Stream: Internet Engineering Task Force (IETF)
RFC: 10005
Category: Standards Track
Published: June 2026
ISSN: 2070-1721
Authors:
P. Mohapatra, *Google LLC*
R. Das, Ed., *HPE*
S. Mohanty, Ed., *Zscaler*
S. Krier, *Cisco Systems*
R.J. Szarecki, *Google LLC*
A. Gattani, *Arista Networks*

RFC 10005
BGP Link Bandwidth Extended Community

Abstract

This document defines a BGP extended community, the Link Bandwidth Extended Community, which carries bandwidth information to enable weighted load-balancing in multipath scenarios. It specifies the format and processing rules for this extended community type.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc10005.

Copyright Notice

Table of Contents

1. Introduction

Load-balancing is a critical aspect of network design, enabling efficient utilization of available bandwidth and improving overall network performance. Traditional equal load-balancing does not account for the varying capacities of different paths. This document suggests that the bandwidth be carried in the network using one of two new extended communities [RFC4360]: the transitive and non-transitive Link Bandwidth Extended Community. The Link Bandwidth Extended Community carries the bandwidth information of a directly connected link or multi-hop/multipath next hop as advertised by a router. This mechanism facilitates maximizing utilization of network resources.

1.1. Requirements Language

The key words "**MUST**", "**MUST NOT**", "**REQUIRED**", "**SHALL**", "**SHALL NOT**", "**SHOULD**", "**SHOULD NOT**", "**RECOMMENDED**", "**NOT RECOMMENDED**", "**MAY**", and "**OPTIONAL**" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Link Bandwidth Extended Community

The Link Bandwidth Extended Community is defined as a BGP extended community that carries the bandwidth information of a router (represented by BGP next hop) that is connecting to a remote network. This community can be used to inform other routers about the available bandwidth through a given route.

The Link Bandwidth Extended Community can be either transitive or non-transitive. Therefore, the value of the high-order octet of the extended Type field can be 0x00 or 0x40, respectively. The value of the low-order octet of the extended Type field for this community is 0x04.

The Global Administrator sub-field in the Value field **SHOULD** be set to the Autonomous System Number (ASN) of the router attaching the Link Bandwidth Extended Community, but it **MAY** contain any 2-octet value. If the ASN cannot be represented in 2 octets, AS_TRANS [RFC6793] **SHOULD** be used in the Global Administrator sub-field. The encoding of the full 4-octet ASNs is not supported by the Link Bandwidth Extended Community. Such a capability, should the operational need for it arise, may be provided by a new BGP extension. The value in the Global Administrator sub-field does not affect the use or semantics of the Link Bandwidth Extended Community. This approach maintains consistency with 2-octet community registries and remains operationally familiar.

The bandwidth value is expressed as 4 octets in the floating point format of [IEEE.754-2019], with units being bytes (not bits!) per second. It is carried in the Local Administrator sub-field of the Value field.

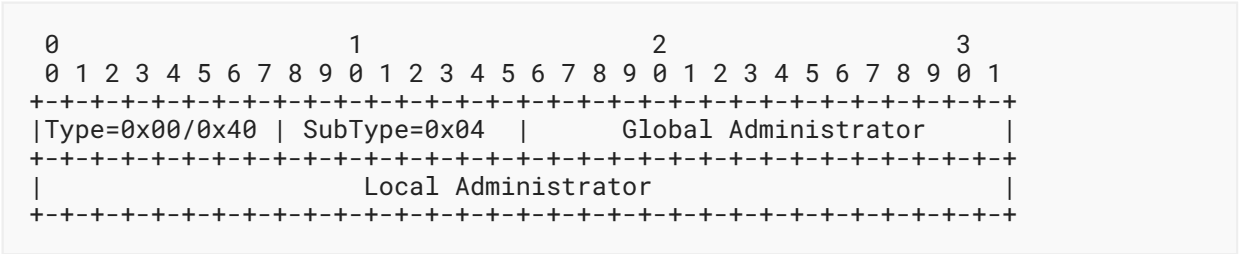

Figure 1: Link Bandwidth Extended Community

Type: A 1-octet field that **MUST** be set to 0x00 or 0x40 to indicate transitive/non-transitive.

SubType: A 1-octet field that **MUST** be set to 0x04 to indicate "link-bandwidth".

Global Administrator sub-field: A 2-octet field that represents an operator-assigned 2-octet value. For example, this can be a 16-bit ASN.

Local Administrator sub-field: Bandwidth value (bytes per second) encoded as 4 octets in the 32-bit floating point format of [IEEE.754-2019].

3. Protocol Procedures

The procedures cover both the transitive and non-transitive variants of the Link Bandwidth Extended Community so that implementations can handle both variants, ensuring that implementations can interoperate correctly across all deployments. Please refer to Section 5 and Appendix A for more details.

3.1. Sender (Originating Link Bandwidth Extended Community)

A BGP speaker that attaches a Link Bandwidth Extended Community **SHOULD** be able to advertise either a transitive or a non-transitive Link Bandwidth Extended Community. Implementations **SHOULD** provide the configuration to set the transitivity type of the Link Bandwidth Extended Community, as well as the Global Administrator and bandwidth values in the Local Administrator sub-field, by using local policy. Different implementations **MAY** use different default values for the transitivity type of the Link Bandwidth Extended Community. The provided configuration **SHOULD** allow operators to override the default transitivity value as needed. Likewise, implementations **SHOULD** expose their default value.

An implementation **MAY** advertise bandwidth value as zero. An operator may, for example, set the Link Bandwidth Extended Community to zero to indicate that the path should not attract traffic during maintenance. However, as per Section 3.2, it is up to the local policy of the receiver to decide how a link-bandwidth value of zero is handled.

Generally, a single Link Bandwidth Extended Community of the transitivity type desired in a deployment is attached to a route. However, during transition (refer Section 7 for details), a BGP speaker **MAY** attach one Link Bandwidth Extended Community per transitivity (transitive/non-transitive); the bandwidth value included in both communities **SHOULD** be the same.

A Link Bandwidth Extended Community **MAY** be attached or updated for a BGP route upon receipt during Adj-RIB-In processing. The Link Bandwidth Extended Community **MAY** be attached or updated for a BGP route's Adj-RIB-Out entry while being advertised to a neighboring BGP speaker. (Adj-RIB-In and Adj-RIB-Out are as defined in [RFC4271].)

Implementations **MAY** provide a configuration option to send non-transitive Link Bandwidth Extended Communities on external BGP sessions.

3.2. Receiver (Receiving Link Bandwidth Extended Community)

A BGP receiver that supports the Link Bandwidth Extended Community **MUST** support processing of both the transitive and non-transitive types. The receiver **MUST NOT** flap or treat the route as malformed based on the transitivity of the Link Bandwidth Extended Community and/or BGP session type (internal versus external).

Implementations **MAY** provide configuration to accept non-transitive Link Bandwidth Extended Communities from external BGP sessions.

A BGP update with an attached Link Bandwidth Extended Community with a bandwidth value of zero is valid. When all contributing paths have a non-zero value in the Link Bandwidth Extended Community, the bandwidth values of those paths (or their ratio) can be utilized as weights to enable weighted load-balancing. Details of weighted load-balancing are outside the scope of this document. Refer to [LINK-BW-USE-CASES], which describes some of the weighted load-balancing aspects. However, in the case where the paths have a mix of zero and non-zero values, or all zero values, the behavior is determined by local policy. For example, implementations may exclude the paths with a zero value from weighted load-balancing formation as long as at least one path with a non-zero value exists, or they may fall back to equal load-balancing. The bandwidth value, however, **SHOULD NOT** be used as an input to the BGP best path selection process.

Between transitive and non-transitive types of Link Bandwidth Extended Communities that have the same bandwidth value, the transitivity does not matter for the purpose of computing weighted load-balancing or programming to the Forwarding Information Base (FIB).

3.3. Re-Advertisement Procedures

This section describes the procedures to be followed when a BGP speaker receives a route with an attached Link Bandwidth Extended Community and subsequently re-advertises that route.

3.3.1. Re-Advertisement with Next Hop Change

When a BGP speaker re-advertises a route received with a Link Bandwidth Extended Community and sets the next hop to itself or to another address, it **MAY** do any one of the following as its default behavior: remove the Link Bandwidth Extended Community, re-advertise it unchanged,

or regenerate it with an updated value. Implementations **SHOULD** provide a local configuration method to alter their default behavior to the other options with per-session granularity. Likewise, implementations **SHOULD** expose their default value.

When regenerating the Link Bandwidth Extended Community, the same procedures as outlined in Section 3.1 apply. Please also refer to Section 3.4 for use in a BGP multipath environment.

3.3.2. Re-Advertisement with Next Hop Unchanged

A BGP speaker that receives a route with a Link Bandwidth Extended Community and re-advertises or reflects the same without changing its next hop **SHOULD NOT** change the Link Bandwidth Extended Community in any way.

3.4. Link Bandwidth Extended Community Arithmetic and BGP Multipath

In a BGP multipath environment, the bandwidth value that is sent or re-advertised **MAY** be calculated based on the Link Bandwidth Extended Community associated with each constituent path contributing to multipath in the Local Routing Information Base (Local-RIB). This topic is beyond the scope of this document. Refer to [LINK-BW-USE-CASES], which describes how this could be done in specific scenarios.

4. Error Handling

If a BGP speaker receives a route with more than one Link Bandwidth Extended Community and uses the route to compute weighted load-balancing, it **SHOULD** use the extended community with the lowest bandwidth value (including zero), ignoring the transitivity. Implementations **MAY** provide configuration to change the above preference.

A negative value in a Link Bandwidth Extended Community **SHOULD NOT** be attached or originated by any BGP speaker. If a BGP receiver encounters a Link Bandwidth Extended Community that contains a negative link-bandwidth value, the Link Bandwidth Extended Community **SHALL** be ignored.

Link Bandwidth Extended Communities with a zero value **MUST NOT** be considered malformed.

If any of the paths lack a valid Link Bandwidth Extended Community, equal load-balancing **SHOULD** be used unless overridden by local configuration.

5. IANA Considerations

In the "Transitive Two-Octet AS-Specific Extended Community Sub-Types" registry [IANA-TransExComm] (Type 0x00), IANA has updated Sub-Type 0x04 to:

Name: Link Bandwidth

In the "Non-Transitive Two-Octet AS-Specific Extended Community Sub-Types" registry [IANA-Non-TransExComm] (Type 0x40), IANA has updated Sub-Type 0x04 to:

Name: Link Bandwidth

Both updates reference this document.

6. Security Considerations

This extension to BGP has similar security implications as BGP extended communities [RFC4360].

The Link Bandwidth Extended Community conveys bandwidth and capacity information that may be sensitive. Exporting this community outside of an administrative domain can expose private network resource details. When propagating the routes with a Link Bandwidth Extended Community towards an untrusted network or outside of an administrative domain, it is recommended operators use policy to filter out this community.

7. Operational Considerations

7.1. Inconsistent Deployment

Prior deployments of the feature specified in this document have involved implementations that only understood one of the two extended community transitivity types. As a result, such implementations would ignore the other transitivity type that they don't understand. The procedures in this document govern how multiple transitivity types for bandwidth should operate.

Inconsistent behavior could occur when networks have deployed a mixture of implementations supporting this document's procedures for both transitivity types, or have deployed older implementations that only understand one transitivity type. A prime example is when a route received by a BGP speaker contains both a transitive and a non-transitive Link Bandwidth Extended Community, and that BGP speaker performs an operation that updates only one of the Link Bandwidth Extended Communities, so then the other community may have an inconsistent value. As a result, downstream BGP speakers that may receive such routes may perform inappropriate weighted load-balancing.

To mitigate such issues, when operators are aware that older implementations are present in their networks, they may wish to take actions to address such inconsistencies. One option would be to filter the unsupported transitivity type of the Link Bandwidth Extended Community at advertisement time on the older BGP speaker, if the implementation is capable of such filtering. Alternatively, a receiving BGP speaker, knowing that the sending speaker is incapable of doing such operations, could strip the Link Bandwidth Extended Community type that is unsupported by the sender.

Ideally, this operational consideration is short lived until all the routers in the network have been upgraded to implementations that consistently support the procedures in this document.

7.2. Bandwidth Value

How the bandwidth value is computed or determined is out of scope of this document. Refer to [LINK-BW-USE-CASES], which describes how this could be done in specific scenarios. It is recommended that implementations provide mechanisms to limit the churn caused by frequently changing bandwidth values, because rapid fluctuations could impact protocol stability and network operations.

8. References

8.1. Normative References

[IEEE.754-2019] IEEE, "IEEE Standard for Floating-Point Arithmetic", IEEE Std 754-2019, DOI 10.1109/IEEESTD.2019.8766229, 22 July 2019, <https://ieeexplore.ieee.org/document/8766229>.

[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>.

[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, <https://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, <https://www.rfc-editor.org/info/rfc4360>.

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

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

8.2. Informative References

[IANA-Non-TransExComm] IANA, "Non-Transitive Two-Octet AS-Specific Extended Community Sub-Types", <https://www.iana.org/assignments/bgp-extended-communities>.

[IANA-TransExComm] IANA, "Transitive Two-Octet AS-Specific Extended Community Sub-Types", <https://www.iana.org/assignments/bgp-extended-communities>.

[LINK-BW-USE-CASES] Litkowski, S., Ed., Mohanty, S R., Ed., Vayner, A., Gattani, A., Kini, A., Tantsura, J., and R. Das, "BGP link bandwidth extended community use cases", Work in Progress, Internet-Draft, draft-ietf-bess-ebgp-dmz-10, 8 April 2026, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-ebgp-dmz-10>.

Appendix A. Document History

The BGP Link Bandwidth Extended Community has evolved over several versions of the IETF draft. In the earlier versions up to draft-ietf-idr-link-bandwidth-08, only the non-transitive version of the Link Bandwidth Extended Community was supported. However, starting from draft-ietf-idr-link-bandwidth-09, both transitive and non-transitive versions of the Link Bandwidth Extended Community are supported.

A BGP speaker (sender or receiver) needs to be upgraded to support the procedures defined in this document to provide full interoperability for both transitive and non-transitive versions of the Link Bandwidth Extended Community. In order to simplify implementations, it is not a goal to provide interoperability by upgrading only the Route Reflector (RR).

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 Jeff Haas for all the discussions and providing text for operational considerations.

The authors would like to thank Bruno Decraene, Robert Raszuk, Joel Halpern, Aleksi Suhonen, Randy Bush, Stephane Litkowski, Mankamana Mishra, Moshiko Nayman, Keon Vafai, Ketan Talaulikar, Yingzhen Qu, Anoop Ghanwani, Dongjie (Jimmy), and John Scudder for their comments and contributions.

Contributors

Kaliraj Vairavakkalai
HPE
1133 Innovation Way
Sunnyvale, CA 94089
United States of America
Email: kaliraj.vairavakkalai@hpe.com

Natrajan Venkataraman
HPE
1133 Innovation Way
Sunnyvale, CA 94089
United States of America
Email: natrajan.venkataraman@hpe.com

Rex Fernando
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
United States of America
Email: rex@cisco.com

Authors' Addresses

Pradosh Mohapatra
Google LLC
Email: pradosh@gmail.com

Reshma Das (EDITOR)
HPE
1133 Innovation Way
Sunnyvale, CA 94089
United States of America
Email: reshma.das@hpe.com

Satya Mohanty (EDITOR)
Zscaler
120 Holger Way
San Jose, CA 95134
United States of America
Email: smohanty@zscaler.com

Serge Krier
Cisco Systems
Pegasus Parc, De Kleetlaan 6a
Belgium
Email: sekrier@cisco.com

Rafal Jan Szarecki
Google LLC
1160 N Mathilda Ave
Sunnyvale, CA 94089
United States of America
Email: rszarecki@gmail.com

Akshay Gattani
Arista Networks
5453 Great America Parkway
Santa Clara, CA 95054
United States of America
Email: akshay@arista.com