IDR Working Group
Internet-Draft
Intended status: Informational
Expires: August 26, 2021

A. Wang
W. Wang
China Telecom
G. Mishra
Verizon Inc.
H. Wang
S. Zhuang
J. Dong
Huawei Technologies
February 22, 2021

Analysis of VPN Routes Control in Shared BGP Session
draft-wang-idr-vpn-routes-control-analysis-00

Abstract

This draft analyzes some scenarios and the necessaries for VPN routes control in the shared BGP session, which can be the used as the base for the design of related solutions.

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 August 26, 2021.

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 overwhelm the process of VPN routes in other VRFs. That is to say, the excessive VPN routes advertisement should be controlled individually for each VRF in such shared BGP session.

The following sections analyzes the scenarios that are necessary to such mechanism.

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]

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

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

4. Inter-AS VPN Option B/AB Scenario

For inter-AS VPN deployment option B/AB scenario, as described in Figure 1, there is one BGP session between ASBR1 and ASBR2, which is used to advertise the VPN routes from VPN1 and VPN2 VRF. Normally the operator will deploy the BGP maximum prefixes feature under different address families between the ASBR1 and ASBR2, but the threshold must be set very high to cope with the situation when all the VRFs in each family reach their VPN routes limit simultaneously. In case VPN routes in only one of VRF, for example VPN1 in PE3, advertises excess VPN routes(with RD set to RD1 and RT import/export set to RT1. Configurations on other PEs are similar) into the network, but VPN routes advertisement in other VRFs are in normal, the prefix bar set between the ASBRs will not take effect. Such excessive VPN routes will be advertised into the AS1, to PE1 and PE2 respectively.

PE1 in this example, provides the services for VPN2 at the same time. If it receives the excessive VPN routes for VPN1 from ASBR1, although such VPN routes have exceeded the limit within the VRF VPN1, it can't break the BGP session with ASBR1 directly. All it can do is to receive and process the excessive BGP updates continuously, parse the excessive VPN routes for VPN1 and drop it, extract the VPN routes for VPN2 and install it.

Doing so can certainly influence the performance of PE1 to serve the other VPN services on it, considering that there are hundreds of VRFs deployed on it.

PE1 should have the capability to control the advertisement of specified excessive VPN routes from its BGP peer. The ASBR should also have such capability.

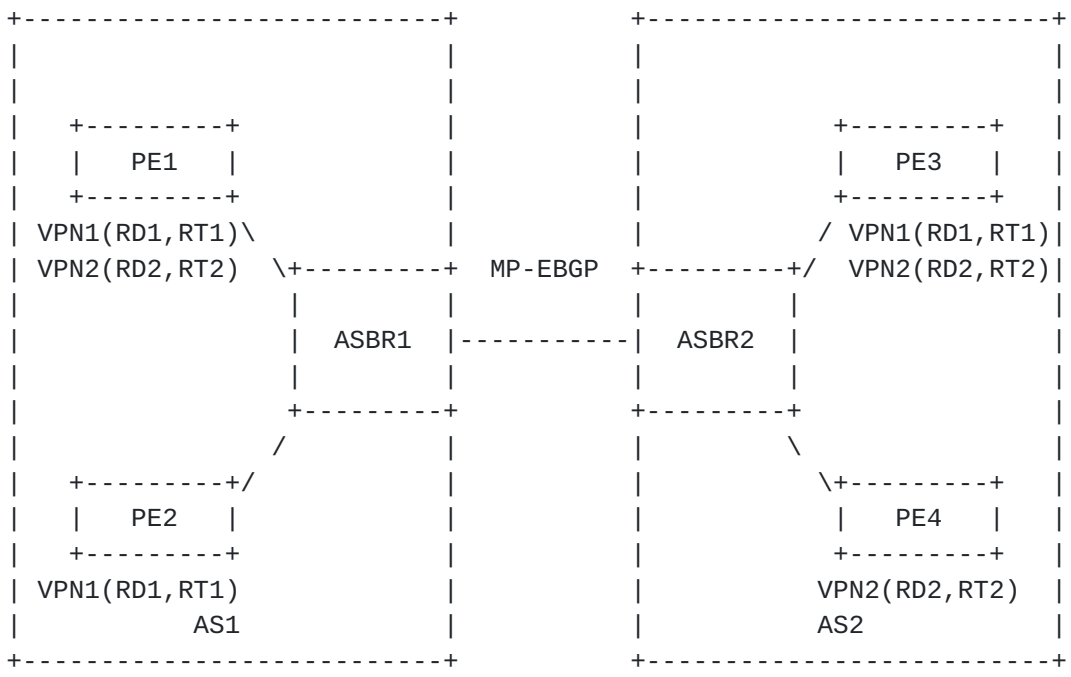

Figure 1: The Option B/Option AB cross-domain scenario

5. Inter-AS VPN Option C Scenario

For inter-AS VPN deployment option C scenario, as that described in Figure 2, there is one BGP session between RR1 and RR2, which is used to advertise the VPN routes from all the VRFs that located on the edge routers(PE1 and PE2). The BGP maximum prefix bar can't also prevent the excessive advertisement of VPN routes in one VRF, and such abnormal behavior in one VRF can certainly influence the performances of PEs to serve other normal VRFs.

PE and RR should all have some capabilities to control the specified excessive VPN routes to be advertised from its upstream BGP peer.

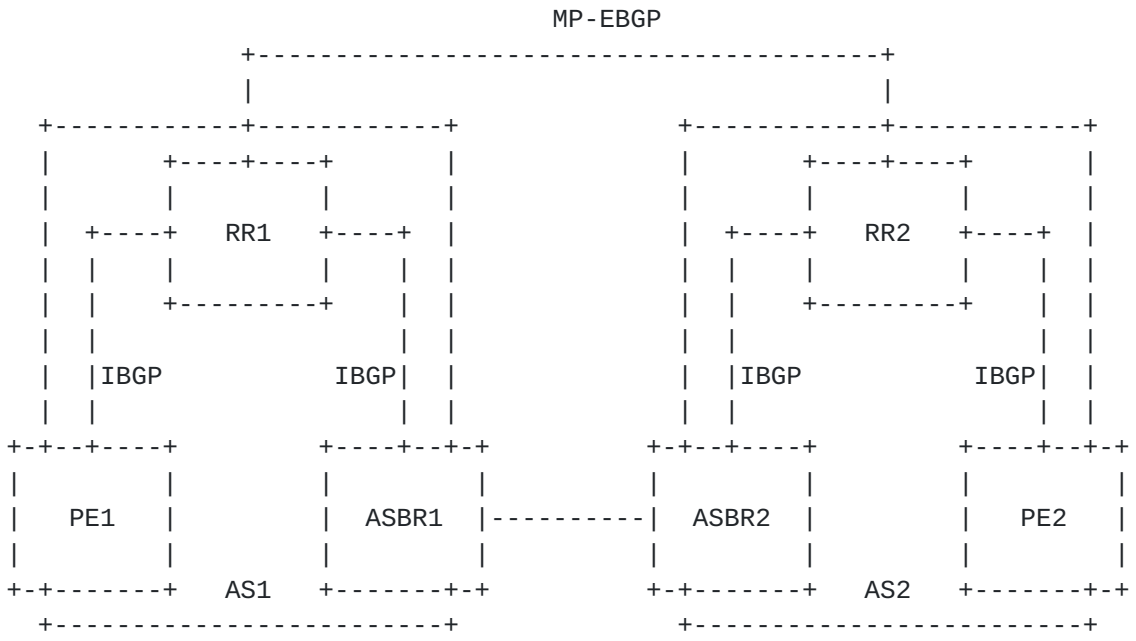

Figure 2: The Option C cross-domain scenario

6. Intra-AS VPN RR Deployment Scenario

For intra-AS VPN deployment, as depicted in Figure 3, if the RR is present, the above excess VPN routes advertisement churn can also occurs. For example, if PE3 receives excessive VPN routes for VPN1 VRF(there may be several reasons for this to occur, for example, multiple CEs connect to PE3 advertising routes simultaneously causing a wave of routes, redistribution from VRF to VRF, or from GRT to VRF on PE3 etc.), it will advertise such excessive VPN routes to RR and then to PE1. The BGP session between RR and PE3, and the BGP session between RR and PE1 can't prevent this to occur.

When PE1 in this figure receives such excessive VPN routes, it can only process them, among the other normal BGP updates. This can certainly influence process of VPN routes for other normal services, the consequences on the receiving PE1 may be the one or more of the followings:

a) PE1 can't process a given number of routes in time period X leading to dropping of routes

b) Delayed processing that may result in an incomplete number of inputs to the BGP Best Path decision.

c) L3VPN customers experiencing an incorrect VPN specification for some time period Y.

d) The convergence of control plane processing impacts the traffic forwarding

PE and RR should all have some capabilities to control the specified excessive VPN routes to be advertised from its upstream BGP peer.

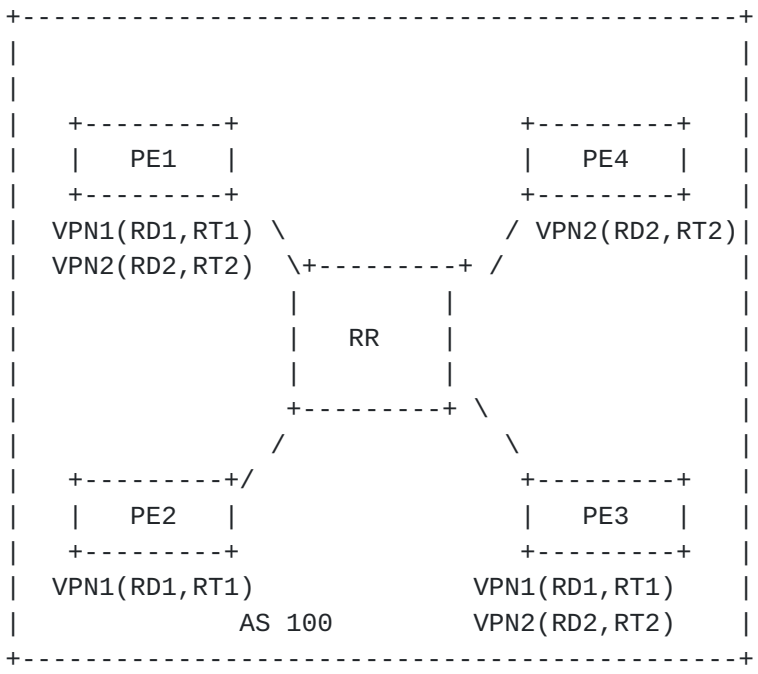

Figure 3: Intra-AS VPN RR deployment scenario

7. VPN Routes Shared on one PE

The scenarios described above are mainly in device level, that is to say, if the receiving PE has some mechanism to control the excess VPN routes advertisement from its BGP neighbor, the failure churn effect can be controlled then. But there are also situations that the granular control should be took place within the receiving PE itself.

Figure 4 below describes such scenario. There are four VRFs on PE, and three of them import the same VPN routes that carry route target RT3. Such deployment can occur in the inter-VRF communication scenario. If the threshold of VPN route-limit for these VRFs is set different, for example, are max-vpn-routes-vrf1, max-vpn-routes-vrf2, max-vpn-routes-vrf3, max-vpn-routes-vrf4 respectively, and these values have the following order, as max-vpn-routes-vrf1<max-vpn-routes-vrf2< max-vpn-routes-vrf3<max-vpn-routes-vrf4.

If the VPN routes that associates with RT3 is overwhelming, the VRF1 will reach its maximum VPN threshold first. At such stage, the PE device can't send the control message to its BGP neighbor on behalf

of all the VRFs on it, because other VRFs have still the desire to receive such VPN routes and have the capacities to store them.

In such situation, the PE device should have some mechanisms to control the distribution of global VPN routes to its individual VRF table. Only when all of VRFs on it don't want some VPN routes, then the PE device can send the VPN routes filter control message to its BGP neighbor (RR in this example).

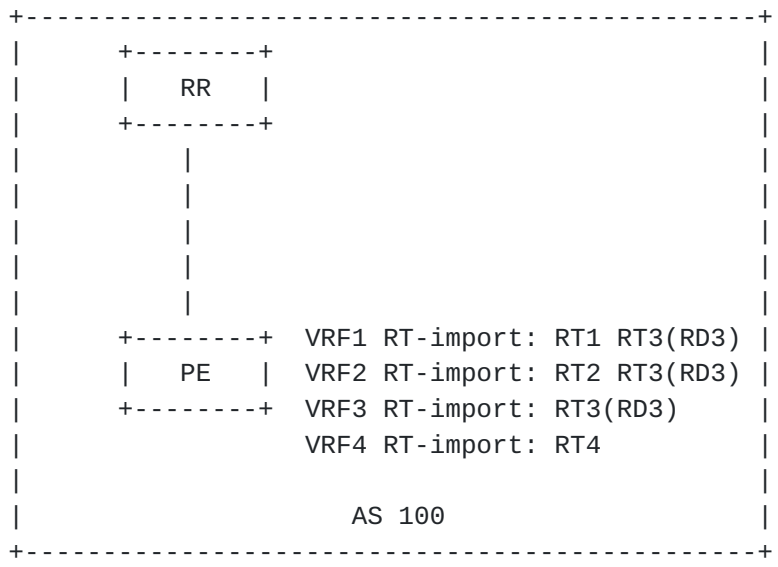

Figure 4: The scenario of several VRFs in a PE import VPN routes carries the same RT

.

8. Requirements for the solutions

Based on the above scenarios description, the potential solutions should meet the following requirements:

a) The control message for the specified VPN routes should be triggered automatically upon the excessive VPN routes reach its limit.

b) The control message should be sent only out the device when all the VRFs on it can't or don't want to process it, or the process of such excessive routes has exceed its own capability.

b) For RR and ASBR devices, such control message should be only flooded to its upstream BGP neighbor when all its downstream BGP peers can't or don't want to process it, or the process of such excessive routes has exceed its own capability.

c) The trigger and removal of such control message should avoid the possible flapping of excessive VPN routes advertisement.

d) The excessive VPN routes should be identified in fine-granular manner.

e) Control of the excessive VPN routes should not affect the existing VPN services on the affected PE.

9. Security Considerations

TBD.

10. IANA Considerations

This document requires no IANA considerations.

11. Acknowledgement

Thanks Robert Raszuk, Jim Uttaro, Jakob Heitz for their valuable comments and discussions of scenarios described in 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>.

[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>.

[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>.

Authors' Addresses

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

Email: wangaj3@chinatelecom.cn

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

Email: wangw36@chinatelecom.cn

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

Haibo Wang
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing, Beijing 100095
China

Email: rainsword.wang@huawei.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