

IDR
Internet-Draft
Updates: [4684](#) (if approved)
Intended status: Standards Track
Expires: September 14, 2017

D. Duan
J. Heitz
Cisco
K. Patel
Arrcus
J. Hass
Juniper Networks
March 13, 2017

Persistent Route Oscillation in BGP Constrained Route Distribution
draft-idr-bgp-rt-oscillation-01

Abstract

[RFC4684](#) defines Multi-Protocol BGP (MP-BGP) procedures that allow BGP speakers to exchange Route Target reachability information (RT-Constrain) to restrict the propagation of Virtual Private Network (VPN) routes. In network scenarios where hierarchical route reflection (RR) is used, the existing RT-Constrain mechanism may result in persistent route oscillations within RRs. This document describes the problem and proposes a solution.

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 September 14, 2017.

Copyright Notice

Copyright (c) 2017 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 (<http://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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1.	Introduction	2
2.	Requirements Language	3
3.	Problem Statement - Persistent Route Oscillations	3
4.	Solution	5
5.	IANA Considerations	5
6.	Security Considerations	5
7.	Acknowledgements	5
8.	Normative References	5
	Authors' Addresses	6

[1.](#) Introduction

[RFC4684] defines Multi-Protocol BGP (MP-BGP) procedures that allow BGP speakers to exchange Route Target reachability information to restrict the propagation of Virtual Private Network (VPN) routes.

[RFC4684] [section 3.2](#) defines a route advertisement rule for Route Target membership information. When advertising an RT membership NLRI to a non-client peer, if the best path as selected by the path selection procedure described in [Section 9.1 of \[RFC4271\]](#) is a route received from a non-client peer, and if there is an alternative path to the same destination from a client peer, then the attributes of

the client path are advertised to the peer. [RFC4684] does not clarify which path to choose in case there are multiple client paths to the same destination.

In network scenarios where hierarchical route reflection (RR) is used, and multiple such client paths exist, persistent route oscillations might be formed based on which client path attributes are advertised to the non-client peers.

2. 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].

3. Problem Statement - Persistent Route Oscillations

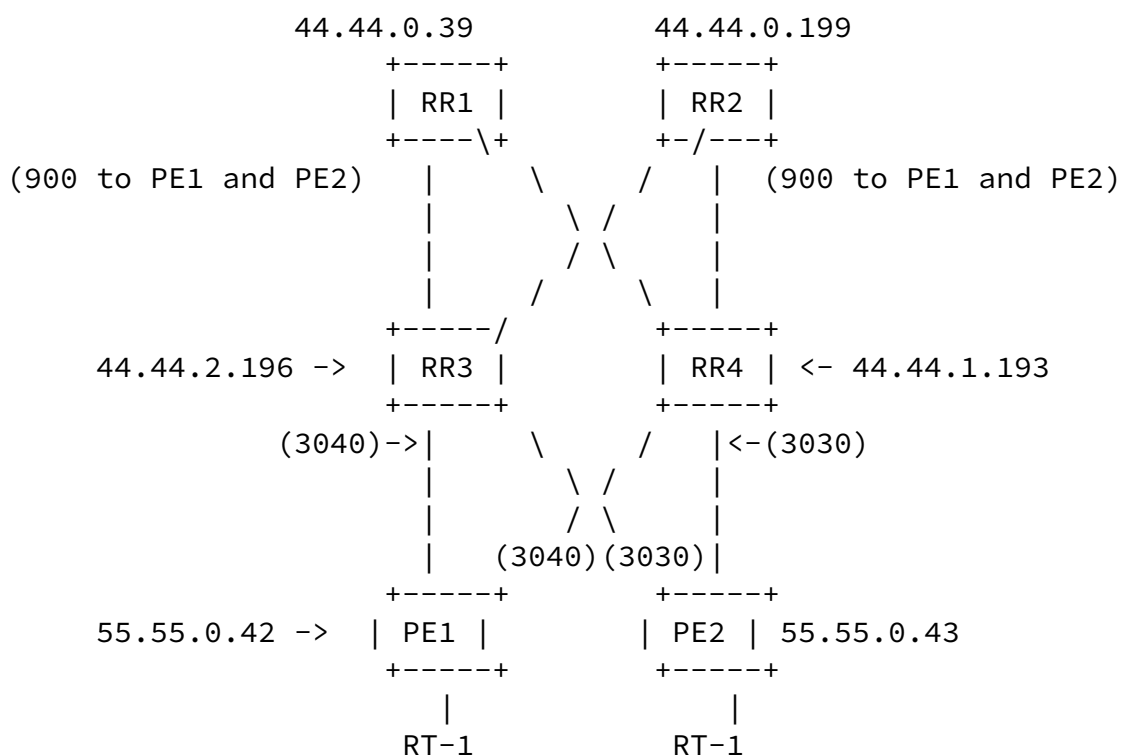


Figure 1. RT-Constrain with Hierarchical Route-reflector

In Figure 1, Hierarchical RRs are deployed. RR3 and RR4 are first level Router Reflectors and RR1 and RR2 are the second level Route Reflectors. Each RR is using its own router-id as its cluster-id. PE1 and PE2 are Route Reflector clients of RR3 and RR4 while RR3 and RR4 are Route Reflector clients of RR1 and RR2. Both PE1 and PE2 are advertising the route-target information RT-1 to the first level Router Reflectors RR3 and RR4. RR3 and RR4 are also advertising route-target information to the second level Router Reflectors RR1 and RR2. The numbers in the parentheses are nexthop metrics.

At step #1, RR3 has two paths for RT-1: one from PE1 and the other from PE2. The path from PE1 has next hop metric 3040 and the path from PE2 has next hop metric 3030. RR3 and RR4 select the path from PE2 as the best path (lower metric). RR3 and RR4 advertise their best paths for RT-1 to the second level router reflectors RR1 and RR2.

At step #2, RR1 has two paths for RT-1: one path from RR3 and the other path from RR4. The next hop metric to reach PE1 and PE2 are both 900. On RR1 and RR2, if both paths have the same ORIGINATOR_ID, then the lower peer address will be used to select the best path. RR1 selects the path from RR4 as best path because it has a lower peer address than RR3. RR1 advertises RT-1 back to RR3.

When announcing RT-1 to its client (RR3), RR1 will set the ORIGINATOR_ID to itself according to [\[RFC4684\] section 3.2.i](#).

At step #3, RR3 has four paths: the first from PE1, the second from PE2, the third from RR1 and a fourth from RR2. For the purposes of this discussion, the path from RR2 is equivalent to that from RR1. The result is the same if either is chosen. RR3 selects the non-client path from RR1 to PE2 as best path because the ORIGINATOR_ID is lower than that of the paths from PE1 and PE2. Since there are client paths available to reach RT-1, RR3 advertises the path attribute of a client path to RR1 according to [\[RFC4684\] section 3.2.ii](#). RR3 could choose either the path from PE1 or PE2. RR3 chooses the path attribute of RT-1 from PE1 at random.

At step #4, RR1 receives the updates and recalculates the best path.

RR1 has a path from RR4 with ORIGINATOR_ID set to PE2's router-id and a path from RR3 with ORIGINATOR_ID set to PE1's router-id. RR1 selects the path from RR3 as best path because of lower ORIGINATOR_ID. RR1 sets the ORIGINATOR_ID to its own router-id and sends it back to RR3.

At step #5, RR3 receives the updates from RR1 and drops the updates since its own cluster-id is in the cluster list. Now RR3's routing state goes back to that at step #1 with 2 paths from its clients and the whole cycle starts again.

The same thing happens on RR4 as on RR3 and the same thing happens on RR2 as on RR1.

These iterations results in a persistent route oscillation for RT-1 prefix of RT-Constrain address-family on RR1, RR2, RR3 and RR4.

[4.](#) Solution

The solution is for the Route Reflector always to prefer the client paths when selecting a best path. This preference MUST be expressed before step f) of the BGP Decision Tie Breaking rules in [Section 9.1.2.2 of \[RFC4271\]](#). It MAY be expressed at a higher step. So at Step #3 on RR3, the best path is still the path from PE2. The oscillation terminates with PE2's path.

Note that the scenario can not happen if RR1 and RR2 are in the same cluster. So at step #3, RR3 only has two client paths. The update from the top level Route Reflector will be dropped because of the cluster id check. The oscillation never happens with such a topology.

[5.](#) IANA Considerations

This draft makes no request of IANA.

[6.](#) Security Considerations

This extension to BGP does not change the underlying security issues

inherent in the existing [[RFC4271](#)] and [[RFC4684](#)].

7. Acknowledgements

The authors would like to thank Shyam Sethuram, Nitin Kumar, Sameer Gulrajani, Mohammed Mirza and Mike Dubrovskiy.

8. 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, <<http://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, <<http://www.rfc-editor.org/info/rfc4271>>.
- [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, <<http://www.rfc-editor.org/info/rfc4684>>.

Duan, et al.

Expires September 14, 2017

[Page 5]

Internet-Draft Route Oscillation in BGP RT-Constraint

March 2017

Authors' Addresses

Dongling Duan
Cisco
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: duan@cisco.com

Jakob Heitz
Cisco
170 West Tasman Drive
San Jose, CA 95054

USA

Email: jheitz@cisco.com

Keyur Patel
Arrcus, Inc

Email: keyur@arrcus.com

Jeffrey Hass
Juniper Networks
1194 N. Mathida Ave
Sunnyvale, CA 94089
USA

Email: jhaas@juniper.net