IDR
Internet-Draft
Intended status: Standards Track
Expires: August 7, 2016

S. Shah
K. Patel
Cisco Systems
S. Bajaj
Juniper Network
L. Tomotaki
Verizon
M. Boucadair
Orange

February 4, 2016

Inter-domain SLA Exchange Attribute
draft-ietf-idr-sla-exchange-07.txt

Abstract

Network administrators typically enforce Quality of Service (QoS) policies according to Service Level Agreement (SLA) with their providers. The enforcement of such policies often relies upon vendor-specific configuration language. Both learning of SLA, either thru SLA documents or via some other out-of-band method, and translating them to vendor specific configuration language is a complex, often manual, process and prone to errors.

This document specifies an optional transitive attribute to signal SLA parameters in-band, across administrative boundaries (considered as Autonomous Systems (AS)), thus simplifying and facilitating some of the complex provisioning tasks in situations where BGP is available as a routing protocol.

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 August 7, 2016.

Copyright Notice

Copyright (c) 2016 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 3
2. Terminology 4
3. QoS Attribute Definition 4
  3.1. SLA, QoS attribute sub-type, Definition 5
  3.2. SLA SubType Value Field 6
  3.3. SLA-Content per Event Field 8
    3.3.1. Supported IPFIX values for Classifier Elements . . . 12
    3.3.2. Traffic Class Service TLVs 13
4. Originating SLA Notification 20
  4.1. SLA Contexts 20
    4.1.1. SLA Advertisement for Point-to-Point Connection . . . 21
    4.1.2. SLA Advertisement for Destination AS Multiple Hops Away 21
5. SLA Attribute Handling at Forwarding Nodes 22
  5.1. BGP Node Capable of Processing QoS Attribute 22
  5.2. SLA Attribute Handling at Receiver 22
6. Error Handling 23
7. Traffic Class Mapping 23
8. Deployment Considerations 24
9. Acknowledgements 25
10. IANA Considerations 25
11. Security Considerations 27
12. References 27
  12.1. Normative References 27
  12.2. Informative References 28
Authors' Addresses 29

1. Introduction

Typically there is a contractual Service Level Agreement (SLA) for QoS established between a customer and a provider or between providers [RFC7297]. This QoS SLA defines the nature of the various traffic classes and services needed within each traffic class. The contract may include full line-rate or sub line-rate without additional traffic classes, or it may contain additional traffic classes and service definitions for those traffic classes. Finer granular traffic classes may be based on some standard code points (e.g., based on DSCP (Differentiated Services Code Point)) or specific set of prefixes.

Once the contractual QoS SLA is established, QoS SLA parameters are enforced in some or all participating devices by deriving those parameters into configuration information on respective devices. The network administrator translates the QoS SLA to QoS policies using router (vendor) specific provisioning language. In a multi-vendor network, translating SLAs into technology-specific and vendor-specific configuration requires the network administrator to consider specific configuration of each vendor. There does not exist any standard protocol to translate SLA agreements into technical clauses and configurations and thus both the steps of out of band learning of negotiated SLA and provisioning them in a vendor specific language can be complex and error-prone.

SLA parameters may have to be exchanged through organizational boundaries, thru SLA documents or via some other off-band method, to an administrator provisioning actual devices.

For example, to provide voice services, the provider may negotiate QoS parameters (like min/max rates) for such traffic classified under the EF (Expedited Forwarding) codepoint in Diffserv-enabled [RFC2475] networks. The Administrator at the CE (Customer Edge) not only will have to know that provider's service for voice traffic is EF-based but will also have to know how to implement DSCP EF classification rule along with Low Latency Service, and possibly min/max rate enforcement for the optimal use of bandwidth, as per vendor specific provisioning language.

The Inter-domain exchange of QoS SLA exchange policy described in this document does not require any specific method for the provider in establishing SLAs. It only requires that the provider wishes to send the QoS SLA policy via BGP UPDATE [RFC4271] messages from the provider to a set of receivers (BGP peers) who will enact the policy. In reaction to, a receiving router may translate that to relevant QoS policy definition on the device.

This document defines a new optional transitive BGP QoS attribute which has as one of its sub-types the SLA policy. The BGP speakers of the originating AS send the BGP Attribute for prefixes this QoS SLA Policy applies to in a BGP UPDATE message that will be distributed to a list of destination ASes. The QoS SLA policy can be for inbound traffic to the advertising AS or outbound traffic from the advertising AS, or both.

The SLA negotiation and assurance is outside the scope of this document. In the future, other sub-types of the QoS Attribute may deal with QoS other than SLA Policy for traffic.

Protocols and data models are being created to standardize setting routing configuration parameters within networks. YANG data models [RFC6020] are being developed so that NETCONF ([RFC6241] or RESTConf ([I-D.ietf-netconf-restconf]) can set these standardized in configuration mechanisms. For ephemeral state, the I2RS protocol is being developed to set ephemeral state. While these protocol provide valid configuration within a domain or across domains, some providers desire to exchange QoS parameters in-band utilizing BGP peering relationships. This is similar to the distribution of Flow Specification information via BGP peering relationships (see [RFC5575] and [RFC7674]).

2. Terminology

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].

3. QoS Attribute Definition

The QoS Attribute is an optional transitive attribute (TBD - attribute code to be assigned by IANA). SLA is defined as one of the sub-types in the QoS attribute. The QoS attribute is only applicable to the NLRIs advertised in the BGP UPDATE message this attribute is included in.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Attr flag   | Attr type QoS |                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               |
~                                                               ~
|                     QoS Attr length/Value                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.......................
```

Attribute flags

highest order bit (bit 0) -
MUST be set to 1, since this is an optional attribute

2nd higher order bit (bit 1) -
MUST be set to 1, since this is a transitive attribute

Figure 1 - QoS BGP attribute

3.1. SLA, QoS attribute sub-type, Definition

The value field of the QoS Attribute contains the following:

QoS Attribute flags, and

Tuple of (SLA sub-type, length, value).

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| QoS Attr flags|      subType  |         subtype Length        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
|                        SubType-Value                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.......................+
```

Figure 2 - Format of BGP QoS Attribute

QoS Attr flags 1 octet. All the bits are currently un-used. The space is provided for future use. All bits MUST be set to zero when sent. The values (0x01-0xFF) are reserved, and MUST be ignored when received.

SubType 1 octet field with the following values:

0x00 = reserved

0x01 = SLA

0x02 - 0x0f = reserved for future use (Standards Action)

SubType length - 2 octet field with length of sub-type value.

SubType Value variable length field containing information about: sender and receiver(s), and SLA parameters described in Section 3.2.

3.2. SLA SubType Value Field

The format of SubType Value field is shown in Figure 3.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    32-bit Source AS (Advertiser)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  32-bit Destination AS count                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                variable list of destination AS                |
~                            ....                               ~
|                            ....                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Event |             SLA id            |      SLA length       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  SLA-Content per SLA Event                    |
~                                                               ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3 - The format of SLA SubType of the BGP QoS attribute

Source AS: 32-bit source AS number. This is the AS that is advertising SLA

0 = ignore Source and Destination AS list from this value field

Note that AS = 0, used in message outside of QoS attribute, is illegal in normal BGP operations. AS = 0 inside the QoS attribute may be used simply as a flag to indicate to the receiver to ignore Source and Destination AS list from inside the QoS attribute.

Destination AS count: 32-bit destination AS count to take variable length AS list. This count has no functional value when Source AS is 0.

0 = QoS attribute is relevant to every receiver of the message

Destination AS list:

32-bit destination AS number

....

.... [as many as AS count]

SLA Event:

4-bits with values

0x0 = reserved

0x1 = ADVERTISE

0x2 to 0xf - Reserved for future use

SLA Id: A 16-bit field containing identifier which is unique in scope of source AS

The significance of an SLA identifier is in the context of the source that is advertising SLA parameters. The SLA identifier is not globally unique but it MUST be unique within the source AS (advertiser).

If the advertised SLA id is different from earlier advertised one, for the same prefix, previous SLA content MUST be replaced with the new advertised one.

The SLA ID applies aggregate for all the traffic to prefixes for a given AFI/SAFI that share same source AS and SLA id.

SLA Length: A 12-bit field indicating the length of SLA-Content. The SLA-content is optional for each advertised SLA id. If the SLA-content field does not exist, the SLA length field value is zero.

SLA-Content per SLA Event: A variable length field (optional field).

If SLA field exists, it follows the format described in Section 3.3.

If the SLA-Content field does not exist in a BGP UPDATE message that contains the QoS attribute with an SLA Sub-type, then receiver MUST inherit the previously advertised SLA content for the same SLA id from the same Source AS.

If there does not exist any prior SLA to relate to the advertised SLA id, then receiver can ignore the SLA advertisement and process the rest of the BGP message.

The lack of a valid prior SLA-Content field does not make this attribute invalid, so the attribute MUST be forwarded as a valid BGP optional transitive attribute.

3.3. SLA-Content per Event Field

The only event described below is ADVERTISE. The format of SLA-Content for the ADVERTISE event is shown in Figure 4.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|dir|       Traffic Class count     | Class Desc Len|           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                  Traffic Class Description                    ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Elements Count|                                               |
+-+-+-+-+-+-+-+-+                                               ~
|                                                               |
~              Traffic Class Elements (TLVs)                    ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Service  Count|                                               |
+-+-+-+-+-+-+-+-+                                               ~
|               Traffic Class Service TLVs                      |
~                                                               ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~  Repeat from Traffic Class Description for next Traffic Class ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~    Repeat from direction for SLA in the other direction       ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4 - SLA-Content for event ADVERTISE

SLA-content contains set of Traffic Class Elements (Classifiers) and Service TLVs for a list of Traffic Classes. This list of Traffic Classes MUST be specified for one direction first and then optionally followed by for the opposite direction.

dir (Direction): 2 bit field that indicates Direction of the SLA. The following values are defined:

0x0 = reserved

0x1 = incoming, to source AS from destination AS

0x2 = outgoing, from source AS towards destination AS

0x3 = for future use

Traffic Class (Classifier Group) count: 16 bit field with the count of number of classifier groups. The value of zero (0x00)in this field is a special value which invalidates previous advertised SLA (if any exist).

Class Descr Len: 8 bit field that contains the length of the Traffic Class Description field. The value of zero in this field indicates that no Traffic Class Description field follows.

Traffic Class Description: The description field MUST carry UTF-8 encoded description.

Elements (Classifier) Count: 8 bit field containing the count of traffic elements. The value zero (0x00) means there are no elements in the traffic class, and thus the traffic class is to classify rest of the traffic not captured otherwise by other traffic classes in the set for a given direction.

It is RECOMMENDED that Traffic Class that has 0 elements is present last in the advertised list of Traffic Classes.

If an advertised message has it positioned somewhere else, then the receiver MUST re-order it, for the forwarding purpose, to the last position in the advertised list of Traffic Classes from a given source AS.

The QoS attribute advertised from a specific source MUST NOT have more than one such Traffic Classes (Traffic Class with 0 elements count). If there are more than one such Traffic Classes present then it is an error condition which should follow handling of such BGP message as described in the Error handling section.

Classifier Element TLVs: (optional) variable length field containing as many TLVs specified by the Elements count field. Each TLV has the following format:

IPFIX Element Identifier: (8 bit type field) IPFIX Identifiers listed in Table 1.

Size of Value field: (8 bit field) - Indicates the length of the value field.

Value: A variable field containing a value appropriate for the IPFIX element. It is an error if the IPFix field does not contain the appropriate format (see BGP error handling in section 6). Only the IPFIX elements shown in Table 1 are supported.

Any Traffic Class element advertised in the QoS attribute only applies to the NLRI advertised (AFI/SAFIs) within the BGP UPDATE message the QoS attribute is contained in. If a receiver receives a BGP UPDATE message with QoS/SLA attribute for an NLRI that it does not support then the receiver MUST NOT install that advertised SLA and continue to forward this attribute as an optional transitive attribute.

Service Count: 8 bit count of Traffic Class Service TLVs following this count. A value of zero is a special value indicating "no bounded service" (a.k.a., Best Effort (BE)).

Traffic Class Service TLVs: (optional) variable length field with the following format for the TLVs

Traffic Class Service type: 16-bit field which specifies a service type. Each service type is detailed in Section 3.3.2. The list of available service types are,

0x00 = reserved

0x01 = TRAFFIC_CLASS_TSPEC

0x02 = L2_OVERHEAD

0x03 = MINRATE_IN_PROFILE_MARKING

0x04 = MINRATE_OUT_PROFILE_MARKING

0x05 = MAXRATE_IN_PROFILE_MARKING

0x06 = MAXRATE_OUT_PROFILE_MARKING

0x07 = DROP_THRESHOLD

0x08 = RELATIVE_PRIORITY

0x09 = SUB_TRAFFIC_CLASSES

Length of value field: 08-bit field that specifies the size of a value field to follow.

TRAFFIC_CLASS_TSPEC type has a fixed size length of a value. It is 96 bits specifying Tspec described in Section 3.3.2.1.

L2_OVERHEAD type has a fixed size length of a value. It is 8 bits as described in Section 3.3.2.2.

MINRATE_IN_PROFILE_MARKING type has a variable length value (see Section 3.3.2.3).

MINRATE_OUT_PROFILE_MARKING type has a variable length value (see Section 3.3.2.4).

MAXRATE_IN_PROFILE_MARKING type has a variable length value (see Section 3.3.2.5).

MAXRATE_OUT_PROFILE_MARKING type has a variable length value (see Section 3.3.2.6).

DROP_THRESHOLD type has a variable length value (see Section 3.3.2.8).

RELATIVE_PRIORITY has a fixed size length of 4 bits specifying the priority value. (see Section 3.3.2.9).

0x09 = SUB_TRAFFIC_CLASSES is a variable length field which allows sub-classes to be specified under traffic classes (see Section 3.3.2.10).

Value field: field containing value appropriate for one of the Service Types. It is an error if this field does not contain the appropriate format (see BGP error handling section for details).

3.3.1. Supported IPFIX values for Classifier Elements

IPFIX [RFC7012] has well defined identifier set for a large number of packet attributes; an IPFIX IANA registry maintains values for packet classifier attributes (https://www.ietf.org/assignments/ipfix"). Only the IPFIX attributes listed in Table 1 are supported by BGP SLA exchange. Any new attribute to be supported by SLA QOS MUST be added by a Standards Action.

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority
8	sourceIPv4Address
27	sourceIPv6Address
9	sourceIPv4PrefixLength
29	sourceIPv6PrefixLength
44	sourceIPv4Prefix
170	sourceIPv6Prefix
12	destinationIPv4Address
28	destinationIPv6Address
13	destinationIPv4PrefixLength
30	destinationIPv6PrefixLength
45	destinationIPv4Prefix
169	destinationIPv6Prefix
4	protocolIdentifier
7	sourceTransportPort
11	destinationTransportPort

Table 1

3.3.2. Traffic Class Service TLVs

3.3.2.1. Traffic Class TSPEC TLV

The TRAFFIC_CLASS_TSPEC TLV consists of:

type = 0x01

length = 96 bits (12 octets) TSpec field

value = 96 bits, TRAFFIC_CLASS_TSPEC value consists of the (r), (b), (p) parameters as described in Invocation Information section of [RFC2212] and shown in Figure 5. Note that inheriting the definition of TSpec here does not enable RFC2212 functionality. Only the values of the Traffic Specification are used in this specification.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Minimum Rate (r) (32-bit IEEE floating point number)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Burst Size (b) (32-bit IEEE floating point number)           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Maximum Rate (p) (32-bit IEEE floating point number)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5 - Traffic Class TSPEC

Format of Parameters (r), (b) and (p): are 32-bit IEEE floating point numbers. Positive infinity is represented as an IEEE single precision floating-point number with an exponent of all ones and a sign mantissa of all zeros. The format of IEEE floating-point numbers is further summarized in [RFC4506].

Parameter (r): indicates min-rate of the traffic class. This rate indicates the minimum rate, measured in octets of Layer 2 (L2) datagrams per second, that the service advertiser is providing for a given class of traffic on advertiser's hop. Note that it does not necessarily translate to a minimum rate service to the receiver of an SLA unless the traffic class definition clearly represents a sole receiver of an SLA. If there is no SLA for min-rate, the value of (r) MUST be set to 0.

Parameter (b): indicates maximum burst size, measured in octets of L2 datagram size. Since queuing delay can be considered a function of burst size (b) and min-rate (r), in presence of non-zero parameter (r), parameter (b) represents bounded delay for the Traffic Class. This delay is a single hop queuing delay when SLA is to be implemented at the resource constrained bottleneck. In other words this burst size can be considered as a buffer size. Value of 0 for parameter (b) means the advertiser does not mandate specific bounded delay.

Parameter (p): indicates max-rate of the traffic class. Just like min-rate, max-rate, measured in octets of L2 packets per second, field here also indicates service provided by advertiser. If advertiser does not have any specific value to set for a given class of traffic, it MAY be set to physical interface line rate or any other indirect limit that may affect this class' maximum rate. In absence of any such known value, it MUST be set to positive infinity. Value 0 is considered an error.

3.3.2.2. Traffic Class L2 Overhead

The L2_OVERHEAD traffic class consists of:

Type = 0x02 (L2_OVERHEAD)

Length = 1 octet

value = 8 bits, count of L2 overhead from sender in bytes

By default the packet rate and other packet size related parameters advertised in an SLA include the L2 packet overhead. For the receiver (of the SLA at next hop),this overhead is the L2 overhead of the local link where advertised SLA is received.

However, in cases where advertised SLA is for a receiver multiple hops away, L2 overhead from the source perspective may be different from the local L2 overhead at the receiver. In such cases, the explicit notification of size of L2 overhead from a sender is necessary for the a receiver to be able to know the L2 overhead required by the sender. When the receiver chooses to react to an advertised SLA and if the L2 Overhead service type is present in advertised SLA, the receiver MUST use the explicit advertised L2 overhead rather than the local L2 overhead.

If SLA is required to consider only IP packet size, the sender MAY advertise this service with a value of 0.

3.3.2.3. Traffic Class for MINRATE_IN_PROFILE_MARKING

The MINRATE_IN_PROFILE_MARKING traffic class consists of:

Type = 0x03 = MINRATE_IN_PROFILE_MARKING

Length = 2 octets

Value:

Marking code-point type = 8 bits (1 octet) IPFIX Element Identifier.

Marking code-point value = 8 bits (1 octet) code-point number.

The marking code-point type of 0x00 is a drop identifier; the length in this case is zero.

The following table lists the supported IPFIX Identifiers:

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 2

3.3.2.4. Traffic Class for MINRATE_OUT_PROFILE_MARKING

The MINRATE_OUT_PROFILE_MARKING traffic class consists of:

Type = 0x04 = MINRATE_OUT_PROFILE_MARKING

Length = 2 octets

Value:

Marking code-point type = 8 bits (1 octet) IPFIX Element Identifier.

Marking code-point value = 8 bits (1 octet) code-point number.

The marking code-point type of 0x00 is a drop identifier; the length in this case is zero.

The following table lists the supported IPFIX Identifiers:

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 3

3.3.2.5. Traffic Class for MAXRATE_IN_PROFILE_MARKING

The MAXRATE_IN_PROFILE_MARKING traffic class consists of:

Type = 0x05 = MAXRATE_IN_PROFILE_MARKING

Length = 2 octets

Value:

Marking code-point type = 8 bits (1 octet) IPFIX Element Identifier.

Marking code-point value = 8 bits (1 octet) code-point number.

The marking code-point type of 0x00 is a drop identifier; the length in this case is zero.

The following table lists the supported IPFIX Identifiers:

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 4

3.3.2.6. Traffic Class for MAXRATE_OUT_PROFILE_MARKING

The MAXRATE_OUT_PROFILE_MARKING traffic class consists of:

Type = 0x06 = MAXRATE_OUT_PROFILE_MARKING

Length = 2 octets

Value:

Marking code-point type = 8 bits (1 octet) IPFIX Element Identifier.

Marking code-point value = 8 bits (1 octet) code-point number.

The marking code-point type of 0x00 is a drop identifier; the length in this case is zero.

The following table lists the supported IPFIX Identifiers:

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 5

3.3.2.7. Precedence between MINRATE and MAXRATE

The precedence between MINRATE_IN_PROFILE_MARKING, MINRATE_OUT_PROFILE_MARKING, MAXRATE_IN_PROFILE_MARKING, and MAXRATE_OUT_PROFILE_MARKING when all four are advertised is:

- o MINRATE_IN_PROFILE_MARKING takes highest precedence (that is over MAXRATE_IN_PROFILE_MARKING),

- o MAXRATE_IN_PROFILE_MARKING takes precedence over MINRATE_OUT_PROFILE_MARKING, and

- o MAXRATE_OUT_PROFILE_MARKING takes precedence over MINRATE_OUT_PROFILE_MARKING

3.3.2.8. Traffic Class for DROP_THRESHOLD

The DROP_THRESHOLD traffic class consists of:

Type = 0x07 - DROP_THRESHOLD

Length = 1 octet that specifies total length of all set of drop thresholds.

A set of drop threshold contains list of code-points of a specific type sharing a threshold in unit of bytes. There MAY be more than one set of such threshold for this Service Type per Traffic Class.

Value: number of set of thresholds and values in form of a sub-TLV for each set.

Number of set of thresholds = 1 octet

sub-TLV for each set: Each sub-TLV specifies a code-point type/values that the burst size is applicable to. The sub-TLV is in the form of a (code-point type, value length, value) where value = list of code-points + burst size in unit of bytes applicable to that code-points.

sub-TLV code point type = 8 bits (1 octet) IPFIX Element Identifier from the list shown in Table 6.

sub-TLV Length = 1 octet that specifies total length for set of code-points and burst size.

sub-TLV Value: variable length field with

sequence of code points - one code-point value specified in 1 octet.

4 octets burst size - 32 bit (4 octets) IEEE Floating point number.

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 6

3.3.2.9. Traffic Class for Relative Priority

The RELATIVE_PRIORITY traffic class consists of:

Type = 0x08 - RELATIVE_PRIORITY

Length = 4 bits

Value:

A value from range of 0 - 15. Lower the value means higher the priority.

Relative priority indicates scheduling priority of this traffic class. Voice traffic, for example, which requires lowest latency compared to any other traffic, may have lowest value advertised in relative priority. For two different traffic classification groups where one application group may be considered more important than the other but from a scheduling perspective does not require to be distinguished with a different priority, relative priority for those classification groups may be advertised with the same value.

3.3.2.10. Traffic Class for Sub-Traffic Classes

The SUB_TRAFFIC_CLASSES traffic class consists of:

Type = 0x09 - SUB_TRAFFIC_CLASSSES

length = the combined length of a set of traffic Class TLVs included in a Sub-Traffic Classes grouping

value = sequence of traffic class TLVs

For SLAs where a specific Traffic Class may further have different sub-services for a sub-group of Classifier Elements, this service type SHOULD be used to further divide Traffic Class in multiple sub-classes. Each sub-class is then defined with their own classifier elements and service types.

4. Originating SLA Notification

The QoS attribute MUST only be added by the originator and MUST NOT be added during BGP propagation.

BGP UPDATE message with the QoS Attribute carrying SLA parameters SHOULD NOT be sent periodically just for the purpose of KEEPALIVE between two points. Some sort of SLA policy change may be considered as a trigger for the advertisement.

For any modified SLA parameters, the originator MUST re-advertise the entire set of SLA parameters. There is no provision to advertise partial set of parameters. To invalidate previously advertised SLA parameters, a message MUST be sent with the same SLA id for the same source with the Traffic Class count set to 0.

4.1. SLA Contexts

In certain cases, the advertisement of a QoS Attribute in a BGP UPDATE message may relate to an SLA for aggregate traffic over a point-to-point connection between a specific destination and a specific source. A point-to-point connection may be the physical link, that connects two BGP peers, or may be a virtual link (e.g. a tunnel). In such cases, a BGP update message with source AS number and NLRI prefix of source end-point can uniquely identify physical/virtual link in order to establish the context for the advertised SLA's for that point to point link.

In the simplest case where provider (e.g., PE) and Customer (e.g., CE) devices are directly connected via a physical link and have only

a single link between them, the CE can uniquely identify the forwarding link to PE with the following:

- o AS number of the PE,
- o NLRI prefix being an IP address of PE,
- o next hop to CE (that is the next hop address from CE to PE).

The SLA advertised in the QoS attribute in the BGP UPDATE message sent from the PE to a CE, along with the PE's AS number and IP address, establishes SLA context for the aggregate traffic through CE-to-PE link.

The SLA advertised in the QoS attribute in the BGP UPDATE message from PE to CE, with PE's AS number and any other prefix establishes SLA for that specific prefix's traffic as a subset of traffic under CE-to-PE link.

Even though this example is in the context of IP prefixes, QoS attribute's SLA exchange does not have to be limited to the IP address family (IPv4 and IPv6). SLA advertisement is generic to all forms of NLRI types that are supported by the BGP specification (like IPv4, IPv6, VPN-IPv4, VPN-IPv6).

4.1.1. SLA Advertisement for Point-to-Point Connection

When BGP UPDATE message with the QoS Attribute with SLA is used to advertise the QoS SLA for a point-to-point connection (physical or logical), the next hop in the BGP message is used with the prefix of the source end-point of the point to point connection.

The destination AS number in the QoS SLA attribute is typically set to the AS of the BGP peer's IP-Address.

If the source AS number in the QoS SLA Attribute is set to zero, the source AS and Destination AS fields in the QoS SLA attribute are ignored. This occurs if the BGP peer is sending an UPDATE message with the QOS SLA directly to a BGP peer (next-hop BGP peer).

4.1.2. SLA Advertisement for Destination AS Multiple Hops Away

When a BGP UPDATE message with a QoS SLA attribute is to be sent by a BGP peer beyond next hop peer, the value of source AS in the QoS attribute MUST be set by the originator of the UPDATE message. If such an update is meant to be for a specific list of AS(es) as receivers, then the list of destination AS(es) MUST be explicitly

described in the QoS attribute message to avoid flooding of the QoS attribute data in the network beyond those destinations.

If a new prefix is added in an AS for which SLA parameters have already been advertised before for other existing prefixes, and if traffic to this new prefix is subject to the same SLA advertised earlier, then the BGP UPDATE for this new prefix may include QoS attribute containing just an SLA id for an SLA id that was advertised earlier. This BGP UPDATE message does not require to have the whole SLA content. The SLA id is sufficient to relate SLA parameters to new advertised prefixes.

5. SLA Attribute Handling at Forwarding Nodes

The propagation of the QoS Attribute in the BGP UPDATE depends on the rules detailed in the following sub-sections for forwarding the QoS Attribute.

5.1. BGP Node Capable of Processing QoS Attribute

If a BGP peer is capable of processing a QoS attribute in a BGP UPDATE message, it MAY process the QoS attribute. If UPDATE has a QoS Attribute with a list of destination ASes, it MAY trim the list and adjust the count of the destination AS to exclude ones that are not required in further announcement of BGP UPDATE messages.

BGP peer MUST drop SLA related sub-type from the QoS attribute, if there are no more ASes from the QoS Attribute's destination list in the forwarding path. The rest of the QoS attribute contents MAY be forwarded if there exist other sub-types of QoS attribute and forwarding rules meets other sub-types requirements. If there is no other sub-types in the QoS attribute content then the node MUST drop the entire QoS attribute all together. The other attributes and NLRI information MAY be announced further if they meet rules defined by other attributes and BGP specification.

Except extracting the entire SLA sub-type of the QoS attribute and trimming the list of destination AS list, all other content MUST NOT be modified by any intermediate receivers of the message.

5.2. SLA Attribute Handling at Receiver

Reception of and processing of advertised QoS SLA content are optional for the receiver. While reacting to SLA advertisement in a QoS attribute,

- o the receiving BGP peer SHOULD invalidate previous advertised SLA parameters if one exists for the same SLA id and source AS. If

the new advertised SLA has a non-zero traffic count, then the new advertised SLA SHOULD be installed. If new advertised SLA update is with Traffic Class count 0, then no further action is required.

- o When BGP UPDATE messages are triggered only as a result of SLA policy change, BGP UPDATE message forwarding beyond intended BGP peer receivers is not necessary. If the receiver device implementation supports policy based filtering, then the receiver MAY implement a policy to filter such messages based on the prefix and attribute.

If QoS attribute with the SLA is advertised to the next hop BGP peer who is a neighbor, the receiver MAY implement advertised SLA for the whole link; the link could be a physical or virtual connected to the neighbor. If the QoS Attribute with the SLA is advertised to a BGP peer which is not the next hop neighbor, then receiver may establish advertised SLA for that specific prefix list under the relevant link. It is completely up to the receiver to decide for which prefixes it should accept advertised SLA and for which ones it will not accept.

6. Error Handling

Error conditions, while processing of the QoS attribute, MUST be handled with the approach of attribute-discard as described in [RFC7606]. In such condition, the receiver SHOULD also cleanup any previously installed SLA state for the same prefix.

7. Traffic Class Mapping

It is possible that forwarding methods used in two different ASes could be different. For example, provider may tunnel a customer's IP traffic thru an MPLS infrastructure. In such cases, the traffic class definition for QoS services may differ between the ASes. For the meaningful use of advertised SLA in such cases, the receiver is required to map the remote traffic class to the local traffic class.

In the example given, traffic classification in Customer AS could be IP Diffserv-based whereas traffic classification in Provider AS could be MPLS TC-based. Thus for advertised MPLS TC-based SLA would require to map traffic class from IP Diffserv-based to MPLS TC type [RFC3270].

There are well-defined recommendations that exist for traffic class mapping between two technologies (e.g. [RFC3270] maps between DSCP and MPLS TC). Receiver MAY use those defined recommendations for traffic class mapping or MAY define its own as per its network Traffic Class service definition to map to advertised Traffic

Classes. It is completely up to the receiver how to define such traffic class mapping.

8. Deployment Considerations

One of the use cases is for a provider to advertise contracted SLA parameters to a Customer Edge (CE) in cases where eBGP is deployed between PE and CE. The SLA parameters may already be provisioned by the provider on the PE device (facing CE). This provisioned SLA parameters are then advertised thru proposed BGP QoS attribute to the CE device. The CE device may read the attribute and SLA sub-type content to implement the QoS policy on the device.

Contracted SLA from PE to CE may be full line-rate or sub line-rate or finer granular controlled services. The advertised SLA can be useful when contracted service is sub-rate of a link and/or when for finer granular traffic classes that are controlled (e.g. voice, video services may be capped to certain rate).

```
                          _______________
        __________         /               \
       /          \       /                 \
      /            \     /                   \
      |CustomerSite|-----|      Provider     |
      \           C/E   P\E                  /
       \__________/       \                 /
                           \_______________/
           AS 3                   AS 2


          SLA_ADVERTISE: AS2 to AS3
          NLRI = PE ip address
```

Figure 6 - Example 1

Another use case can be to advertise SLAs among different network sites within one Enterprise network. In Hub and Spoke deployments, Administrator may define SLAs at spoke and advertise QoS SLA parameters to the Hub thru BGP updates. In Figure 7, each spoke (AS1 and AS2) are connected to Hub (AS3) via a VPN tunnel. As shown in Figure 7, AS2 can advertise SLA to AS3 in the context of that tunnel ip address.

```
                                                AS 2
                     _______________        ________
                    /               \      /        \
       _____       /                 \-----| Spoke2 |
     /      \     /                   \    \________/
     | Hub  |-----|      Provider     |     ________
     \______/     \                   /    /        \
                   \                 /-----| Spoke1 |
      AS 3          \_______________/      \________/
                                             AS 1


        SLA_ADVERTISE: AS2 to AS3
                       NLRI = AS2 tunnel address

        SLA_ADVERTISE: AS1 to AS3
                       NLRI = AS1 tunnel address
```

Figure 7 - Example 2

Deployment options are not limited to involving CEs, PE-to-CE or CE-to-CE, only. For any contract between two providers, SLA parameters may be advertised from one to the other.

9. Acknowledgements

Thanks to Fred Baker, David Black, Sue Hares and Benoit Claise for their suggestions and to Christian Jacquenet, Ken Briley, Rahul Patel, Fred Yip, Lou Berger, Brian Carpenter, Bertrand Duvivier, Bruno Decraene for the review.

10. IANA Considerations

This document defines a new BGP optional transitive path attribute, called QoS Attribute. IANA action is required to allocate a new code-point in the BGP path Attributes registry.

IANA is requested to create a registry for QoS Attribute subTypes. This is a registry of 1 octet value, to be assigned on a standards action/early allocation basis. The initial assignments are:

```
    QoS Attribute subTypes
        ============= ========
    Reserved       0x00
    SLA            0x01
        Reserved       0x02-0xff (Standards Action)
```

IANA is requested to create a registry for SLA Event Types. This is a registry of 4-bits value, to be assigned on a standards action or early allocation basis. The initial assignments are:

QoS Attribute	SLA Event Types
Reserved	0x00
ADVERTISE	0x01

IANA is requested to create a registry to define QoS SLA Direction. This is the direction in forwarding path, advertised QoS SLA is applicable to. The values for QoS SLA direction are:

QoS SLA Direction	Value
Reserved	0x00
To source AS from destination AS	0x01
From source AS to destination AS	0x02

QoS SLA Traffic Class Element Types will be referring to existing IPFIX IANA types as listed in Table 1.

IANA is requested to create a registry for QoS SLA Traffic Class Service Types. This is a registry of 2 octet values, to be assigned on a standards action/early allocation basis. The initial assignments are:

Traffic Class Service Type	Value
Reserved	0x00
TRAFFIC_CLASS_TSPEC	0x01
L2_OVERHEAD	0x02
MINRATE_IN_PROFILE_MARKING	0x03
MINRATE_OUT_PROFILE_MARKING	0x04
MAXRATE_IN_PROFILE_MARKING	0x05
MAXRATE_OUT_PROFILE_MARKING	0x06
DROP_THRESHOLD	0x07
RELATIVE_PRIORITY	0x08
SUB_TRAFFIC_CLASSES	0x09

11. Security Considerations

The QOS attribute defined in this document SHOULD be used by the managed networks for enforcing Quality of Service policies and so there should not be any risks for identity thefts. To strengthen the security for the QoS attribute, RPKI based origin validation [RFC7115] MAY be used. In addition to the RPKI based origin validation, BGP Path Validation (e.g., [I-D.ietf-sidr-bgpsec-protocol]) procedures could be used over BGP QoS attribute and its associated prefix in producing the digital signature that can be carried within the signature SLA for the messages. This would help prevent any man- in-the-middle attacks.

12. References

12.1. 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>.

[RFC2212] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, DOI 10.17487/RFC2212, September 1997, <http://www.rfc-editor.org/info/rfc2212>.

[RFC3270] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, DOI 10.17487/RFC3270, May 2002, <http://www.rfc-editor.org/info/rfc3270>.

[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>.

[RFC4506] Eisler, M., Ed., "XDR: External Data Representation Standard", STD 67, RFC 4506, DOI 10.17487/RFC4506, May 2006, <http://www.rfc-editor.org/info/rfc4506>.

[RFC7012] Claise, B., Ed. and B. Trammell, Ed., "Information Model for IP Flow Information Export (IPFIX)", RFC 7012, DOI 10.17487/RFC7012, September 2013, <http://www.rfc-editor.org/info/rfc7012>.

[RFC7115] Bush, R., "Origin Validation Operation Based on the Resource Public Key Infrastructure (RPKI)", BCP 185, RFC 7115, DOI 10.17487/RFC7115, January 2014, <http://www.rfc-editor.org/info/rfc7115>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <http://www.rfc-editor.org/info/rfc7606>.

12.2. Informative References

[I-D.ietf-netconf-restconf]
Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", draft-ietf-netconf-restconf-09 (work in progress), December 2015.

[I-D.ietf-sidr-bgpsec-protocol]
Lepinski, M., "BGPsec Protocol Specification", draft-ietf-sidr-bgpsec-protocol-14 (work in progress), December 2015.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <http://www.rfc-editor.org/info/rfc2475>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <http://www.rfc-editor.org/info/rfc5575>.

[RFC6020] Bjorklund, M., Ed., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)", RFC 6020, DOI 10.17487/RFC6020, October 2010, <http://www.rfc-editor.org/info/rfc6020>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <http://www.rfc-editor.org/info/rfc6241>.

[RFC7297] Boucadair, M., Jacquenet, C., and N. Wang, "IP Connectivity Provisioning Profile (CPP)", RFC 7297, DOI 10.17487/RFC7297, July 2014, <http://www.rfc-editor.org/info/rfc7297>.

[RFC7674] Haas, J., Ed., "Clarification of the Flowspec Redirect Extended Community", RFC 7674, DOI 10.17487/RFC7674, October 2015, <http://www.rfc-editor.org/info/rfc7674>.

Authors' Addresses

Shitanshu Shah
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
US

Email: svshah@cisco.com

Keyur Patel
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
US

Email: keyupate@cisco.com

Sandeep Bajaj
Juniper Network
1194 N. Mathilda Avenue
Sunnyvale, CA 94089
US

Email: sbajaj@juniper.net

Luis Tomotaki
Verizon
400 International
Richardson, TX 75081
US

Email: luis.tomotaki@verizon.com

Mohamed Boucadair
Orange
Rennes
35000
France

Email: mohamed.boucadair@orange.com