IDR Working Group
Internet-Draft
Intended status: Experimental
Expires: 2 August 2026

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

29 January 2026

VPN Prefix Outbound Route Filter (VPN Prefix ORF) for BGP-4
draft-ietf-idr-vpn-prefix-orf-25

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 Forwarding (VRF) instances are exchanged through a single shared Border Gateway Protocol (BGP) session.The purpose of VPN Prefix ORF mechanism is to control the overload of VPN routes based on RT. With this mechanism, the overload can be limited within the minimum range.

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 2 August 2026.

Copyright Notice

Table of Contents

1. Introduction

The BGP Maximum Prefix feature [RFC4486] is often used at the network boundary to control the number of prefixes injected into the network. However, in scenarios where VPN routes from multiple VRFs are advertised over a shared BGP session, there is a lack of appropriate methods to control route flooding within one VRF. This flooding can overwhelming the processing of VPN routes in other VRFs, consequently degrading their performance (e.g., causing route drops, processing delays, and abnormal customer services). Therefore, it is desirable to control excessive VPN route advertisements individually for each VRF within such a shared BGP session.

Several solutions 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, each existing solution has its own limitation as described in Section 4.

This draft defines a new type of Outbound Route Filter (ORF), called the VPN Prefix ORF. This 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 identifies the VPN prefix (Route Distinguisher (RD), Route Target (RT), source PE, etc.) of the overload VPN routes and sends a VPN Prefix ORF message to the BGP peer that announced these overload VPN routes. Upon receiving a VPN Prefix ORF entry, the BGP speaker filters and withdraws any overload VPN routes that were previously announced to its peer.

The purpose of this mechanism is to control the overload within the minimum scope and avoid route churn effects when a VRF overflows. The VPN Prefix ORF mechanism is applicable when VPN routes from different VRFs are exchanged via a shared BGP session.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. 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 [RFC4271]

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

* MPLS: Multi-Protocol Label Switching.

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

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

* 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.

3. Existing Solutions

3.1. Route Target Constraint (RTC)

RTC can only filter the VPN routes from any uninterested VRFs, if the route overload comes from an interested VRF, the RTC mechanism can’t filter them.

3.2. 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.

3.3. 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 overload of VPN prefixes dynamically.

3.4. PE-CE edge peer Maximum Prefix

The BGP Maximum-Prefix feature controls the number of prefixes received from a neighbor. Configuring it on every PE-CE link can prevent VPN route overflow. However, relying solely on sender-side protection is insufficient. If the sender has not configured Maximum Prefix, the VPN Prefix ORF mechanism can still prevent VPN route overflow.

3.5. Configuring the Maximum Prefix for each VRF on edge nodes

When a VRF overflows, some implementations may stop importing routes. Any additional VPN routes are held in the Routing Information Base (RIB). However, PEs still need to parse the incoming BGP messages, which consumes CPU cycles and further burdens the overflowed PE.

The VPN Prefix ORF mechanism improves upon this by enabling the overloaded PE to signal the specific overload routes back to the sender. The sender can then suppress these routes at the source, eliminating wasted processing and preserving resources for healthy VRFs.

4. VPN Prefix ORF Encoding

In this section, we describe the encoding of VPN Prefix ORF entries. The VPN Prefix 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 format of a ROUTE-REFRESH message carrying VPN Prefix ORF entries is as follow:

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

* SAFI (1 octet). 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. It is applicable for all types of EVPN routes as mentioned in [RFC7432].

* When-to-refresh (1 octet): the value MUST be 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 encoding of the common part is shown in Figure 1.

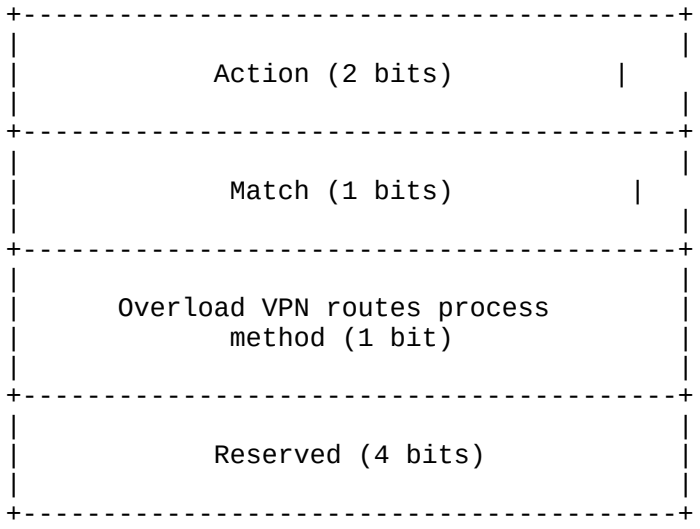

Figure 1: VPN Prefix ORF type-specific encoding

* Action (2 bits): the value is ADD, REMOVE or REMOVE-ALL as described in [RFC5291].

* Match (1 bit): the value is PERMIT or DENY as described in [RFC5291].

* Overload VPN routes process method (1 bit): if the value is set to 0, it means all overload 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 overload 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 2.

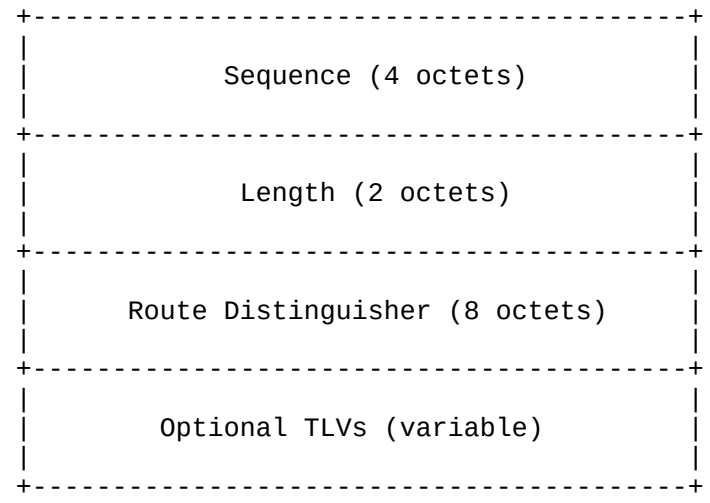

Figure 2: VPN Prefix ORF type-specific encoding

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

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

* Route Distinguisher: Distinguishes different user routes. The VPN Prefix ORF filters the VPN routes it intends to send based on Route Distinguisher. If the RD is set to 0, it indicates all VPN prefixes.

* Optional TLVs: Carries potential additional information to provide extensibility for the VPN Prefix ORF mechanism. Its format is shown in Figure 6. If one or more TLV(s) are unrecognized, the entire VPN Prefix ORF entry SHOULD be discarded.

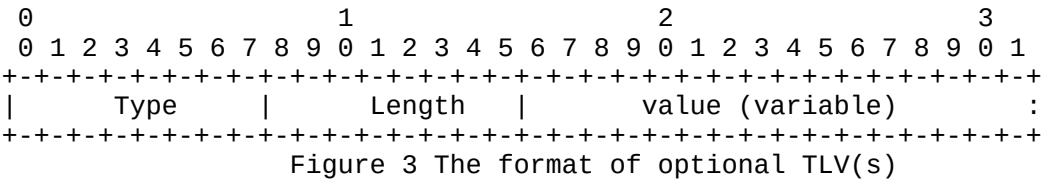

Figure 3 The format of optional TLV(s)

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 purpose of VPN Prefix ORF is to block unwanted VPN prefixes; therefore, the "Action" of a valid entry SHOULD be set to "DENY". To allow other permitted VPN prefixes to pass the filter, a default last-resort entry SHOULD be installed in advance in the VPN Prefixes ORF table, with the RD set to 0 and the corresponding Sequence set to 0xFFFFFFFF.

According to [RFC5291], if any field in a VPN Prefix ORF entry in the message contains an unrecognized value, the entire 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 by using the Outbound Route Filtering Capability defined in [RFC5291].

4.1. Source PE TLV (including 3 types)

The Source PE TLV is defined to identify the originator of the VPN routes. The sender of the VPN Prefix ORF MUST check for the existence of the Source PE Extended Community (SPE EC) on the VPN route being matched. If the SPE EC exists, the sender MUST include its value in the Source PE TLV. Otherwise, the value of Source PE TLV SHOULD be set to the Next Hop address.

The Source PE TLV SHOULD appear at most once within an individual ORF entry. If an ORF entry contains multiple Source PE TLVs, the entire ORF entry MUST be ignored.

The source PE TLV supports the following types:

* IPv4 Source PE TLV: Type = 1, Length = 4 octets, value = Next Hop address in IPv4 format.

* IPv6 Source PE TLV: Type = 2, Length = 16 octets, value = Next Hop address in IPv6 format (global IPv6 address only).

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

4.2. Source AS TLV

The Source AS TLV is defined to identify the source AS number of the source PE. It is only required in inter-domain scenarios.

The Source AS TLV SHOULD appear at most once within an individual ORF entry. If an ORF entry contains multiple Source AS TLVs, the entire entry SHOULD be ignored.

The encoding of Source AS TLV is as follows:

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

4.3. Route Target TLV

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

If this TLV contains only one RT but multiple RTs are configured on the VPN route, the device SHOULD check whether the RT included in this TLV exists among the configured RTs. If so, the device SHOULD filter out the VPN route.

The Route Target TLV supports the following type:

> Type = 5, Length = 8*n (where n is the number of RTs associated with the overloaded VPN routes) octets, value = the RT value(s) of the overload VPN routes. If multiple RTs are included, an exact match is required.

5. The general procedures of VPN Prefix ORF mechanism

5.1. Process of VPN Prefix ORF mechanism on sender

The operation of the VPN Prefix ORF mechanism on each device is independent. Each device 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 according to their specific circumstances.

This section describes the procedures for the receiving BGP peer to process VPN route information from the sending BGP peer. The VPN information includes updated VPN routes and their corresponding VPN instance identification information. Based on this identification information, the receiving BGP peer determines the newly added VPN routes and checks whether the number of these routes causes 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 (identified by the VPN instance identification information) is reached or exceeded, the receiving BGP peer sends a VPN Prefix ORF message to the sending BGP peer. This message indicates that the sender SHOULD stop transmitting the corresponding VPN routes identified by the information.

Before originating a VPN Prefix ORF message, the device SHOULD compare the list of RTs carried by the VPN routes with those imported by other VRFs on the device. If a 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, the sender SHOULD send a "default" entry to the VPN Prefix ORF receiver, to allow other permitted VPN prefixes to pass the filter. The "default" entry should be pre-installed in the VPN Prefixes ORF table, with the overload VPN route processing method set to 0, sequence set to 0xFFFFFFFF, the length set to 8, and Route Distinguisher set to 0.

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

The instruction information sent by the receiving BGP peer includes the following details:

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

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

* A Match field (in ORF entries) set to a value that instructs the sending BGP peer to deny VPN route updates matching the corresponding ORF entries.

* An RD value (identifying the above mentioned VPN instance) added to the type-specific part of the ORF entries.

When multiple VRFs on a PE receive VPN routes with a specific RD, the PE sends a VPN Prefix ORF message if one of these VRFs exceeds its limit for routes with that RD. This prevents other non-exceeded VRFs from receiving VPN routes containing the same RD, thereby avoiding communication disruptions between these VRFs and the rejected VPN routes. In order to more finely control VPN routing, if 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 alarm information to network operators.

The procedures for senders of VPN Prefix ORF entries are described below:

```
S01. For each VRF v that receives updated VPN routes {
S02.     If (the total number of prefixes in VRF v exceeds its
         configured prefix limit) {
S03.         RT_set = the set of Route Targets imported by VRF v.
S04.         overload_RD_source_pairs = all <RD, Source PE>
             tuples from the newly received routes that belong
             to VRF v.

             // Check if any RT in RT_set is also imported by
                another VRF that has NOT exceeded its limit
S05.         conflict_exists = FALSE;
S06.         For each RT r in RT_set {
S07.             For each other VRF u on this device {
S08.                 If (r is in the import RT list of VRF u)
                     AND (the prefix count of VRF u <= its
                     prefix limit) {
S09.                     conflict_exists = TRUE;
S10.                 }
S11.             }
S12.         }

S13.         If (conflict_exists == TRUE) {
```

```
S14.             // Cannot send ORF: would block routes needed
                    by healthy VRFs
S15.             Send warning message to the operator.
S16.         }

S17.         // Safe to send ORF entries
S18.         For each <RD_x, PE_y> in overload_RD_source_pairs {
S19.             Collect all RTs carried by routes with RD=RD_x
                 from source PE_y that are imported into VRF v.

S20.             Construct a VPN Prefix ORF entry with:
S21.                 Action = ADD,
S22.                 Match = DENY,
S23.                 Overload VPN routes process method = 0,
S24.                 Sequence = Generate unique sequence number,
S25.                 Route Distinguisher = RD_x,
S26.                 Optional TLVs include:
S27.                     Source PE TLV = PE_y,
S28.                     Route Target TLV = RT_list.

S29.             Send a BGP ROUTE-REFRESH message containing this
                 ORF entry to the upstream BGP peer (e.g., RR).
S30.             Send an alarm message to the operator indicating
                 VRF v overflow and ORF transmission.
S31.         }
S32.     } Else {
S33.         // No overflow in this VRF; no ORF triggered
S34.         Continue normal route processing.
S35.     }
S36. }
```

5.1.1. Intra-domain Scenarios and Solutions

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

* unique RD (per VPN, per PE).

* the same RD (per VPN, same on all PEs)

Detailed descriptions about the above solutions are in provided Appendix B.

5.2. Protocol 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 a VPN Prefix ORF entry, it MUST check first whether the "Match" bit is "DENY" or not.

If the "Match" bit is "PERMIT", and is the "default" entry (the overload 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 MUST be discarded and a warning MUST 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 an RR, ASBR or PE) performs the following actions upon receiving a VPN Prefix ORF entry from its BGP peer:

```
S01. The receiver checks the combination of <AFI/SAFI, ORF-Type,
     Sequence, Route Distinguisher> in 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 if (Action is set to ADD) {
S05.         The receiver overwrite the old VPN Prefix ORF entry
             with the new one.
S06. } else if (Action is set to REMOVE) {
S07.         The receiver removes the corresponding VPN Prefix ORF
             entry from the ORF-Policy table.
S07. } else {
             The receiver SHOULD remove all VPN Prefix ORF entries
             from the ORF-Policy table.
S08. }
```

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

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

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

```
S01. RR or ASBR checks if there are matching filtering conditions
     in the ORF-Policy table for the VPN routes.
S02. If (no matching filtering conditions exist) {
S03.     The RR/ASBR sends the VPN routes.
S04. } else {
S05.     If (the "Overload 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 stops 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 stops sending
             the corresponding VPN routes to the sender of the
             VPN Prefix ORF entry.
S09. }
```

The procedure above can be used for route refresh processing after receiving an ORF update and the usual VPN route propagation. A change to the ORF prefixes triggers a rescan of the relevant routing information, followed by a route refresh. In contrast, regular individual VPN route updates are only subject to matching against the existing ORF rules.

6. Source PE Extended Community

Next Hop does not always identify the source as seen 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 an Option B inter-domain scenario, the ASBR will change the Next Hop.

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

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

The format of SPE EC is shown as Figure 4.

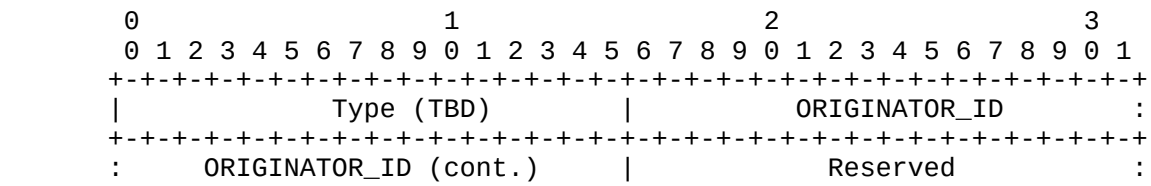

Figure 4 The format of SPE EC

Where:

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

* ORIGINATOR_ID: Specifies the identifier of the source PE.

* Reserved: MUST be set to zero on transmission.

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

* Check for the existence of the SPE EC. If it exists, the RR/ASBR MUST NOT change it.

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

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

This section extends route reflection behaviours, meaning that if support for this feature extension is required, the RR MUST perform the additional actions specified above.

7. Operational Considerations

7.1. Quota value calculation

The VPN Prefix ORF mechanism is designed for intra-domain BGP/MPLS IP VPN [RFC4364] and BGP/MPLS Ethernet VPN (EVPN) [RFC7432] deployments where multiple VRFs on a Provider Edge (PE) router exchange VPN routes via a single shared iBGP session (typically with a Route Reflector).

This mechanism operates in two modes:

* Basic mode: Triggered solely by VRF-level prefix limits. No per-source quota configuration is required. In this mode, the PE sends a VPN Prefix ORF only if all VRFs that import the same Route Target(s) have exceeded their respective prefix limits.

* Granular mode (optional): Enabled when operators configure per-<Route Distinguisher, Source PE> quotas via their Network Management System (NMS) or CLI. This enables finer-grained control, allowing ORF triggering even if only one VRF exceeds its limit while others sharing the same RT remain healthy, provided that the overload routes originate from a specific source.

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).

When the granular mode is enabled, an operator may configure a quota for each <RD, Source PE> tuple imported into a VRF. This quota represents the maximum number of prefixes allowed from that specific source for the given RD.

The quota value can be derived based on historical traffic patterns, service level agreements (SLAs), or static provisioning via NMS/CLI. It is not a prerequisite for the VPN Prefix ORF mechanism to operate; the mechanism defaults to VRF-level prefix limit enforcement if no per-source quotas are configured.

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.

7.2. 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 "overload 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.

8. Security Considerations

On devices that support VPN Prefix ORF mechanism, it is necessary to enforce a per-peer limit on the number of VPN Prefix ORF entries. Once this limit is exceeded, the peer will ignore all newly received VPN Prefix ORF entries.

Others security considerations are aligned with [RFC4271].

9. IANA Considerations

9.1. VPN Prefix Outbound Route Filter

This document defines a new Outbound Route Filter type, entitled "VPN Prefix Outbound Route Filter (VPN Prefix ORF)", and assigns a value of 66 from the BGP Outbound Route Filtering (ORF) Types space which is under the "Border Gateway Protocol (BGP) Parameters" registry group.

Value	Description	Reference
66	VPN Prefix ORF	This document

9.2. VPN Prefix ORF TLV types

This document defines a new "VPN Prefix ORF TLV Type" subregistry in the "Border Gateway Protocol (BGP) Parameters" registry. The registration policies, per [RFC8126], for this subregistry are as follows:

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

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

IANA should make initial assignments as follows:

Value	Description	Reference
0	Reserved	This document
1	IPv4 Source PE TLV	This document
2	IPv6 Source PE TLV	This document
3	Source PE Identifier TLV	This document
4	Source AS TLV	This document
5	Route Target TLV	This document
6-127	Unassigned	
128-255	Unassigned	

9.3. Source PE Extended Community

This document defines a new BGP Transitive Extended Community Type called "Source PE Extended Community". Codepoint 0x0d is suggested to be allocated to the Source PE Extended Community.

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

9.4. Commen part of ORF entry

IANA needs to make a new "ORF Entry Bits" registry in the "Border Gateway Protocol (BGP) Parameters" registry. The registration policy for this subregistry is IETF Review.

IANA should make initial assignments as follows:

Bit Position	Name	Description	Reference
0-1	Action	The value of this field is 0 for ADD,1 for REMOVE, and 2 for REMOVE-ALL.	RFC5291
2	Match	The value of this field is 0 for PERMIT and 1 for DENY.	RFC5291
3	Overload VPN routes process method	The value of this field is 0 for withdrawn all overload VPN routes, and 1 for refusing to receive new overload VPN routes.	This document
4-7	Reserved		RFC5291

10. Contributor

Shunwan Zhuang

Huawei Technologies

Huawei Building, No.156 Beiqing Rd.

Beijing

Beijing, 100095 China

11. 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.

12. 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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

Appendix A. Experimental topology

The experimental topology is shown in Figure 5.

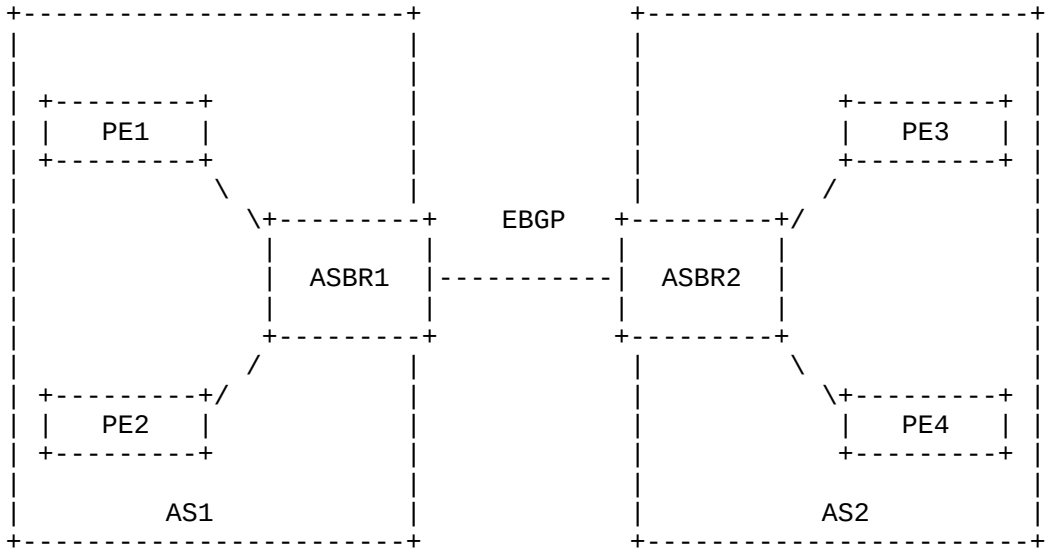

Figure 5 The experimental topology

This topology can be used to verify the following:

* whether the VPN Prefix ORF mechanism can block overload routes in intra-domain scenarios.

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

* whether the quota value leads to route flapping.

This draft is experimental in order to determine if the proposed mechanism could block the overload routes as expected or not, and whether it would cause other potential network failures or operational challenges. The status of the document may be changed to proposed standard once there is sufficient deployment experience and issues identified, if any, are addressed.

Appendix B. Intra-domain Scenarios and Solutions

This section describes the workflow of some example scenarios for illustrative purposes.

B.1. Scenario 1: unique RD (per VPN, per PE)

In this scenario, PE1-PE4 and RR are iBGP peers. RD is allocated per VPN per PE. The overload VPN routes only carry one RT. We assume that the network topology is shown in Figure 6.

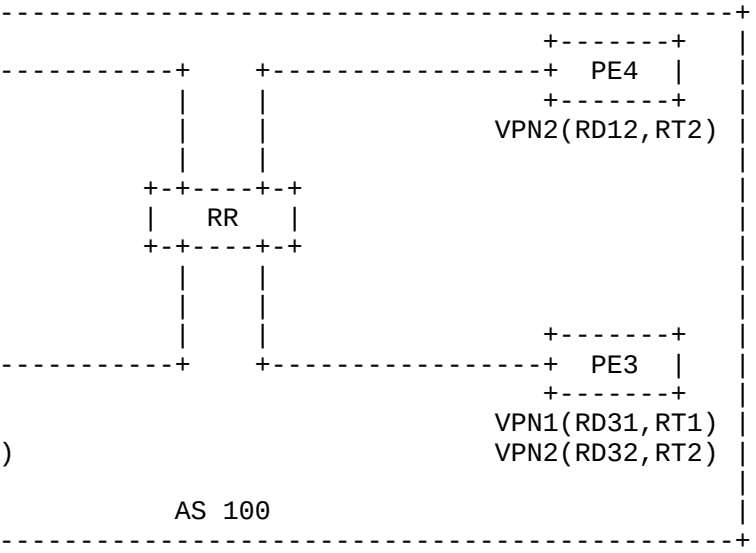

Figure 6 Network Topology of Scenario 1

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

a) PE1

If no quota value is set on PE1 and each VRF on PE1 has a prefix limit, when PE1 receives VPN routes from its BGP peer, it performs the following actions:

```
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 carries the
         following parameters: RD set to RD31, RT
         value set to RT1, and source PE set to PE3.
         The RR processes the overload VPN routes
         and controls the number of VPN routes based
         on the value of the "Overload VPN routes
         process method" field.
S03. } else {
S04.     PE1 does not trigger the VPN Prefix ORF
         mechanism and only performs VPN route
         filtering for the target VRF.
S05. }
```

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

If a quota is configured for each <RD31, source PE3> tuple imported into a VRF and each VRF has a prefix limit, when PE1 receives VPN routes from its BGP peer, it performs the following actions:

```
S01. If (VPN routes associated with <RD31, PE3>
     tuple exceed the quota) {
S02.     If (the prefix limit of the VPN1 VRF
         is not exceeded) {
S03.         PE1 sends a warning message to the
             operator, and the VPN Prefix ORF
             mechanism is not triggered.
S04.     } else {
S05.         PE1 generates a BGP ROUTE-REFRESH
             message containing a VPN Prefix ORF
             entry with the parameters (RD = RD31,
             source PE = PE3, RT = RT1), and sends
             this entry to the RR.
             The RR proceses the overload VPN
             routes based on the value of the
             "Overload VPN routes process method".
S06.     }
S07. }
```

b) PE2

If no quota value is set on PE2 and each VRF on PE2 has a prefix limit, when PE2 receives VPN routes from its BGP peer, it performs the following actions:

```
S01. If (the prefix limit for the VPN1 VRF is exceeded) {
S02.     If (the prefix limit for the 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 specifies the RD set to RD31 and the RT
             value set to RT1. The RR processes the overload VPN
             routes and controls the number of VPN routes based
             on the value of the "Overload VPN routes process
             method" field.
S04.     } else {
S04. } else {
S05.     PE2 does not trigger the VPN Prefix ORF mechanism and
         only performs VPN route filtering for the target VRF.
S06. }
```

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

If a quota is configured for each <RD31, source PE3> tuple imported into a VRF and each VRF has a prefix limit, when PE2 receives VPN routes from its BGP peer, it performs the following actions:

```
S01. If (the VPN routes associated with the <RD31, PE3> tuple
     exceed the quota) {
S02.     If (the prefix limit of the VPN1 VRF is not exceeded) {
S03.         PE2 sends a warning message to the operator, and the
             VPN Prefix ORF mechanism is not triggered.
S04.     } else {
S05.         If (the prefix limit of the VPN2 VRF is not exceeded)
             {
S06.             PE2 does not trigger the VPN Prefix ORF mechanism
                 and only performs VPN route filtering for the
                 target VPN1 VRF, stopping the import of VPN routes
                 associated with <RD31, PE3>.
S07.         } else {
S08.             PE2 generates a BGP ROUTE-REFRESH message
                 containing a VPN Prefix ORF entry with the
                 parameters (RD31, source PE = PE3, RTs = RT1 and
                 RT2), and sends this entry to the RR. The RR
                 processes the overload VPN routes based on the
                 value of the "Overload VPN routes process method"
                 field.
S09.         }
S10.     }
S11. }
```

B.2. Scenario 2: the same RD (per VPN, same on all PEs)

In this scenario, PE1-PE4 and RR are iBGP peers. RD is allocated per VPN. One/Multiple RTs are associated with the overload VPN routes and are imported into different VRFs on other devices. We assume the network topology is shown in Figure 7.

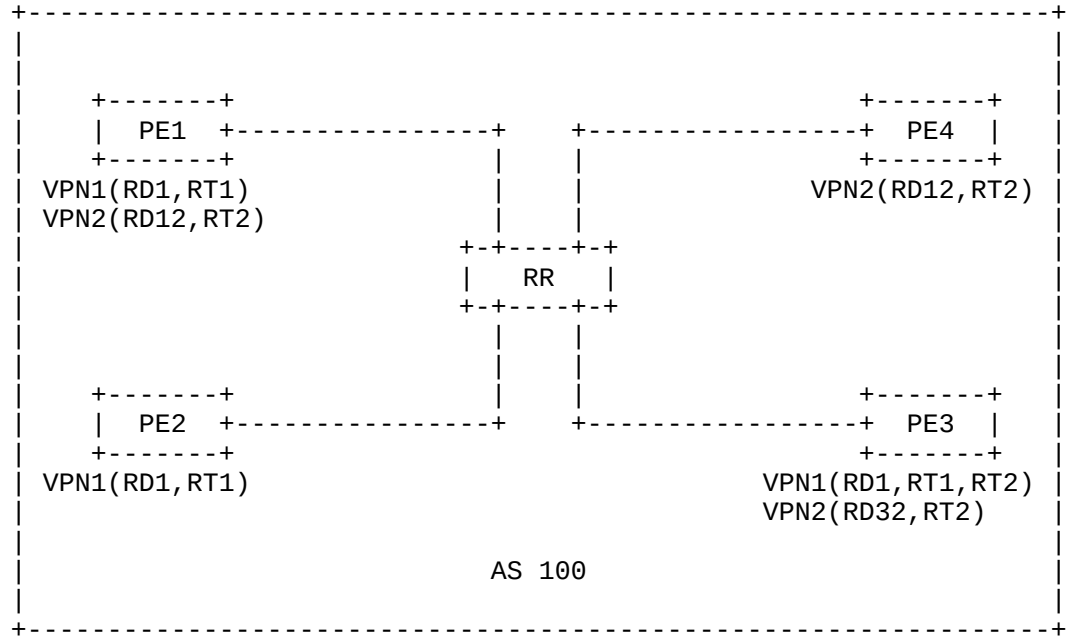

Figure 7 Network Topology of Scenario 2

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 overload VPN routes can affect the performance of the VRFs on PEs.

a) PE1

If no quota value is set on PE1 and each VRF on PE1 has a prefix limit, PE1 does not directly trigger the VPN Prefix ORF mechanism when the prefix limit of the VPN1 VRF is exceeded, because the VPN2 VRF imports VPN routes with RT2. This case is similar to that of PE2 without a quota in Scenario 1, with modifications 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 specifies the RD set to RD1, the RT values set to RT1 and RT2, and the source PE identified as PE3. The RR processes the overload VPN routes and controls the number of VPN routes based on the value of the "Overload VPN routes process method" field.

If a quota is configured for each <RD1, source PE3> tuple imported into a VRF and each VRF has a prefix limit, this case is similar to that of PE2 with a quota in Scenario 1, with modifications as follows:

S08. PE1 generates a BGP ROUTE-REFRESH message containing a VPN Prefix ORF entry with the parameters (RD1, source PE = PE3, RTs = RT1 and RT2), and sends this entry to the RR. The RR processes the overload VPN routes based on the value of the "Overload VPN routes process method" field.

b) PE2

If no quota value is set on PE2 and each VRF on PE2 has a prefix limit, since only the VPN1 VRF needs to import VPN routes with RT1, this case is similar to that of PE1 without a quota in Scenario 1, with modifications 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 specifies the RD set to RD1, the RT values set to RT1 and RT2, and the source PE identified as PE3. The RR processes the overload VPN routes and controls the number of VPN routes based on the value of the "Overload VPN routes process method" field.

If a quota is configured for each <RD31, source PE3> tuple imported into a VRF and each VRF has a prefix limit, this case is similar to that of PE1 with a quota in Scenario 1, with modifications as follows:

S05. PE2 generates a BGP ROUTE-REFRESH message containing a VPN Prefix ORF entry with the parameters (RD1, source PE = PE3, RTs = RT1 and RT2), and sends this entry to the RR. The RR processes the overload VPN routes based on the value of the "Overload VPN routes process method" field.

Appendix C. Applicability

Using scenario 1 in Appendix B, we demonstrate how to determine each field when the sender generates a VPN Prefix ORF entry. Assuming an IPv4 network, after PE1-PE4 and RR have advertised the Outbound Route Filtering Capability, each of PE1-PE4 needs to 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
* Overload 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 that 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
* Overload 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

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