

BGP Class of Service Interconnection
draft-knoll-idr-cos-interconnect-14

Abstract

This document focuses on Class of Service Interconnection at inter-domain interconnection points. It specifies two new transitive attributes, which enable adjacent peers to signal Class of Service Capabilities and certain Class of Service admission control Parameters. The new "CoS Capability" is deliberately kept simple and denotes the general EF, AF Group BE and LE forwarding support across the advertising AS. The second "CoS Parameter Attribute" is of variable length and contains a more detailed description of available forwarding behaviours using the PHB id Code encoding. Each PHB id Code is associated with rate and size based traffic parameters, which will be applied in the ingress AS Border Router for admission control purposes to a given forwarding behaviour.

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](#)].

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 November 17, 2015.

Copyright Notice

Copyright (c) 2015 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.

Table of Contents

1.	Introduction	2
2.	Definition and Usage of the CoS Capability	3
2.1.	Extended Community Type	3
2.2.	Structure of the CoS Capability Attribute	4
2.3.	Usage of the CoS Capability Attribute	6
3.	Definition and Usage of the CoS Parameter Attribute	6
3.1.	Definition of the CoS Parameter Attribute	6
3.2.	Usage of the CoS Parameter Attribute	8
4.	Confidentiality Considerations	8
5.	IANA Considerations	9
6.	Security Considerations	9
7.	References	9
7.1.	Normative References	9
7.2.	Informative References	10
	Author's Address	10

[1.](#) Introduction

AS interconnection is currently based on best effort interconnection only. BGP-4 [[RFC4271](#)] is the de-facto interconnection protocol used to exchange reachability information. There is no standardized set of supported traffic classes, no standardized packet marking and no standardized forwarding behaviour, which cross-domain traffic could rely on. QoS policy decisions are taken by AS providers independently and in an uncoordinated fashion. However, many AS providers make use of the Differentiated Services Architecture [[RFC2475](#)] as AS internal QoS mechanism. Within this architecture, there are 64 codepoints and an unlimited number of Per Hop Behaviours (PHBs) available. Some PHBs have been defined in separate RFCs, which will be focused on in this document.

Knoll

Expires November 17, 2015

[Page 2]

A Basic Set of supported Classes, called "Basic CoS" is defined inhere, which consists of the primitive "Best Effort (BE)" PHB, the "Expedited Forwarding (EF)" PHB [[RFC3246](#)], the "Assured Forwarding (AF)" PHB Group [[RFC2597](#)] and the "Lower Effort" Per-Domain Behavior (PDB) [[RFC3662](#)]. AS providers, which can support this Basic CoS are asked to signal this capability to their interconnection partners by means of the new CoS Capability Extended Community defined in [Section 2](#) of this draft.

4 AF PHB classes have been defined so far, which will be grouped into the generally signalled "AF Group". That is, as long as the AS provider can support at least one out of the 4 AF classes in his externally supported CoS Set, this AS is regarded as AF capable.

A second transitive attribute is defined in [Section 3](#), which is used for parameter signalling about the applied access control within the ingress AS border router. The reason for this traffic limitation is the fact, that certain high quality forwarding behaviours can only be achieved, if the percentage of high priority traffic within the traffic mix lies below a certain threshold. This attribute informs the interconnection partner about the applied limitation, which can in turn be used to perform traffic shaping at the neighbouring AS' egress. The attribute allows this limitation signalling either associated to the NLRI within the same UPDATE message or with "global" scope to describe the generally applied ingress limitation.

Both attributes are likely to be used together, if ingress class limitation is used for the respective AS.

More detailed signalling of forwarding behaviour distinction and associated cross-layer marking can be achieved using the QoS Marking Attribute approach [[I-D.knoll-idr-qos-attribute](#)].

[2. Definition and Usage of the CoS Capability](#)

[2.1. Extended Community Type](#)

The new CoS Capability is encoded as a BGP Extended Community [[RFC4360](#)]. Extended Community Attributes are transitive optional BGP attributes with Type Code 16. An adoption to the simple BGP Community Attribute encoding [[RFC1997](#)] is not defined in this document. The actual encoding within the BGP Extended Community Attribute is as follows.

Knoll

Expires November 17, 2015

[Page 3]

The CoS Capability is transitive and of regular type which results in a 1 octet Type field followed by 7 octets for the CoS Capability structure. The Type is IANA-assignable (FCFS procedure) and marks the community as transitive across ASes. The type number has been assigned by IANA to 0xYY (0x00-0x3f).

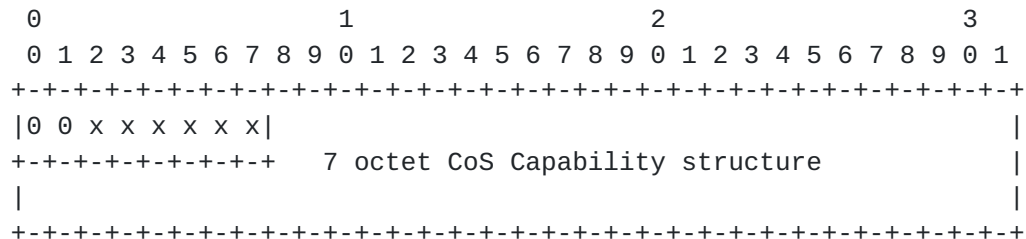


Figure 1

2.2. Structure of the CoS Capability Attribute

The CoS Capability structure is deliberately kept very simple and is defined as follows.

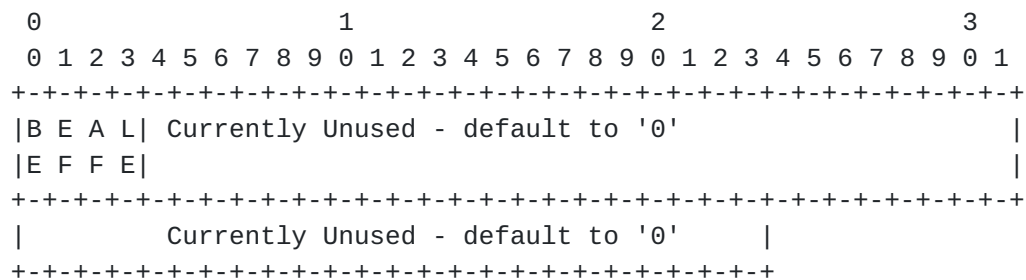
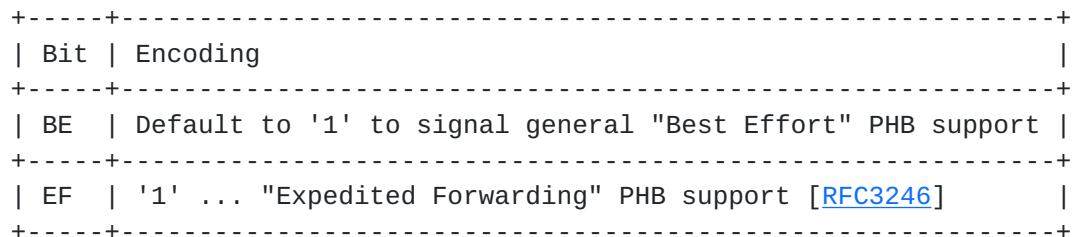


Figure 2

The Currently Unused bits default to '0' and MUST be ignored on reception.

Leading "BE, EF, AF and LE" encoding.

This encoding signals the BE, EF, AF Group and LE support of the respective AS.



Knoll

Expires November 17, 2015

[Page 4]

[illegible]

LE:

```

  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 0   0   1   0   0   0 | 0   0   0   0   0   0   0   0   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Figure 3

2.3. Usage of the CoS Capability Attribute

The CoS Capability is used as primitive means to signal the general availability of the set of "Basic CoS" PHBs in the advertising AS. This Extended Community is included within the attribute section of an BGP UPDATE message and is therefore associated to the NLRI information of the same message. Whether the Basic CoS is available and is therefore advertised can easily being judged on for all prefixes, which originate from the advertising AS.

All other reachability information MUST be signalled together with this CoS Capability if they were received together with such an Extended Community by neighbouring peers.

NLRI MUST NOT be marked as supporting "Basic CoS" by means of the CoS Capability, if it were not received together with such an attribute.

3. Definition and Usage of the CoS Parameter Attribute

3.1. Definition of the CoS Parameter Attribute

The CoS Parameter Attribute is an optional transitive BGP attribute.

The attribute contains one or more of the following:

```

  0                               1                               2                               3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| PHB id Code                               | Flags       | Reserved = '0'|
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                               ASN of sending AS                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Token Bucket Rate [r] (32-bit IEEE floating point number) |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Token Bucket Size [b] (32-bit IEEE floating point number) |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Peak Data Rate [p] (32-bit IEEE floating point number)   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Minimum Policed Unit [m] (32-bit integer)                 |

```

Knoll

Expires November 17, 2015

[Page 6]

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Maximum Packet Size [M] (32-bit integer)   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Figure 4

PHB ID:

This field specifies the targeted Per Hop Behaviour limitations and follows the defined encoding of [\[RFC3140\]](#) as listed in Figure 3.

Flags:

```

  0  1  2  3  4  5  6  7
+---+---+---+---+---+---+---+
|G |DR|0 |0 |0 |0 |0 |0 |
+---+---+---+---+---+---+---+

```

Only two flags are defined. The remaining bits default to '0' and MUST be ignored on reception.

The 'G' flag signals, whether the limitations have global scope on all incoming traffic ('1') or are associated to traffic that is destined to destinations within the NLRI of the UPDATE message ('0'). NLRI specific limitation will supersede globally signalled ones for traffic destined to those NLRI destinations.

The 'DR' flag signals the applied handling of non-confirming traffic. DR='0' signals strict dropping of excess traffic. DR='1' signals the performed remarking of excess traffic packets to Best Effort traffic marking.

ASN of sending AS:

Depending on the 2-octet or 4-octet AS peering type, the sending AS of the attribute MUST encode its AS number as right-aligned 32bit number.

Peak Data Rate, Token Bucket Rate, Token Bucket Size, Minimum Policed Unit and Maximum Packet Size:

The rates and sizes are given in 4 octet IEEE floating point format [\[IEEE\]](#) or 4 octet integer format, respectively. They are parameters to a token bucket ingress filter, which is applied to the packets belonging to the stated PHB id. The parameters follow the definition given in [\[RFC2210\]](#) and [\[RFC2215\]](#).

3.2. Usage of the CoS Parameter Attribute

The signalled parameters are used for PHB id Code based ingress limitation. Depending on which PHB id Codes a BGP peer signals in this attribute to its neighbour, it is said, that the respective PHB id Code is supported and will experience the defined limitations.

Those limitations can be applied to all incoming traffic of a specific PHB id Code (marked as 'G') or only for incoming traffic, that is destined for the NLRI of the given UPDATE message.

The resulting treatment for non-confirming traffic is signalled through the 'DR' flag.

To withdraw a previously signalled limitation, a CoS Parameter Attribute for the respective PHB id Code MUST be sent with a rate value [r] of zero. Using the 'G' flag, this can be withdrawn globally for all traffic of the given PHB id Code or withdrawn only for traffic destined to the prefixes given in the NLRI of the UPDATE. Previously signalled non-global (i.e. NLRI specific) limitations are also waived, if the same prefix is advertised without a CoS Parameter Attribute later on. In this case, the missing attribute is considered as the above described 'rate zero update' for those prefixes. Waived prefix specific limitations do not supersede global limitations for the respective PHB id Code. In turn, a withdrawal of a global limitation does also withdraw any possibly existing prefix specific ones for the respective PHB id Code.

All limitations have AS local scope for the advertising AS and the neighbouring AS might or might not adopt its sending behaviour to those advertised limitations.

Despite the transitive nature of the new attribute, its usage for ingress limitation is confined to neighbouring ASes. Processing of the conveyed parameters is only valid for peers, who are peering with the AS specified in the ASN field of the attribute.

The attribute SHOULD NOT be transitively relayed to non-adjacent interconnection partners.

4. Confidentiality Considerations

The disclosure of confidential AS intrinsic information by means of the signalled Basic CoS support is of low key security concern. The disclosure of information through CoS Parameter signalling is more detailed. However, all included parameters are exchanged with direct interconnection partners and are the free choice of each AS provider.

5. IANA Considerations

This document defines a new BGP Extended Community, which needs to be assigned a number by IANA within the Extended Community list. The new CoS Capability is a BGP Extended Community of regular type. It is IANA-assignable (FCFS procedure) and is transitive across ASes. A number assignment application within the numbering range of 0x00-0x3f is made to IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

This document defines a new BGP attribute. This attribute is optional and transitive.

6. Security Considerations

This extension to BGP does not change the underlying security issues inherent in the existing BGP version.

The signalled attributes are transitive with limited relay operation in the CoS Parameter Attribute case. AS peers, which use egress traffic shaper on the signalled limitations SHOULD exhaust all available BGP security features to make sure, that the signalled limitation is actually sent by the adjacent peer.

7. References

7.1. Normative References

- [IEEE] IEEE, "IEEE Standard for Binary Floating-Point Arithmetic", ISBN 1-5593-7653-8, 1985.
- [RFC1997] Chandrasekeran, R., Traina, P., and T. Li, "BGP Communities Attribute", [RFC 1997](#), August 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", [RFC 2210](#), September 1997.
- [RFC2215] Shenker, S. and J. Wroclawski, "General Characterization Parameters for Integrated Service Network Elements", [RFC 2215](#), September 1997.
- [RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", [RFC 2597](#), June 1999.

- [RFC3140] Black, D., Brim, S., Carpenter, B., and F. Le Faucheur, "Per Hop Behavior Identification Codes", [RFC 3140](#), June 2001.
- [RFC3246] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", [RFC 3246](#), March 2002.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), January 2006.
- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", [RFC 4360](#), February 2006.

7.2. Informative References

- [I-D.knoll-idr-qos-attribute]
Knoll, T., "BGP Extended Community for QoS Marking", [draft-knoll-idr-qos-attribute-15](#) (work in progress), January 2015.
- [RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", [RFC 2475](#), December 1998.
- [RFC3662] Bless, R., Nichols, K., and K. Wehrle, "A Lower Effort Per-Domain Behavior (PDB) for Differentiated Services", [RFC 3662](#), December 2003.

Author's Address

Thomas Martin Knoll
TU Chemnitz

Email: thomas.m.knoll@gmail.com

