

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 26, 2013

P. Mohapatra
R. Fernando
Cisco Systems
January 22, 2013

BGP Link Bandwidth Extended Community
draft-ietf-idr-link-bandwidth-06.txt

Abstract

This document describes an application of BGP extended communities that allows a router to perform unequal cost load balancing.

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 July 26, 2013.

Copyright Notice

Copyright (c) 2013 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 [3](#)
 [1.1.](#) Requirements Language [3](#)
[2.](#) Link Bandwidth Extended Community [3](#)
[3.](#) Deployment Considerations [3](#)
[4.](#) Acknowledgments [4](#)
[5.](#) IANA Considerations [5](#)
[6.](#) Security Considerations [5](#)
[7.](#) Normative References [5](#)
Authors' Addresses [5](#)

Internet-Draft

Link Bandwidth

January 2013

1. Introduction

When a BGP speaker receives multiple paths from its internal peers, it could select more than one path to send traffic to. In doing so, it might be useful to provide the speaker with information that would help it distribute the traffic based on the bandwidth of the external (DMZ) link. This document suggests that the external link bandwidth be carried in the network using a new extended community [[RFC4360](#)] - the link bandwidth extended community.

1.1. 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 [RFC 2119](#) [[RFC2119](#)].

2. Link Bandwidth Extended Community

When a BGP speaker receives a route from an external neighbor and advertises this route (via IBGP) to internal neighbors, as part of this advertisement the router may carry the cost to reach the external neighbor. The cost can be either configured per neighbor or derived from the bandwidth of the link that connects the router to a directly connected external neighbor. This value is carried in the Link Bandwidth Extended Community. No more than one link bandwidth extended community SHALL be attached to a route. Additionally, if a route is received with link bandwidth extended community and the BGP speaker sets itself as next-hop while announcing that route to other peers, the link bandwidth extended community should be removed.

The extended community is optional non-transitive. The value of the high-order octet of the extended Type Field is 0x40. The value of the low-order octet of the extended type field for this community is 0x04. The value of the Global Administrator subfield in the Value Field SHOULD represent the Autonomous System of the router that

attaches the Link Bandwidth Community. If four octet AS numbering scheme is used [[RFC6793](#)], AS_TRANS should be used in the Global Administrator subfield. The bandwidth of the link is expressed as 4 octets in IEEE floating point format, units being bytes (not bits!) per second. It is carried in the Local Administrator subfield of the Value Field.

[3.](#) Deployment Considerations

The usage of this community is restricted to the cases where BGP multipath can be safely deployed. If the path between the load

sharing router and the exit point is not tunneled, then the IGP distance between the load balancing router and the exit points should be the same.

If the path between the load sharing router and the exit point is tunneled, then the choice to use this community is a purely local matter to the load sharing router.

In the context of BGP/MPLS VPNs [[RFC4364](#)], link bandwidth community could be used to support inbound load balancing for multihomed sites, as follows. Consider a site that is connected to PE1 and PE2. Both PE1 and PE2 would advertise VPN-IP routes associated with the destinations within the site. One way to enable other PEs to receive all these routes is to require the RD of the routes advertised by PE1 to be different from the RD of the routes advertised by PE2. The VPN-IP routes advertised by PE1 should carry the link bandwidth community; likewise for the VPN-IP routes advertised by PE2. The bandwidth value carried in the community could be locally determined by PE1 and PE2. Alternatively CEs of the site, when advertising IP routes to PE1 and PE2, could add the link bandwidth community to these advertisements, in which case PE1 and PE2, when originating VPN-IP routes, would use the bandwidth value from the IP routes they received from the CEs to construct the link bandwidth community carried by these VPN-IP routes.

An ingress PE, when sending traffic to destinations within the site, can use the bandwidth value carried in the community of the routes advertised by PE1 and PE2 to perform load sharing, where some of the traffic would go via PE1, while other traffic would go via PE2.

If there are multiple paths to reach a destination and if only some of them have link bandwidth community, the load sharing router should not perform unequal cost load balancing based on link bandwidths.

[4.](#) Acknowledgments

The authors would like to thank Yakov Rekhter, Srihari Sangli and Dan Tappan for proposing unequal cost load balancing as one possible application of the extended community attribute.

The authors would like to thank Bruno Decraene, Robert Raszuk, Joel Halpern, Aleksi Suhonen, Randy Bush, and John Scudder for their comments and contributions.

[5.](#) IANA Considerations

This document defines a specific application of the two-octet AS specific extended community. IANA is requested to assign a sub-type value of 0x04 for the link bandwidth extended community.

Name	Value
----	-----
non-transitive Link Bandwidth Ext. Community	0x4004

[6.](#) Security Considerations

There are no additional security risks introduced by this design.

[7.](#) Normative References

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

- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", [RFC 4360](#), February 2006.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", [RFC 4364](#), February 2006.
- [RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", [RFC 6793](#), December 2012.

Authors' Addresses

Pradosh Mohapatra
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Phone:
Email: pmohapat@cisco.com

Mohapatra & Fernando

Expires July 26, 2013

[Page 5]

Internet-Draft

Link Bandwidth

January 2013

Rex Fernando
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Phone:
Email: rex@cisco.com

