

9. Applicability

Using the scenario in Section 4.1.1, we demonstrate how to determine each field when the sender generates a VPN Prefix ORF entry. Assuming it is an IPv4 network, after PE1-PE4 and RR have advertised the Outbound Route Filtering Capability, each of PE1-PE4 SHOULD send a VPN Prefix ORF entry that means "PERMIT-ALL" as follows:

* AFI is equal to IPv4

* SAFI is equal to MPLS-labeled VPN address

* When-to-refresh is equal to IMMEDIATE

* ORF Type is equal to VPN Prefix ORF
* Length of ORF entries is equal to 22
* Action is equal to ADD
* Match is equal to PERMIT
* Offending VPN routes process method is equal to 0
* Sequence is equal to 0xFFFFFFFF
* Length is equal to 8
* Route Distinguisher is equal to 0

When the VPN Prefix ORF mechanism is triggered on PE1, PE1 generates a VPN Prefix ORF entry contains the following information:

* AFI is equal to IPv4
* SAFI is equal to MPLS-labeled VPN address
* When-to-refresh is equal to IMMEDIATE
* ORF Type is equal to VPN Prefix ORF
* Length of ORF entries is equal to 45
* Action is equal to ADD
* Match is equal to DENY
* Offending VPN routes process method is equal to 0
* Sequence is equal to 1
* Length is equal to 31
* Route Distinguisher is equal to RD31
* Optional TLV:
  - Type is equal to 1 (Source PE TLV)
  - Length is equal to 4

- value is equal to PE3’s IPv4 address
- Type is equal to 4 (Source AS TLV)
- Length is equal to 4
- value is equal to PE3’s source AS number
- Type is equal to 5 (Route Target TLV)
- Length is equal to 8
- value is equal to RT1

10. Operational Considerations

10.1. Quota value calculation

Quota is a threshold to limit the number of VPN routes under specific granularities (such as <PE>, <RD, Source AS>).

In deployment, quota values SHOULD be set and delivered by the Network Management System (NMS). The quota value can be set with different granularity, such as by <PE>, <RD, Source AS>, etc. If the quota value is set to (VRF prefix limit/the number of PEs), whenever a new PE access to the network, the quota value SHOULD be re-evaluated or adjusted accordingly.

To avoid frequent changes to the quota value, the value SHOULD be set based on the following formula:

Quota=MIN[(Margins coefficient)*<PE,CE limit>*<Number of PEs within the VPN, includes the possibility expansion in futures>, VRF Prefixes Limit]

It SHOULD be noted that the above formula is only an example, the operators can use different formulas based on actual needs in management plane.

11. Implementation Consideration

11.1. Implementation status

Currently, H3C has implemented VPN Prefix ORF mechanism related functions as follows:

* By configuring quota, achieve the use of RD and Source PE to control VPN routing.

* Generating, transmitting and processing Type 1 and Type 2 Source PE TLV.

* Using the Offending VPN routes process method to revoke all routes.

Besides, we also implemented the following functions based on the open-source BGP implementation (FRR):

* VPN Prefix ORF mechanism triggered based on VRF limit in intra-domain scenarios.

* RD based VPN routing filtering in intra-domain scenarios.

11.2. Experimental topology

The experiments will test whether the VPN Prefix ORF blocks the offending routes in the following scenarios:

* Intra-domain as a standalone mechanism

* Adding the VPN Prefix ORF to existing mechanisms for intra-domain VPNs

12. Security Considerations

This draft adds no new security considerations beyond those of [RFC5291].

13. IANA Considerations

This document defines a new Outbound Route Filter type - VPN Prefix Outbound Route Filter (VPN Prefix ORF).

We would want to refer to the text from [RFC5291]: This new ORF is exchanged using outbound route filtering capability defined in [RFC5291] (for the sake of completeness).

under "BGP Outbound Route Filtering (ORF) Types"
Registry: "VPN Prefix Outbound Route Filter (VPN Prefix ORF)"
Registration Procedure(s): First Come, First Served
Value: 66

This document also define a VPN Prefix ORF TLV type under "Border Gateway Protocol (BGP) Parameters", four TLV types are defined:

under "Border Gateway Protocol (BGP) Parameters"
Registry: "VPN Prefix ORF TLV"

Range	Registration Procedures
0-127	IETF Review
128-255	First Come First Served

Registry	Type	Meaning
Reserved	0(suggested)	Reserved
IPv4 Source PE TLV	1(suggested)	IPv4 address for source PE.
IPv6 Source PE TLV	2(suggested)	IPv6 address for source PE.
Source PE Identifier TLV	3(suggested)	ORIGINATOR_ID in Source PE Extended Community for source PE
Source AS TLV	4(suggested)	Source AS for source PE
Route Target TLV	5(suggested)	Route Target of the offending VPN routes

This document also requests a new Transitive Extended Community Type. The new Transitive Extended Community Type name SHALL be "Source PE Extended Community".

Under "BGP Transitive Extended Community Types:"
Registry: "Source PE Extended Community" type
0x0d(suggested) Source PE Extended Community

14. Contributor

Shunwan Zhuang

Huawei Technologies

Huawei Building, No.156 Beiqing Rd.

Beijing

Beijing, 100095 China

15. Acknowledgement

Thanks Jeffrey Haas, Robert Raszuk, Jim Uttaro, Jakob Heitz, Jeff Tantsura, Rajiv Asati, John E Drake, Gert Doering, Shuanglong Chen, Enke Chen, Srihari Sangli and Igor Malyushkin for their valuable comments on this draft.

16. 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>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <https://www.rfc-editor.org/info/rfc4456>.

[RFC4486] Chen, E. and V. Gillet, "Subcodes for BGP Cease Notification Message", RFC 4486, DOI 10.17487/RFC4486, April 2006, <https://www.rfc-editor.org/info/rfc4486>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <https://www.rfc-editor.org/info/rfc4684>.

[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>.

[RFC5292] Chen, E. and S. Sangli, "Address-Prefix-Based Outbound Route Filter for BGP-4", RFC 5292, DOI 10.17487/RFC5292, August 2008, <https://www.rfc-editor.org/info/rfc5292>.

[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, <https://www.rfc-editor.org/info/rfc6514>.

[RFC7024] Jeng, H., Uttaro, J., Jalil, L., Decraene, B., Rekhter, Y., and R. Aggarwal, "Virtual Hub-and-Spoke in BGP/MPLS VPNs", RFC 7024, DOI 10.17487/RFC7024, October 2013, <https://www.rfc-editor.org/info/rfc7024>.

[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>.

[RFC7543] Jeng, H., Jalil, L., Bonica, R., Patel, K., and L. Yong, "Covering Prefixes Outbound Route Filter for BGP-4", RFC 7543, DOI 10.17487/RFC7543, May 2015, <https://www.rfc-editor.org/info/rfc7543>.

Appendix A. Experimental topology

The experimental topology is shown in Figure 6.

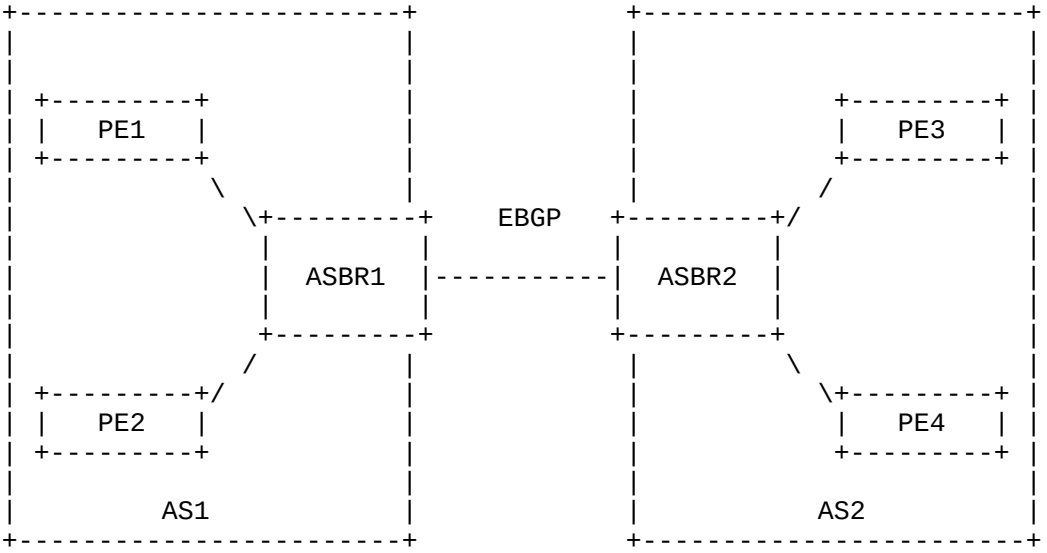

Figure 6 The experimental topology

This topology can be used to verify as follows:

* whether the VPN Prefix ORF mechanism could block the offending routes in intra-domain scenario.

* whether the VPN Prefix ORF mechanism conflicts with the existing mechanism and cause failure.

* whether the quota value leads to flapping.

Authors' Addresses

Wei Wang
China Telecom
Beiqijia Town, Changping District
Beijing
Beijing, 102209
China
Email: weiwang94@foxmail.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing
Beijing, 102209
China
Email: wangaj3@chinatelecom.cn

Haibo Wang
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing
Beijing, 100095
China
Email: rainsword.wang@huawei.com

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring, MD 20904
United States of America
Email: gyan.s.mishra@verizon.com

Jie Dong
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing
Beijing, 100095
China
Email: jie.dong@huawei.com