

IDR Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 4, 2022

W. Wang
A. Wang
China Telecom
H. Wang
Huawei Technologies
G. Mishra
Verizon Inc.
S. Zhuang
J. Dong
Huawei Technologies
March 3, 2022

VPN Prefix Outbound Route Filter (VPN Prefix ORF) for BGP-4
draft-wang-idr-vpn-prefix-orf-02

Abstract

This draft defines a new Outbound Route Filter (ORF) type, called the VPN Prefix ORF. The described VPN Prefix ORF mechanism is applicable when the VPN routes from different VRFs are exchanged via one shared BGP session (e.g., routers in a single-domain connect via Route Reflector).

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 September 4, 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents

(<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Conventions used in this document	4
3.	Terminology	4
4.	Operation process of VPN Prefix ORF mechanism on sender . . .	4
4.1.	Intra-domain Scenarios and Solutions	5
4.1.1.	Scenario-1 and Solution (Unique RD, One RT)	5
4.1.2.	Scenario-2 and Solution (Unique RD, Multiple RTs) . . .	7
4.1.3.	Scenario-3 and Solution (Universal RD)	8
5.	Operation process of VPN Prefix ORF mechanism on receiver . .	8
6.	Withdraw of VPN Prefix ORF entries	9
7.	VPN Prefix ORF Encoding	9
7.1.	Source PE TLV	10
7.2.	Route Target TLV	11
8.	Implementation Considerations	11
9.	Security Considerations	11
10.	IANA Considerations	12
11.	Acknowledgement	12
12.	Normative References	12
	Authors' Addresses	13

[1.](#) Introduction

[I-D.wang-idr-vpn-routes-control-analysis] analysis the scenarios and necessities for VPN routes control in the shared BGP session. This draft analyzes the existing solutions and their limitations for these scenarios, proposes the new VPN Prefix ORF solution to meet the requirements that described in section 8 of [\[I-D.wang-idr-vpn-routes-control-analysis\]](#).

Now, there are several solutions can be used to alleviate these problem:

- o Route Target Constraint (RTC) as defined in [\[RFC4684\]](#)
- o Address Prefix ORF as defined in [\[RFC5292\]](#)
- o CP-ORF mechanism as defined in [\[RFC7543\]](#)

- o PE-CE edge peer Maximum Prefix
- o Configure 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 the uninterested VRFs, if the "trashing routes" come from the interested VRF, filter on RTs will erase all prefixes from this VRF.

2) Address Prefix ORF

Using Address Prefix ORF to filter VPN routes need to pre-configuration, but it is impossible to know which prefix may cause overflow in advance.

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 the BGP/MPLS Ethernet VPN (EVPN) [[RFC7432](#)] networks, but its main aim is also to get the wanted VPN prefixes and can't be used to filter the 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, which is not desirable for the fining control of the VPN Prefixes advertisement.

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

When a VRF overflows, it stops the import of routes and log the extra VPN routes into its RIB. However, PEs still need to parse the BGP updates. These processes will cost CPU cycles and further burden the overflowing PE.

This draft defines a new ORF-type, called the VPN Prefix ORF. This mechanism is event-driven and does not need to be pre-configured. When a VRF of a router overflows, the router will find out the VPN prefix (RD, RT, source PE, etc.) of offending VPN routes in this VRF, and send a VPN Prefix ORF to its BGP peer that carries the relevant information. If a BGP speaker receives a VPN Prefix ORF entry from its BGP peer, it will filter the VPN routes it tends to send according to the entry.

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 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. Operation process of VPN Prefix ORF mechanism on sender

The operation of VPN Prefix ORF mechanism on each device is independent, each of them makes a local judgement to determine whether it needs to send VPN Prefix ORF to its peers. The operators need to make sure the algorithms in different devices consistent. On PE, each VRF has a prefix limit, and routes associated with each <RD, source PE> tuple has a pre-configured quota.

- o when routes associated with <RD, source PE> tuple past the quota but the prefix limit of VRF is not exceeded, PE should send warnings to the operator, and the VPN Prefix ORF mechanism should not be triggered.
- o when routes associated with <RD, source PE> tuple past the quota and the prefix limit is exceeded and there is no other VRFs on offended PE need VPN routes with this RD, they should be dropped via VPN Prefix ORF mechanism.

When the VPN Prefix ORF mechanism is triggered, the device must send an alarm information to network operators.

4.1. Intra-domain Scenarios and Solutions

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

- o RD is allocated per VPN per PE, each VRF only import one RT(see [Section 4.1.1](#)).
- o 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](#)).
- o 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, RD is allocated per VPN per PE. The offending VPN routes only carry one RT. We assume the network topology is shown in Figure 1.

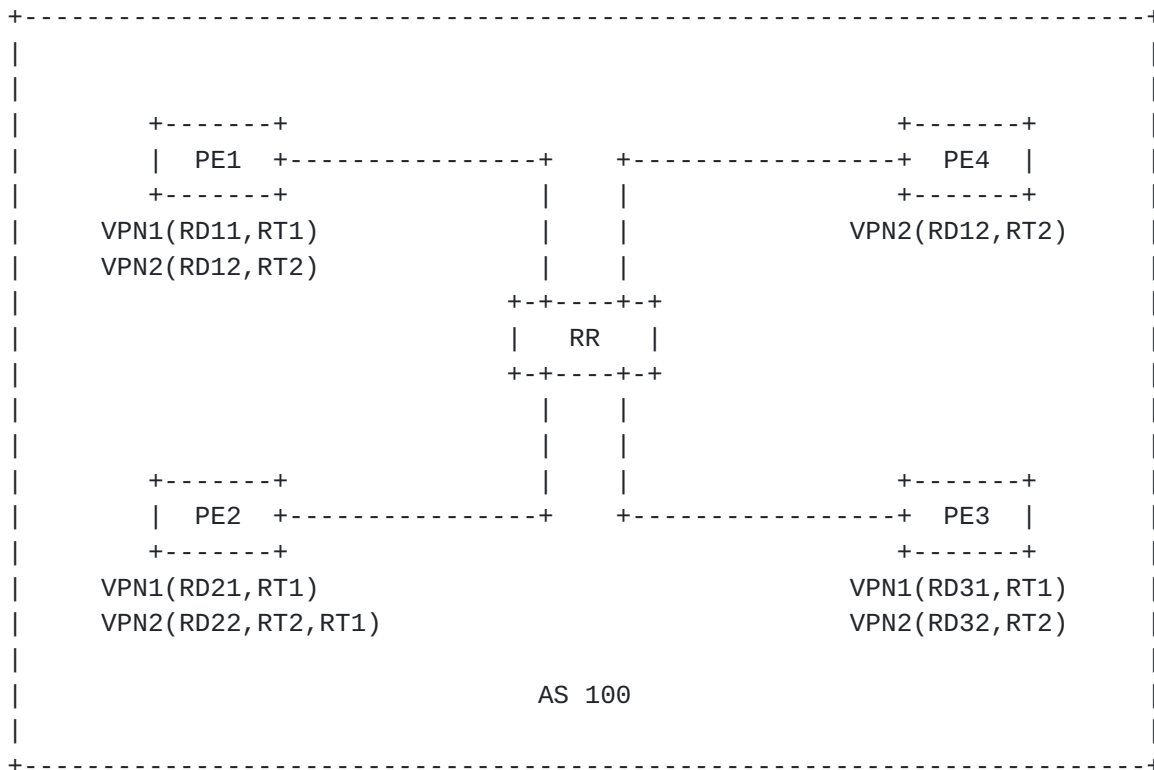


Figure 1 Network Topology of Scenario-1

When PE3 sends excessive VPN routes with RT1, while 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 some mechanisms to identify and control the advertisement of offending VPN routes.

On PE1, each VRF has a prefix limit, and each <RD, source PE> tuple imported into VRF has a quota. When routes associated with <RD31, PE3> tuple past the quota but the prefix limit of VPN1 VRF is not exceeded, PE1 sends a warning message to the operator, and the VPN Prefix ORF mechanism should not be triggered. After the prefix limit of VPN1 VRF is exceeded, due to there is no other VRFs on PE1 need the VPN routes with RT1, PE1 will generate a BGP ROUTE-REFRESH message contains a VPN Prefix ORF entry, and send to RR. RR will withdraw and stop to advertise such offending VPN routes (RD31, RT1, source PE is PE3) to PE1.

On PE2, both VPN1 VRF and VPN2 VRF import VPN routes with RT1. If PE2 triggers VPN Prefix ORF mechanism when VPN1 VRF overflows, VPN2 VRF cannot receive VPN routes with RT1 as well. PE2 should not trigger the VPN Prefix ORF mechanism to RR (RD31, RT1, source PE is PE3) until all the VRFs that import RT1 on it overflow.

On PE2, due to there is only one VRF imports VPN routes with RT1. If it overflows, it will trigger VPN Prefix ORF (RD31, RT1, comes from PE3) mechanisms. RR will withdraw and stop to advertise such offending VPN routes to PE2.

The receiver of VPN Prefix ORF entries may be a RR or PE. As it receives the VPN Prefix ORF entries, it will check <AFI/SAFI, ORF-Type, Sequence, Route Distinguisher> to find if it already existed in its ORF-Policy table. If not, the receiver will add the VPN Prefix ORF entries into its ORF-Policy table; otherwise, the receiver will

discard it. Before the receiver send a VPN route, it will check its ORF-Policy table whether there is a related VPN Prefix ORF entry or not. If not, the receiver will send this VPN route; otherwise, the receiver will stop sending that VPN route to its peer.

6. Withdraw of VPN Prefix ORF entries

When the VPN Prefix ORF mechanism is triggered, the alarm information will be generated and sent to the network operators. Operators should manually configure the network to resume normal operation. Due to devices can record the VPN Prefix ORF entries sent by each VRF, operators can find the entries needs to be withdrawn, and trigger the withdraw process as described in [[RFC5291](#)] manually. After returning to normal, the device sends withdraw ORF entries to its peers who have previously received ORF entries.

7. 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:

- 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 VPN Prefix 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)

VPN Prefix ORF also contains type-specific part. The encoding of the type-specific part is shown in Figure 4.

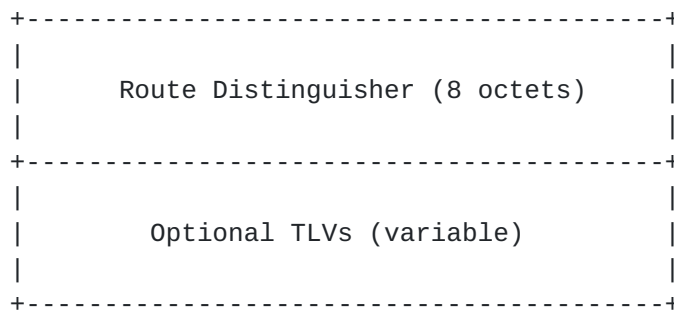


Figure 4: VPN Prefix ORF type-specific encoding

- o Route Distinguisher: distinguish the different user routes. The VPN Prefix ORF filters the VPN routes it tends to send based on Route Distinguisher.
- o Optional TLVs: carry the potential additional information to give the extensibility of the VPN Prefix ORF mechanism.

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:

- o The ORF-Type MUST be set to VPN Prefix 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 set to DENY.

7.1. Source PE TLV

Source PE TLV is defined to identify the source of the VPN routes. Using source PE and RD to filter VPN routes together can achieve more refined route control. The source PE TLV contains the following types:

- o In single-domain or Option C cross-domain scenario, NEXT_HOP attribute is fixed during routing transmission, so it can be used as source address.

Type = 1, Length = 4 octets, value = NEXT_HOP.

Type = 2, Length = 16 octets, value = NEXT_HOP.

- o In Option B or Option AB cross-domain scenario, NEXT_HOP attribute may be changed by ASBRs and cannot be used as the source address. The originator can be traced by the Route Origin Community in BGP (as defined in [Section 5 of \[RFC4360\]](#)).

Type = 3, Length = 6 octets, value = the value field of Route Origin Community.

In principle, when the device can extract Route Origin Community from the received packet, the value of Source PE TLV should be set to Route Origin Community; Otherwise, the value should be set to NEXT_HOP.

7.2. 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 encoding of Route Target TLV is following:

Type = 4, 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.

8. Implementation Considerations

Before originating an VPN Prefix ORF message, the device should compare the list of RD and RT(s) carried by VPN routes signaled for filtering and the RD and RT(s) imported by not affected VRF(s). Once they have intersection, the VPN Prefix ORF message MUST NOT be originated.

Since the motivation of VPN Prefix ORF mechanism is to filter the unwanted VPN routes, "Match" field of VPN Prefix ORF message MUST be set to "DENY". The default behavior of the VPN Prefix ORF MUST be PERMIT-ALL.

9. Security Considerations

A BGP speaker will maintain the VPN Prefix ORF entries in an ORF-Policy table, 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.

10. IANA Considerations

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

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

Registry	Type	Meaning
IPv4 Source PE TLV	1	IPv4 address for source PE.
IPv6 Source PE TLV	2	IPv6 address for source PE.
ROC Source PE TLV	3	Route Origin Community for Source PE.
Route Target TLV	4	Identify the RT of the offending VPN routes

Figure 5: IANA Allocation for newly defined TLVs

11. Acknowledgement

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

12. Normative References

- [I-D.ietf-bess-evpn-inter-subnet-forwarding]
Sajassi, A., Salam, S., Thoria, S., Drake, J. E., and J. Rabadan, "Integrated Routing and Bridging in Ethernet VPN (EVPN)", [draft-ietf-bess-evpn-inter-subnet-forwarding-15](#) (work in progress), July 2021.
- [I-D.wang-idr-vpn-routes-control-analysis]
Wang, A., Wang, W., Mishra, G. S., Wang, H., Zhuang, S., and J. Dong, "Analysis of VPN Routes Control in Shared BGP Session", [draft-wang-idr-vpn-routes-control-analysis-04](#) (work in progress), September 2021.

- [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>>.
- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", [RFC 4360](#), DOI 10.17487/RFC4360, February 2006, <<https://www.rfc-editor.org/info/rfc4360>>.
- [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>>.
- [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>>.
- [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>>.

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

Phone: 301 502-1347
Email: gyan.s.mishra@verizon.com

Shunwan Zhuang
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing, Beijing 100095
China

Email: zhuangshunwan@huawei.com

Jie Dong
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing, Beijing 100095
China

Email: jie.dong@huawei.com