Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 8, 2014

S. Shah
K. Patel
Cisco Systems
S. Bajaj
Juniper Networks
L. Tomotaki
Verizon
M. Boucadair
France Telecom
Nov 04, 2013

Inter-domain SLA Exchange
draft-ietf-idr-sla-exchange-02

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, many times manual, process and prone to errors. This document proposes an in-band method of SLA signaling which can help to simplify some of the complexities.

This document defines an optional transitive attribute to signal SLA details in-band, across administrative boundaries (considered as Autonomous Systems (AS)), thus simplifying and facilitating some of the complex provisioning tasks.

Though the use case with the proposed BGP attribute is explicitly defined in this document, purpose of this attribute is not limited to this use case only.

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 May 8, 2014.

Copyright Notice

Table of Contents

1. Introduction

Typically there is a contractual Service Level Agreement (SLA) established between Customer and Provider or between providers, possibly using one or the other form of the template [CPP]. This contractual agreement usually defines the nature of the various traffic classes (i.e., traffic match conditions) and services needed for each traffic class. The contract may exist at different levels of traffic granularity. The contract could be for the full line-rate or sub line-rate without granular traffic distinction or it could be for finer granular traffic classes, with services defined. Finer granular classes can be based on some standard code-points (like DSCP) or for a specific set of prefixes or for a set of well-known application types.

Once the SLA is established, SLA parameters are enforced in some or all participating devices by deriving SLA parameters into configuration information on respective devices. 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. In a subsequent step, administrator requires to translate 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 to consider specificities 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. As an example for voice service, the Provider may negotiate QoS parameters (like min/max rates) for such traffic based upon the EF code-point in Diffserv-enabled [RFC2475] networks. Administrator at the CE side not only will have to know that Provider's service for voice traffic is EF-based, so that traffic exiting CE is marked properly, 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.

An in-band signaling method of propagating SLA parameters from provider, PE in an example above, to contractual devices, CE in an example above, can help eliminate manual administrative process described above. Provider may have SLA negotiated with the Customer via some defined off-band method (based on the specifics defined by the Provider or using protocols like [CPNP]. The Inter-domain SLA exchange proposal described in this document does not pre-requisite any specific method of establishing SLAs). The Provider provisions established SLA on the Provider device. This SLA instance then can

be signaled to the Customer via in-band signaling protocol. In reaction to this signal, receiver router may translate that to relevant QoS policy definition on the device.

For an in-band signaling, we propose to use BGP as a transport. The details of SLA parameters are specific to the granularity of traffic classes and their respective treatment, which is independent of the BGP protocol itself. Though we find BGP as a suitable transport for inter-domain SLA exchange for the following reasons:

- The need to exchange SLA parameters between domains (Automated Systems (AS)), where in use-cases described in this document, BGP is a suitable protocol for inter-domain exchange [RFC4271] [RFC4364]
- There is no specifically defined protocol available today for SLA exchange
- BGP updates already advertise specific set of prefixes (flow or flow-group). Other QoS-related attributes, apart from the the use of SLA advertisement, can be added to these updates in the future

The proposal is to define a new BGP attribute to advertise/learn SLA details in-band. The proposed attribute is intended to advertise SLA from one AS to a list of destined ASes. The advertised QoS information could be for the incoming traffic to the advertiser, that is advertising SLA or could be for the outgoing traffic from the advertiser or could be for both directions. Reception of and reaction to advertised SLAs are optional for the receiver.

We propose QoS as an optional transitive attribute, keeping SLA advertisement and discovery (request) as one of the sub-types of QoS attribute. This is to keep the QoS attribute open for extensions. For example, SLA Negotiation and Assurance is out of scope of this document but can be envisioned as another sub-type.

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 proposed here is an optional transitive attribute (attribute type code to be assigned by IANA). SLA is defined as one of the sub-types in the QoS attribute.

```
       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

3.1. SLA, QoS attribute sub-type, Definition

The value field of the QoS Attribute contains TLVs, followed to QoS Attribute flags described in the previous section. One of the TLVs that we define is a 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  |         sub type Length       |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       ~                                                               ~
       |                               Value                           |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.........................
```

The first octet in the Value field of the QoS attribute is QoS attribute specific flags

highest order bit (bit 0) -
It defines if update message MUST be dropped (if set to 1) without updating routing information base, when this is the last BGP receiver from the list of destination ASes this attribute is announced to, or MUST announce (if set to 0) further to BGP peers

The purpose of this bit is discussed further in subsequent sections.

Remaining bits are currently unused and MUST be set to 0

```
    subType - 8 bits
        0x00        = reserved
        0x01        = SLA
        0x02 - 0x0f = for future use
```

SLA sub-type specific value field details. These details contain information about 1) sender and receiver(s) and 2) SLA parameters. SLA Parameters include SLA event type (such as Advertise, Request) and contents associated to that event type.

The format of SLA message is,

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    32-bit source AS (Advertiser)              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Optional advertiserid total len|      Advertiser id TLVs       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               ~
   |                                                               |
   ~                                                               ~
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  32-bit destination AS count                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                variable list of destination AS                |
   ~                            ....                               ~
   |                            ....                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Event |             SLA id            |      SLA length       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Content as per SLA Event                   |
   ~                                                               ~
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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.
        Instead refer to Source and Destination AS as defined by BGP
        message.

Optional advertiser id total len

    16-bit Source address identifier (optional).
    0 = No optional identifier

In general any additional qualifier for an advertiser is not required. The SLA definition is in the context of prefix advertised in the NLRI definition. The exception is where a BGP

speaker, in the middle of an update path to the destination AS, aggregates prefixes. We will refer this middle BGP speaker, that aggregates routes, as an Aggregator. Aggregator is then required to insert original NLRI details in the optional advertiser field

Optional Advertiser id TLV
    4-bit type
    0x0 = reserved
    0x1 = ORIGIN_NLRI, variable length
    0x2 to 0xf = for future use,

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 Type
    4-bits
    0x0 = reserved
    0x1 = ADVERTISE
    0x2 = REQUEST
    0x3 to 0xf, for future use

SLA Id
    16-bit identifier unique within the 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).

The SLA content is optional for an advertised SLA id. If SLA content does not exist in BGP update messages with advertised QoS attribute, that contains the SLA sub-type, then receiver MUST inherit prior advertised SLA content for the same SLA id from the same Source AS.

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

SLA is aggregate for all the traffic to prefixes that share

same source AS and SLA id.

SLA Length
12-bits

The format of SLA ADVERTISE event message is,

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|dir|       Traffic Class count     | Class Desc Len|           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+           ~
|                                                               |
~                  Traffic Class Description                    ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~              Traffic Class Elements count/values              ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Service  Count|      service type/value pair                  |
+-+-+-+-+-+-+-+-+                                               ~
|                                                               |
~                                                               ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~  Repeat from Traffic Class Description for next Traffic Class ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~    Repeat from direction for SLA in the other direction       ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Direction
02-bit for incoming or outgoing traffic,
0x0 = reserved
0x1 = incoming, from destination AS towards source AS
0x2 = outgoing, from source AS towards destination AS
0x3 = for future use

Traffic Class count (Classifier Groups count)
16-bit, count of number of classifier groups
00 = Advertisement to invalidate previous advertised SLA if any

Traffic Class Descr Length
08-bit, size of the length

0 = No description

Traffic Class Description
Ascii Description of the Traffic Class

Traffic Class Elements Count in a Traffic Class,

08-bit count of classifier elements in a specific Traffic Class

00 = this has relative definition. It means classify rest all traffic that is not classified via earlier described Traffic Classes.
It is RECOMMENDED that Traffic Class, that has 0 elements, is present last in the advertised list of Traffic Classes. If Advertised message has it positioned some-where else, then 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. QoS attribute advertised from a specific source MUST NOT have more than one such Traffic Classes (Traffic Class with 0 element count). If there are more than one such Traffic Classes present then advertised SLA parameters MUST be ignored. It is okay though to have none Traffic Class with element count 0.

Classifier Element values in a Traffic Class (optional),

08-bit   = IPFIX Element Identifier
variable-length = based on type of the Element

Given IPFIX [RFC5102] has well defined identifier set for a large number of packet attributes, IPFIX IANA registry is "https://www.ietf.org/assignments/ipfix" chosen to specify packet classification attributes. However, since not all identifiers from IPFIX would be applicable to this proposal, only a limited set identified here can be supported by BGP SLA exchange. Any new element identifier, in future, added to the IPFIX IANA registry does not automatically mean supported for this proposal.

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   |
       +----+----------------------------+


    Traffic Class Service count (for a Traffic Class under definition)
        08-bit count of service attributes fields to follow with
               type/value pair
        List of service types and relevant values are discussed below

        00 = no bounded service (also means Best Effort)



    Traffic Class Service (optional),
        16-bit          = type of the field
        variable-length = based on type of the service


    - 0x00 = reserved


    - 0x01 = TRAFFIC_CLASS_TSPEC
      160-bits TSpec Parameter

      The TRAFFIC_CLASS_TSPEC parameter consists of the (r), (b), (p),
      (m) and (M) parameters as described in Invocation Information
      section of [RFC2212]. Note that inheriting definition of TSpec
      here does not enable RFC2212 functionality. It purely is the
      Traffic Specification that is inherited here for the purpose of
      SLA exchange.

       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
 |  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)         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  Minimum Policed Unit (m) (32-bit integer)                    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  Maximum Packet Size (M)  (32-bit integer)                    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Parameter (r) indicates min-rate of the traffic class. This rate indicates the minimum rate, measured in bytes of Layer 2 (L2) datagrams per second, 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 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 bytes 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 bytes 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.

Parameters (r), (b) and (p) are each set as 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].

The minimum policed unit (m) and maximum packet size (M) parameters have no relevance for the purpose of SLA exchange. Thus they MUST be ignored.

- 0x02, L2_OVERHEAD
  08-bit, value

  By default specification of rate and other packet size related parameters, advertised in an SLA, includes L2 overhead. For the receiver 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 consideration from the source perspective may be different from the local L2 overhead at the receiver. Explicit notification of size of L2 overhead from a sender, in such cases, is useful for a receiver to distinguish local L2 overhead from the sender advertised one. When receiver choose to react to an advertised SLA and if this service type is present in advertised SLA, receiver MUST use advertised L2 overhead over local L2 overhead.

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

- 0x03 = MINRATE_IN_PROFILE_MARKING
  08-bit   = IPFIX Element Identifier
  variable-length = based on type of the Element

  00 Identifier = drop, variable-length for this id is 0.

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

- 0x04 = MINRATE_OUT_PROFILE_MARKING
  08-bit   = IPFIX Element Identifier
  variable-length = based on type of the Element

  00 Identifier = drop, variable-length for this id is 0.

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

```
    - 0x05 = MAXRATE_IN_PROFILE_MARKING
      08-bit          = IPFIX Element Identifier
      variable-length = based on type of the Element

      00 Identifier = drop, variable-length for this id is 0.

       +----+----------------------------+
       | ID | Name                       |
       +----+----------------------------+
       |195 | ipDiffServCodePoint        |
       |203 | mplsTopLabelExp            |
       |244 | dot1qPriority              |
       +----+----------------------------+


    - 0x06 = MAXRATE_OUT_PROFILE_MARKING
      08-bit          = IPFIX Element Identifier
      variable-length = based on type of the Element

      00 Identifier = drop, variable-length for this id is 0.

       +----+----------------------------+
       | ID | Name                       |
       +----+----------------------------+
       |195 | ipDiffServCodePoint        |
       |203 | mplsTopLabelExp            |
       |244 | dot1qPriority              |
       +----+----------------------------+


      In the case when MINRATE_IN_PROFILE_MARKING,
      MINRATE_OUT_PROFILE_MARKING, MAXRATE_IN_PROFILE_MARKING and
      MAXRATE_OUT_PROFILE_MARKING all of them are advertised,
          - MINRATE_IN_PROFILE_MARKING takes highest precedence
            (that is over MAXRATE_IN_PROFILE_MARKING)

          - MAXRATE_IN_PROFILE_MARKING takes precedence over
            MINRATE_OUT_PROFILE_MARKING

          - and MAXRATE_OUT_PROFILE_MARKING takes precedence over
            MINRATE_OUT_PROFILE_MARKING


    - 0x07  = DROP_THRESHOLD
      03-bit count of drop-priority fields to follow with
               (type, type-value, burst size) tuple

      04-bit, drop priority type
```

```
08-bit          = IPFIX Element Identifier
variable-length = based on type of the Element
32-bit, Burst Size (32-bit IEEE floating point number)
```

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

This finer granular drop threshold does not require separate buffer space from the aggregate buffer space. It is just an indicator beyond which code-point specific traffic to be discarded when occupancy of aggregate buffers reached to that threshold.

- 0x08 = RELATIVE_PRIORITY
  04-bit, priority value
          lower the value, higher the priority

  Relative priority indicates scheduling priority. For example voice traffic, which requires lowest latency compare 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.

- 0x09 = SUB_TRAFFIC_CLASSES
  variable-length, repeats all content described above from Traffic
                   Class count onwards.

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

4. Originating SLA Notification

The QoS attribute to advertise SLA sub-type MUST be added by the originator of a BGP UPDATE message.

SLA messages SHOULD NOT be sent periodically just for the purpose of keep alive. 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 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). A BGP update message, in such cases, with source AS number and NLRI prefix of source end-point can uniquely identify physical/virtual link and so establishes advertised SLA's context 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 single link between them, CE can uniquely identify the forwarding link to PE with AS number of the PE and NLRI prefix being an IP address of PE, to CE (that is the next hop address from CE to PE). SLA advertised thru BGP update message from PE to CE, with PE's AS number and IP address, establishes SLA context for the aggregate traffic through link CE to PE. SLA advertised thru BGP update message from PE to CE, with PE's AS number and any other prefix establishes SLA for that specific prefix, subset of traffic under CE to PE link.

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

4.1.1. SLA Advertisement for Point-to-Point Connection

When SLA messages are intended to be advertised for the point-to-point connection (physical or logical), the message is destined for

the next hop and advertised message is in the context of the prefix of the source end-point of the point to point connection.

The destination AS number set to, within QoS SLA attribute, typically is of the neighbor BGP speaker's. Alternatively, the originator MAY not encode source/destination AS numbers (that is the source AS is set to 0 and destination AS count is set to 0), in the QoS attribute. The most significant bit of the QoS attribute flag MAY be set to 1, specifically it MUST be set to 1 when intention is to not install route update, at the receiver, for the advertised message.

4.1.2. SLA Advertisement for Destination AS Multiple Hops Away

When SLA messages are to be advertised beyond next hop, value of source AS, in the QoS attribute, MUST be set by the originator of the update message. If such update is meant to be for a specific list of AS(es) as receivers, then the list of destination AS MUST be explicitly described in the QoS attribute message to avoid flooding of the QoS attribute data in the network beyond those destinations.

When a new prefix is added in the AS, 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 BGP update for this new prefix may include QoS attribute containing just an SLA id, an id that was advertised earlier. The corresponding Update message does not require to have the whole SLA content. SLA id is sufficient to relate SLA parameters to new advertised prefix.

When BGP update messages are triggered as a result of SLA policy change and thus only for the purpose of SLA exchange, forwarding BGP update messages beyond intended receivers are not necessary. Highest order bit in the QoS Attribute flag MUST be set to suggest receiver to drop entire BGP update message [Note that it is an indication to drop entire update message, not only QoS attribute], after all intended receivers have processed it. If update message contains a list of destination ASes, then the message MUST be dropped only after all intended receivers (destinations) have received it.

5. SLA Attribute Handling at Forwarding Nodes

5.1. BGP Node Capable of Processing QoS Attribute

If a BGP node is capable of processing QoS attribute, it optionally MAY process the message. If advertised SLA has a list of destination

ASes, it MAY trim list and so count of destination AS to exclude ones that are not required in further announcement of BGP updates.

BGP node MUST drop SLA related sub-type from the QoS attribute, if none of the AS from the destination list is 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 QoS attribute all together. The other attributes and NLRI information may be announced further if they meet rules defined by other attributes and BGP protocol.

If the most significant bit in the QoS attribute flag is set to 1 then the entire BGP update message MUST be dropped if there are no destinations left in the list to advertise to.

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

5.2. BGP Node not Capable of Processing QoS Attribute

If the BGP node is not capable of processing QoS attribute, it MUST forward the QoS attribute message unaltered.

5.3. Aggregator

It is RECOMMENDED to not aggregate prefixes from 2 or more BGP update messages into one BGP update, when original messages contain the QoS attribute with SLA sub-type contents. If Aggregator MUST aggregate them then it MUST copy entire parameter set of an SLA sub-type from the QoS attribute in the new aggregated BGP update message. At the same time, it MUST also insert NLRI information, from the original update message, as an optional advertiser id to go along with source AS inside the QoS attribute.

To support SLA exchange multiple hops away in the path that has one of the forwarding node acting as an Aggregator, it is required that the Aggregator node is capable of processing the QoS attribute.

6. SLA Attribute Handling at Receiver

Reception of and processing of advertised QoS SLA content are optional for the receiver.

While reacting to SLA advertisement

- receiver SHOULD invalidate previous advertised SLA parameters if one exists for the same SLA id and source AS. If new advertised SLA update is with non-zero Traffic Class count, new advertised SLA SHOULD be installed. If new advertised SLA update is with Traffic Class count 0, no action is required.

- If advertised QoS Attribute, inside an update message, is with a flag set indicating to drop that message, a receiver MUST drop message if it is the last receiver, in update path, that message is advertised to.

If the advertised SLA is from the next hop, in the reverse path, the receiver may implement advertised SLA for the whole link, the link could be physical or virtual link, associated with the next hop. If NLRI advertised in update message is not of the next hop, 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 won't.

For cases where if earlier messages have not reached the intended receiver yet, a re-signaling is required. A receiver may intend to request an SLA message from the originator in such case. Since BGP messages are considered reliable, it is assumed that advertised messages always reach intended receivers. Thus discussion of REQUEST message, for this purpose or any other purpose, is considered out of the scope of this document.

To handle error conditions, the approach of "attribute-discard" as mentioned in [IDR-ERR] MAY be used in the event QOS attribute parsing results in any attribute errors. Alternatively, an approach of "treat-as-withdraw" MAY be used as mentioned in [IDR-ERR] if an implementation also wishes to withdraw the associated prefix.

6.1. Traffic Class Mapping

It is possible that switching/routing methods used in 2 different ASes could be different. For example, Provider may tunnel Customer's IP traffic thru MPLS cloud. In such cases traffic class definition for QoS services may differ in both ASes. For the meaningful use of advertised SLA in such cases, receiver is required to map traffic class from one type to the other.

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.

There are well-defined recommendations that exist for traffic class mapping between two technologies. 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.

7. Deployment Considerations

One of the use cases is for a Provider to advertise contracted SLA parameters to Customer Edge (CE). The SLA parameters are 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. 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. SLA advertise 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
```

Another use case can be to advertise SLAs among different network sites within one Enterprise network. In Hub and Spoke deployments, Administrator, being aware of each Spoke's SLA, may define SLAs for each of them at the Hub and advertise them thru BGP updates, where at each Spoke, advertised SLA may translate to a forwarding policy. In a scale network, managing a large number of Spokes can be complex. The proposal in such cases would be to provision SLA parameters at the Hub only and distribute them to each Spoke with SLA exchange protocol described here.

Alternatively, in a network that supports SLA parameters signaling capabilities with the Provider, manual administration can be avoided or minimized even at the Hub. As shown in the figure below, AS2 may first learn its SLA with the Provider from the Provider Edge it is connected to. AS2 can advertise the same or a subset of that SLA to AS3 in the context of tunnel's 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
```

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.

8. 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 for the review.

9. IANA Considerations

The proposal in this document defines a new BGP attribute. IANA maintains the list of existing BGP attribute types. A new type to be added in the list for the QoS attribute.

The proposal also defines a list for Service types associated to Traffic Class. IANA will be required to maintain this list for Traffic Class Service type as a new registry. Where-as Traffic Class Element types, defined in the proposal, refer to existing IPFIX IANA types.

```
      Proposed definition of Traffic Class Service Types
          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
```

10. Security Considerations

There is a potential for mis-behaved AS to advertise wrong SLA, stealing identity of another AS. This resembles to problems already identified and resolved, in the routing world, thru reverse path forwarding check. One proposal, inline to RPF, to resolve such threats is to have each BGP speaker node, in the forwarding path, perform reverse path check on source AS. Since we expect these messages to originate and distributed in the managed network, there should not be any risks for identity theft. Thus reverse path check is not considered in this proposal nor have we considered any alternates. Such solutions can be explored later if any such need.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2212] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private

Networks (VPNs)", RFC 4364, February 2006.

[RFC4506] Eisler, M., "XDR: External Data Representation Standard", STD 67, RFC 4506, May 2006.

[RFC5102] Quittek, J., Bryant, S., Claise, B., Aitken, P., and J. Meyer, "Information Model for IP Flow Information Export", RFC 5102, January 2008.

11.2. Informative References

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[IDR-ERR] Scudder, J., Chen, E., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Message, I-D.draft-ietf-idr-error-handling", June 2012.

[CPP] Boucadair, M., Jacquenet, C., and N. Wang, "IP/MPLS Connectivity Provisioning Profile, I-D.boucadair-connectivity-provisioning-profile", Sep 2012.

[CPNP] Boucadair, M. and C. Jacquenet, "Connectivity Provisioning Negotiation Protocol (CPNP), I-D.boucadair-connectivity-provisioning-protocol", May 2013.

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 Networks
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
France Telecom
Rennes 35000
France

Email: mohamed.boucadair@orange.com