IDR
Internet-Draft
Updates: 4271, 4360, 7153 (if approved)
Intended status: Standards Track
Expires: September 13, 2017

Z. Li
China Mobile
J. Dong
Huawei Technologies

March 12, 2017

Carry congestion status in BGP extended community draft-li-idr-congestion-status-extended-community-04

Abstract

A new extended community is introduced in this document to carry the link congestion status, especially for the exit link of one AS. It is called congestion status extended community. This extended community can be used by the BGP routers or the SDN controllers to steer the Internet-access traffic among the exit links.

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 September 13, 2017.

Copyright Notice

Table of Contents

1. Problem Statement

Typically the architecture of a large scale ISP's network is multi-layered, as illustrated in Figure 1. The national backbone network has its own AS, and each of the province or state network has a specific AS. Backbone network connects all the province or state networks together and has several exit links to access the Internet. The province or state networks usually have direct exit links to the Internet. The total bandwidth of the backbone exit links is usually much bigger than that of the direct exit links in the province or state networks. Thus, the Internet-access traffic is mainly transported through the backbone exit links by deploying route policies on the ASBR routers in the province or state networks. The ASBR routers in the province or state networks, for example, prefer the routes learned from the backbone by setting higher local preference for those routes. However, when the backbone exit links are congested due to traffic increasing or delay of the capacity expansion, the ASBR routers in the province or state networks do not know this, and still deliver Internet-access traffic to the backbone. The customer experience deteriorates, the operator, in turn, will receive more and more complaints for its bad network performance. Then, the operator has to steer some Internet-access traffic to the direct exit links in the province or state networks by deploying route policy on the ASBR routers. This kind of policy should be removed when the capacity expansion of the backbone exit links is done. The ASBR routers do not know this again.

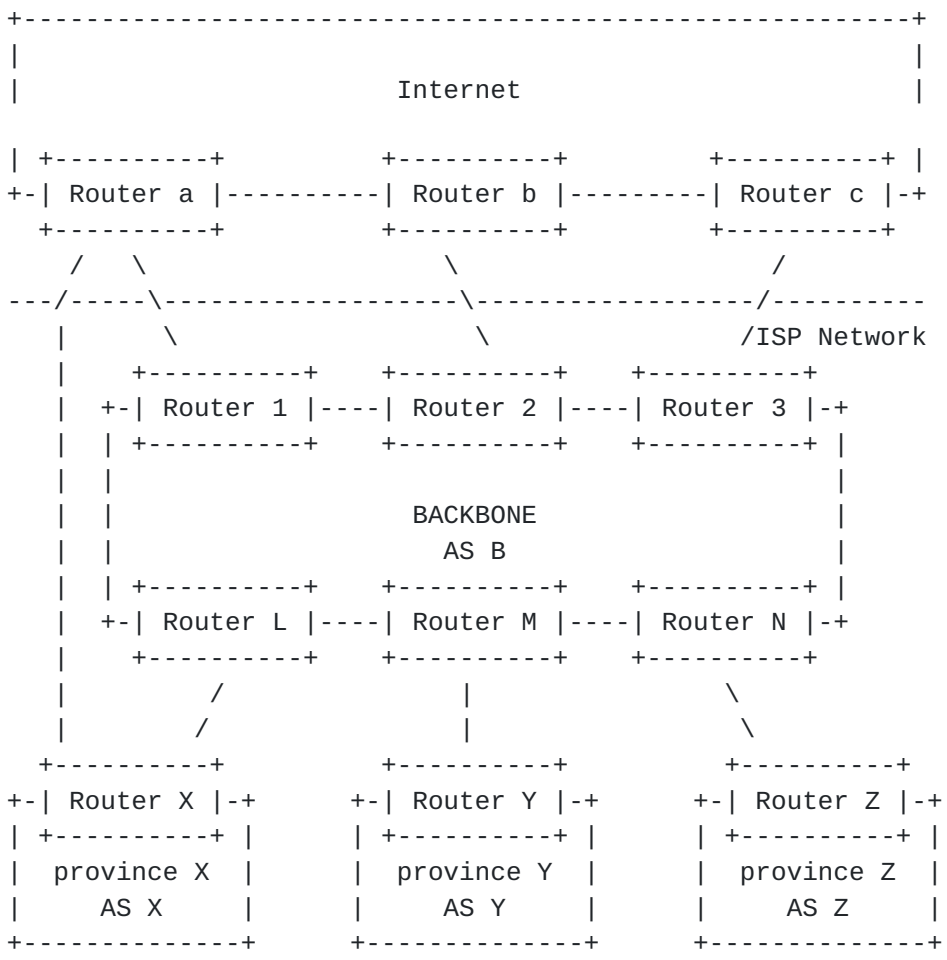

Figure 1: Typical architecture of a large scale ISP's network

In [constrained-multiple-path], authors from France Telecom also specified the requirement to know the congestion status of a link.

2. Solution Overview

This document introduces a new extended community [RFC4360] to deliver the congestion status of the exit link to other BGP speakers. The BGP receiver can then use this extended community to deploy route policy, thus steer Internet-access traffic according to the congestion status of the exit link. Router X in the above figure, for example, can steer some Internet-access traffic to the direct exit link when it knows the backbone exit link is congested. On the other hand, when Router X knows the exit link of AS B is not congested anymore, it can steer all the Internet-access traffic back to the backbone network. The introduced extended community is called congestion status extended community.

Congestion status extended community is good not only to the ASBRs in other AS, but also to the BGP peers within one AS. For instance, Router M in backbone AS chooses Router 2 to transport the Internet-access traffic by default, because the IGP cost from Router M to Router 2 is smallest. When Router M receives congestion status extended communities from Router 1,2,3, which indicate the utilization of the exit link of Router 1,2,3 is 90%, 70%, and 50% respectively, it can choose Router 3 to transport some Internet-access traffic using route policy.

In a network deployed SDN (Software Defined Network) controller, congestion status extended community can be used by the controller to steer the Internet access traffic among all the exit links from the perspective of the whole network.

For the network with Route Reflectors (RRs) [RFC4456], RRs by default only advertise the best route for a specific prefix to their clients. Thus RR clients has no opportunity to compare the congestion status among all the exit links. In this situation, to allow RR clients learning all the routes for a specific prefix from all the exit links, RRs are RECOMMENDED to enable add-path functionality [RFC7911].

3. Congestion Status Extended Community

As described in [RFC4360], the extended community attribute is an 8-octet value with the first one or two octets to indicate the type of this attribute. Since congestion status extended community needs to be delivered from on AS to other ASes, and used by the BGP speakers both in other ASes and within the same AS as the sender, it MUST be a transitive extended community, i.e. the T bit in the first octet MUST be zero.

We only define the congestion status extended community for four-octet AS number [RFC6793], since all the BGP speakers can handle four-octet AS number now and the two-octet AS numbers can be mapped to four-octet AS numbers by setting the two high-order octets of the four-octet field to zero, as per [RFC6793].

Congestion status extended community is a sub-type allocated from Transitive Four-Octet AS-Specific Extended Community Sub-Types defined in section 5.2.4 of [RFC7153]. Its format is as Figure 2.

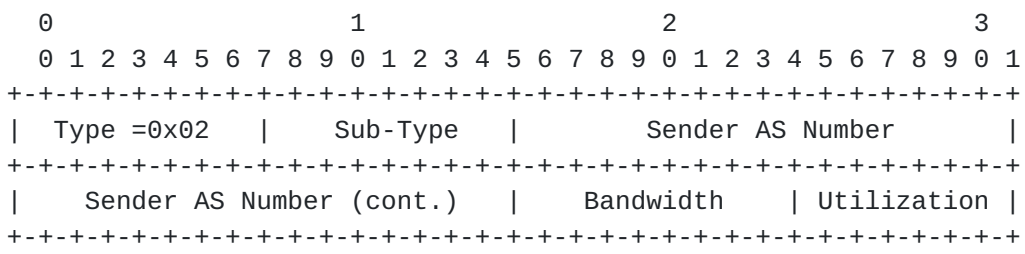

Figure 2: Congestion status extended community

The "Type" field MUST be 0x02, which indicate this is a Transitive Four-Octet AS-Specific Extended Community.

The "Sub-Type" field is used to indicate this is a Congestion Status Extended Community. Its value is to be assigned by IANA.

The "Sender AS Number" field is 4 octets. Its value is the AS number of the BGP speaker who generates this congestion status extended community. If the generator has 2-octct AS number, it MUST encode its AS number in the last (low order) two bytes and set the first (high order) two bytes to zero, as per [RFC6793].

The "Bandwidth" field is 1 octet. Its value is the bandwidth of the exit link in unit of 10 gbps (gigabits per second). The link with bandwidth less than 10 gbps is not suitable to use this feature.

The "Utilization" field is 1 octet. Its value is the utilization of the exit link in unit of percent. We can use the "Utilization" field together with the "Bandwidth" field to calculate the traffic load that we can further steer to this exit link.

4. Application Considerations

To avoid route oscillation, the exit router SHOULD set a threshold. When the utilization change reaches the threshold, the exit router SHOULD generate a BGP update message with congestion status extended community. Implementations SHOULD further reduce the BGP update messages trigered by link utilization change using the method similar to BGP Route Flap Damping [RFC2439]. When link utilization change by small amounts that fall under thresholds that would cause the announcement of BGP update message, implementations SHOULD suppress the announcement and set the penalty value accordingly.

To avoid traffic oscillation, i.e. more traffic than expected is attracted to the low utilized link, and some traffic has to be steered back to other links, route policy can be set at the exit router. Congestion status extended community is only conveyed for

some specific routes or only for some specific BGP peers. Congestion status extended community can also be used in a SDN network. The SDN controller uses the exit link utilization information to steer the Internet access traffic among all the exit links from the perspective of the whole network.

5. Security Considerations

This document only defines a new extended communities to carry the congestion status of the exit link. This new extended community does not directly introduce any new security issues. The same security considerations as for the BGP extended community [RFC4360] applies.

6. IANA Considerations

One sub-type is solicited to be assigned from Transitive Four-Octet AS-Specific Extended Community Sub-Types registry to indicate the Congestion Status Extended Community defined in this document.

7. Acknowledgments

Many thanks to Rudiger Volk, Susan Hares, John Scudder for their review and comments to improve this document.

8. References

8.1. 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>.

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

[RFC7153] Rosen, E. and Y. Rekhter, "IANA Registries for BGP Extended Communities", RFC 7153, DOI 10.17487/RFC7153, March 2014, <http://www.rfc-editor.org/info/rfc7153>.

8.2. Informative References

[constrained-multiple-path]
Boucadair, M. and C. Jacquenet, "Constrained Multiple BGP Paths", October 2010.

[RFC2439] Villamizar, C., Chandra, R., and R. Govindan, "BGP Route Flap Damping", RFC 2439, DOI 10.17487/RFC2439, November 1998, <http://www.rfc-editor.org/info/rfc2439>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <http://www.rfc-editor.org/info/rfc4456>.

[RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", RFC 6793, DOI 10.17487/RFC6793, December 2012, <http://www.rfc-editor.org/info/rfc6793>.

[RFC7911] Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", RFC 7911, DOI 10.17487/RFC7911, July 2016, <http://www.rfc-editor.org/info/rfc7911>.

Authors' Addresses

Zhenqiang Li
China Mobile
No.32 Xuanwumenxi Ave., Xicheng District
Beijing 100032
P.R. China

Email: li_zhenqiang@hotmail.com

Jie Dong
Huawei Technologies
Huawei Campus, No.156 Beiqing Rd.
Beijing 100095
P.R. China

Email: jie.dong@huawei.com