Inter-Domain Routing Working Group	Th. Knoll	
Internet-Draft	Chemnitz University of Technology	
Intended status: Standards Track	June 08, 2008	
Expires: December 10, 2008		

the TOC link

BGP Extended Community Attribute for QoS Marking
draft-knoll-idr-qos-attribute-00

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.
Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.
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."
The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.
This Internet-Draft will expire on December 10, 2008.

Abstract

This document specifies a simple signalling mechanism for inter-domain QoS marking using a BGP Extended Community QoS Attribute. Class based packet forwarding for delay and loss critical services is currently performed in an individual AS internal manner. The new QoS marking attribute makes the QoS class setup within the IP prefix advertising AS known to all access and transit ASes. This enables individual (re-)marking and forwarding treatment adaptation to the original QoS class setup of the respective IP prefix. The attribute provides the means to signal QoS markings on different layers, which are linked together in QoS class sets. It provides inter-domain and cross-layer insight into the QoS class mapping of the source AS with minimal signalling traffic.

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 (Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.) [RFC2119].

Table of Contents

---

1.   Introduction   TOC

A new BGP Extended Community Attribute is defined in this document, which carries QoS marking information for different network layer technologies across ASes. This attribute is called "QoS Marking Attribute". This new attribute provides a mechanism for labelling priority class information carried in BGP-4 [RFC4271] (Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," January 2006.). It allows for the consistent exchange of priority class encoding values between BGP peers for physical, data link, network and transport layer QoS mechanisms. These labels can be used to control the distribution of this information, for the encoding and for treatment adjustments within the AS or for other applications.

Several QoS Marking Attributes MAY be included in a single BGP UPDATE
message. Each QoS Marking Attribute is encoded as 8-octet tuple, as
follows.

---

2. Definition of the QoS Marking Attribute

2.1. Extended Community Type

The new QoS Marking Attribute is encoded as a BGP Extended Community
Attribute [RFC4360] (Sangli, S., Tappan, D., and Y. Rekhter, "BGP
Extended Communities Attribute," February 2006.). It is therefore a
transitive optional BGP attribute with Type Code 16. An adoption to the
simple BGP Community Attribute encoding [RFC1997] (Chandrasekeran, R.,
Traina, P., and T. Li, "BGP Communities Attribute," August 1996.) is
not defined in this document. The actual encoding within the BGP
Extended Community Attribute is as follows.
The QoS Marking Attribute is of regular type which results in a 1 octet
Type field followed by 7 octets for the QoS marking 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-x3f).

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0 0 x x x x x x|                                               |
   +-+-+-+-+-+-+-+-+   7 octet QoS Marking Attribute structure     |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Note to RFC Editor: The actual type number needs to replace the '0xYY
(0x00-x3f)' after the IANA assignment has occurred.

2.2. Structure of the QoS Marking Attribute

The QoS Marking Attributes provides a flexible encoding structure for
various QoS Markings on different layers. This flexibility is achieved
by a Flags, a QoS Set Number and a Technology Type field within the 7
octet structure as defined below.

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    Flags      | QoS Set Number|        Technology Type        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | QoS Marking O | QoS Marking A |   P. Count    | C. Unused ='0'|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Flags:

```
    0  1  2  3  4  5  6  7
   +--+--+--+--+--+--+--+--+
   |L2 L1 L0|R |I |A |0 |0 |
   +--+--+--+--+--+--+--+--+
```

> All used and unused flags default to a value of '0'. The
> following table shows the bit encoding of the Flags field.

Bit	Flag	Encoding
0-2	L2 & L1 & L0	Specification of the applied enumeration list for the Technology Type
3	R	'1' … remarking occurred
4	I	'1' … QoS marking ignored
5	A	'1' … QoS class aggregation occurred
6,7	unused	Default to '0'

> Technology Type Enumeration can be chosen from different sources
> as follows.

L2&L1&L0	Enumeration List
'000'	GMPLS LSP encoding Type see (IANA, "GMPLS Signaling Parameters," June 2008.) [GMPLS-SIG]
'001'	MPLS Pseudowire Type see (Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)," April 2006.) [RFC4446]
'010'	EtherType see (IEEE, "IEEE EtherType registration list," June 2008.) [EtherType]
'011'	IANA Protocol Numbers see (Arkko, J. and S. Bradner, "IANA Allocation Guidelines for the Protocol Field," February 2008.) [RFC5237]
'100'	

	MIB II Interface List [see (McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB," June 2000.)](#) [RFC2863]
'111'	alternative Technology Type [see (Technology Type Enumeration)](#)

> The Flags "R, I and A" are set to '0' in the advertisement by the IP prefix originating AS. Transit ASes MUST change the flag value to '1' once the respective event occurred. If the QoS marking actively used in the transit AS internal forwarding is different from the advertised original one, the 'Remarking (R)' flag is set to '1'. This MUST be done separately for each technology type attribute within the attribute set. The same applies to the 'Ignore (I)' flag, if the respective advertised QoS marking is ignored in the transit AS internal forwarding.
>
> The 'Aggregation (A)' flag MUST be set to '1' by the UPDATE message relaying transit AS, if the respective IP prefixes will be advertised inside an IP prefix aggregate.

QoS Set Number:

> Several single QoS Marking Attributes can be grouped logically into a QoS Marking Attribute Set characterized by a QoS Set Number. This grouping of the single QoS Marking Attributes into a set provides a cross-layer linking between the QoS class encodings. The number of signalled QoS Marking Attributes as well as QoS Marking Attribute Sets is at the operator's choice of the originating AS. The enumerated QoS set numbers have BGP UPDATE message local significance starting with set number 0x00.

Technology Type:
The technology type encoding follows commonly used enumerations as referred to by the L2&L1&L0 bit combinations stated above. If the use of external enumerations is not favoured or the mapping of commonly used technologies is sufficient, the enumeration list [in (Technology Type Enumeration)](#) SHALL be used.
QoS Marking / Enumeration O & A:
The interpretation of these identically approached fields depends on the selected layer and technology. QoS mechanisms using bit encodings for the targeted priority (e.g. IP DSCP, Ethernet User Priority, MPLS EXP etc.) MUST use a copy of the encoding in this attribute field. Unused higher order bits default to '0'. Other technologies, which use other ways of labelling priority information such as L-LSPs, VPI/VCI inferred ATM classes, lambda inferred priority, etc. SHALL use class enumerations as encoding in this attribute field.
There are two QoS Marking fields within the QoS Marking Attribute for the "original (O)" and the "active (A)" QoS marking.
QoS Marking O (Original QoS Marking):

> The IP prefix originating AS copies the internally associated QoS
> encoding of the given Technology Type into this one octet field.
> The field MUST remain unchanged in BGP UPDATE messages of
> relaying nodes.

QoS Marking A (Active QoS Marking):

> QoS Marking A and O MUST be identically encoded by the prefix
> originating AS.

> All other ASes use this Active QoS Marking field to advertise
> their internal QoS encoding of the given class and technology. A
> cleared Marking field signals that this traffic class experiences
> default traffic treatment within the transit AS forwarding
> technology.

Processing Count (P. Count):

> Each BGP instance, which processes the attribute and appends a
> different AS number to the AS_PATH, MUST increase this counter by
> one. The attribute originating instance initializes the counter
> value to '0x00'.

Currently Unused (C. Unused):

> This one octet field is currently unused and MUST be set to
> '0x00'.

2.3. Technology Type Enumeration

A small list of technologies is provided in the table below for the
direct encoding of common technology types.

Value	Technology Type
0x0000	DiffServ enabled IPv4
0x0001	DiffServ enabled IPv6
0x0010	Ethernet using 802.1q priority tag
0x0020	MPLS using E-LSP
0x0021	MPLS using L-LSP
0x0100	GMPLS – time slot encoding
0x0101	GMPLS – lambda encoding
0x0102	GMPLS – fibre encoding

3. Attribute Usage

Providers MAY choose to
process the QoS Marking
Attributes and adopt the
priority encoding according
to their local policy.
Whether this MAY also lead
to different IGP routing decisions or even effect BGP update filters is
out of scope for the attribute definition.

Only the IP prefix originating AS is allowed to signal the QoS Marking
Attributes and Sets. Transit ASes MUST NOT modify or extend the QoS Set
except for the update of each 'QoS Marking A' field contained in the
Attribute Set, the respective "R, I, A" flags and the Processing
Counter.

3.1. QoS Marking Attribute Example

See Appendix A (QoS Marking Attribute Example) for an example QoS
Marking Attribute Set.

3.2. AS Border Packet Forwarding

IP packet forwarding based on packet header QoS encoding might require
remarking of packets in order to match AS internal policies and
encodings of neighbouring ASes.
Identical QoS class sets and encodings between neighbouring ASes do not
require any remarking. Different encodings will be matched on the
outgoing traffic.
Outgoing traffic for a given IP prefix uses the 'QoS Marking A'
information of the respective BGP UPDATE message QoS Marking Attribute
for adopted remarking of the forwarded packet.
If the Process Count is smaller than the number of different AS numbers
in the AS PATH or if the 'I' flag is set for a given encoding, the
outgoing traffic will be remarked using the 'QoS Marking O' information
instead.
'QoS Marking O' information is also used for outgoing remarking, if the
targeted class is not supported by the neighbouring AS.

3.3. IP Prefix Aggregation

Several IP prefixes of different IP prefix originating ASes MAY be
aggregated to a shorter IP prefix in transit ASes. The QoS Marking
Attribute Set of the resulting IP prefix aggregate is chosen as
follows:

1. If only on aggregate member contains an associated QoS Marking
   Attribute Set, than this Set MUST be used as resulting QoS
   Marking Attribute Set for the IP prefix aggregate.

2. If several aggregate members contain an associated QoS Marking
   Attribute Set, than the Set of the shortest IP prefix aggregate
   member MUST be used as resulting QoS Marking Attribute Set for
   the IP prefix aggregate. Equal IP prefix aggregate member
   lengths SHOULD be resolved in favour of the IP prefix with
   lower IP addresses.

In case of IP prefix aggregation, the 'Aggregation (A)' flag of each
QoS Marking Attribute within the Set MUST be set to '1'.

---

4. Confidentiality Considerations

The disclosure of confidential AS intrinsic information is of no
concern since the signalled marking for QoS class encodings can be
adopted prior to the UPDATE advertisement of the IP prefix originating
AS. AS internal cross-layer remarking and policy based update filtering
allows for consistent QoS class support despite made up QoS class set
and encoding information within UPDATE advertisements. In case of such
policy hiding strategy the required originating AS internal ingress and
egress remarking SHALL be done transparently without explicit "Active
Marking" and 'R' flag signalling.

---

5. IANA Considerations

This document defines a new BGP Extended Community Attribute, which
needs to be assigned a number by IANA within the Extended Community
Attribute list.
The new QoS Marking Attribute is a BGP Extended Community Attribute of
regular type. It is IANA-assignable (FCFS procedure) and is transitive
across ASes.
An number assignment application within the numbering range of 0x00 -
x3f is made to IANA.
Note to RFC Editor: this section may be removed on publication as an
RFC.

---

6. Security Considerations

The new QoS marking Attribute does not raise extra security concerns.
Existing BGP security measures are in place by means of the reused
transport within BGP UPDATE messages.

---

7. References

7.1. Normative References

[RFC1997]	Chandrasekeran, R., Traina, P., and T. Li, "BGP Communities Attribute," RFC 1997, August 1996 (TXT).
[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC2863]	McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB," RFC 2863, June 2000 (TXT).
[RFC4271]	Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, January 2006 (TXT).
[RFC4360]	Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute," RFC 4360, February 2006 (TXT).
[RFC4446]	Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)," BCP 116, RFC 4446, April 2006 (TXT).
[RFC5237]	Arkko, J. and S. Bradner, "IANA Allocation Guidelines for the Protocol Field," BCP 37, RFC 5237, February 2008 (TXT).

7.2. Informative References

[EtherType]	IEEE, "IEEE EtherType registration list," June 2008.
[GMPLS-SIG]	IANA, "GMPLS Signaling Parameters," June 2008.

Appendix A. QoS Marking Attribute Example

The example AS is advertising several IP prefixes, which experience equal QoS treatment from AS internal networks. The IP packet forwarding policy within this originating AS defines e.g. 3 traffic classes for IP traffic (DSCP1, DSCP2 and DSCP3). These three classes are also consistently taken care off within AS internal 802.1q Ethernet switches (using UPC1, UPC2 and UPC3) as well as within the EXP bit supporting MPLS tunnel forwarding. The BGP UPDATE message for the announced IP prefixes will contain the following QoS Marking Attribute Set together with the IP prefix NLRI.

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 0 0|0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 1 0 1 1 1 0|0 0 1 0 1 1 1 0|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 0 0|0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 0 1 1 1|0 0 0 0 0 1 1 1|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 0 0|0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 0 1 0 1|0 0 0 0 0 1 0 1|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 0 1|0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 1 1 0 0 0|0 0 0 1 1 0 0 0|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 0 1|0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 0 1 0 1|0 0 0 0 0 1 0 1|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 0 1|0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 0 0 1 1|0 0 0 0 0 0 1 1|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 1 0|0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 1 0 0 0|0 0 0 0 1 0 0 0|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 1 0|0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 0 0 0 1|0 0 0 0 0 0 0 1|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1 1 1 0 0 0 0 0|0 0 0 0 0 0 1 0|0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0 0 0 0 0 0 0 1|0 0 0 0 0 0 0 1|    currently used = '0'       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The class set as well as the example encodings are arbitrarily chosen.

Author's Address

	Thomas Martin Knoll
	Chemnitz University of Technology
	Reichenhainer Str. 70 /331
	Chemnitz, 09126
	GERMANY
Phone:	+49-371-531-33246
Fax:	+49-371-531-833246
Email:	knoll@etit.tu-chemnitz.de

Full Copyright Statement

Intellectual Property