Interdomain Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 20, 2014

S. Litkowski
Orange Business Service
J. Haas
Juniper Networks
June 18, 2014

Inter Domain considerations for Constrained Route distribution
draft-litkowski-idr-rtc-interas-00

Abstract

[RFC4684] defines Multi-Protocol BGP (MP-BGP) procedures that allow BGP speakers to exchange Route Target reachability information in order to limit the propagation of Virtual Private Networks (VPN) Network Layer Reachability Information (NLRI).

[RFC4684] addresses both intra domain and inter domain distributions. Based on operational deployments, the current distribution model defined in [RFC4684] may cause some issue in specific scenarios.

This document refines the route distribution rules for inter domain NLRIs in order to address these specific scenarios.

Requirements Language

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].

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 http://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on December 20, 2014.

Table of Contents

1. Problem statement

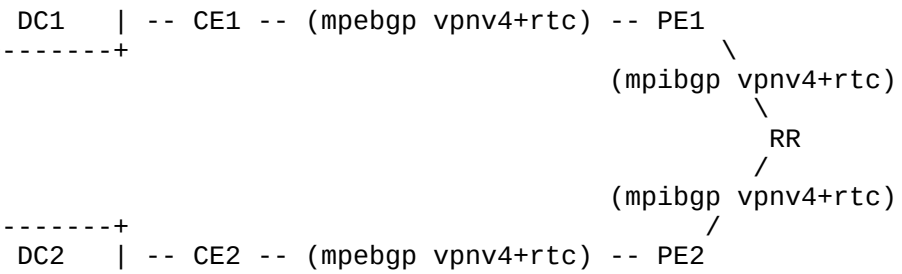

Figure 1

The figure above describes a typical service provider scenario where datacenters are connected through MPLS VPN interas option B with the Service Provider network. Route Target Constraint (RTC) is deployed on MPeBGP sessions as well as internally in the service provider network to ensure optimal distribution of VPN routes (required for scaling reason). In this scenario, both Datacenters are using the same AS number, generally a private ASN (65000) like a typical PE-CE connection. As we expect DCs to communicate between each other, some

features like "as-override" are deployed on PEs to overcome ASPATH loop issue.

[RFC4684] Section 3.1 and 3.2 describes propagation of Route Target NLRI between ASes and inside an AS and distinguish two types of NLRIs :

- Locally originated NLRI where origin-as field of the NLRI is equal to the local AS number.

- External NLRI where origin-as field of the NLRI is different from the local AS number.

Regarding External NLRI, the idea of Section 3.1 and 3.2 is to establish the route distribution tree over the shortest path considering that BGP routing is internally consistent for a given AS.

Extract from [RFC4684] Section 3.2 :

"As indicated above, the inter-AS VPN route distribution graph, for a given route-target, is constructed by creating a directed arc on the inverse direction of received Route Target membership UPDATEs containing an NLRI of the form {origin-as#, route-target}.

Inside the BGP topology of a given autonomous-system, as far as external RT membership information is concerned (route-targets where the as# is not the local as), it is easy to see that standard BGP route selection and advertisement rules [4] will allow a transit AS to create the necessary flooding state."

In the Figure 1, CE1 and CE2 are advertising the RT 1:1 respectively to PE1 and PE2, the generated NLRI would be 65000:2:1:1/96. According to procedures defined in [RFC4684] Section 3.2, both PEs are using the standard BGP route selection and advertising rules. So both PEs are advertising their path for 65000:2:1:1/96 to the route-reflector. The route-reflector would also use the standard BGP route selection to create the RT flooding state. Considering that path from PE1 is the best one, a flooding tree branch for RT 1:1 is created only towards PE1.

Due to this behavior, VPN routes from DC1 would never to send to DC2 because PE2 is not part of the flooding tree and as DC1 and DC2 are disjoint, even if they are using the same ASN, there is no communication possible between them.

The same issue may appear if two MPeBGP sites using the same ASN are connected on the same PE like in figure 2.

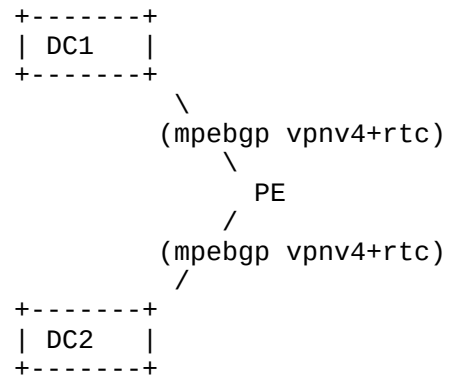

Figure 2

2. Proposal

This document proposes to modify the following procedures defined in [RFC4684] :

1. [RFC4684] Section 3.1 :

"Using RT membership information that includes both route-target and originator AS number allows BGP speakers to use standard path selection rules concerning as-path length (and other policy mechanisms) to prune duplicate paths in the RT membership information flooding graph, while maintaining the information required to reach all autonomous systems advertising the Route Target."

2. [RFC4684] Section 3.2 :

"As indicated above, the inter-AS VPN route distribution graph, for a given route-target, is constructed by creating a directed arc on the inverse direction of received Route Target membership UPDATEs containing an NLRI of the form {origin-as#, route-target}.

Inside the BGP topology of a given autonomous-system, as far as external RT membership information is concerned (route-targets where the as# is not the local as), it is easy to see that standard BGP route selection and advertisement rules [4] will allow a transit AS to create the necessary flooding state."

In order to support our scenario, path pruning may be disabled by configuration for a given origin AS (different from the local AS). Implementations may also permit path pruning to be disabled for private AS numbers by default, but must make provision for it to be selectively enabled if such a feature is present.

This modification in establishing route distribution tree may create unnecessary flooding states in the situations where a real AS is multihomed to a service provider network (as displayed in Figure 3).

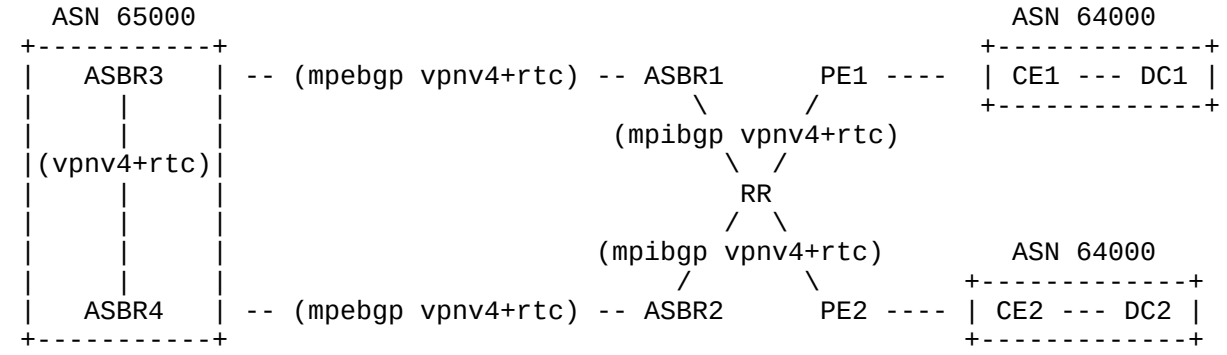

Figure 3

In the figure above, disabling pruning is required for AS64000 but it may be interesting to keep it enabled for AS65000. Implementations may require support for such granularity as proposed previously.

3. Security considerations

This document does not introduce any new security issue compared to [RFC4684].

4. Acknowledgements

5. IANA Considerations

There is no IANA consideration.

6. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[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, November 2006.

Authors' Addresses

Stephane Litkowski
Orange Business Service

Email: stephane.litkowski@orange.com

Jeff Haas
Juniper Networks

Email: jhaas@juniper.net