Network Working Group S. Shah
Internet-Draft K. Patel
Intended status: Standards Track Cisco Systems
Expires: April 22, 2013 S. Bajaj
Juniper Networks
L. Tomotaki
Verizon
M. Boucadair
France Telecom
Oct 19, 2012

Inter-domain SLA Exchange
draft-svshah-interdomain-sla-exchange-03

Abstract

Network administrators typically provision QoS policies for their application traffic (such as voice, video) based on SLAs negotiated with their providers, and translate those SLAs to 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 operational transitive attribute to signal SLA details in-band, across administrative boundaries (considered as Autonomous Systems (AS)), and thus simplify/speed-up some of the complex tasks.

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

Table of Contents

1. Introduction

Typically there is a contractual Service Level Agreement (SLA) negotiated between Customer and Provider or between one Provider to another Provider [CPP]. This contractual agreement 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 full line-rate or sub rate for aggregate traffic. Or it could be even finer granular traffic distinction with services defined for standard code-points or for specific set of prefix or for set of well-known application types.

Once the SLA is negotiated, it needs to be translated into enforcing configuration data and policies on the Provider's Edge (PE) as well as on the Customer's Edge (CE). At the Customer, a person administering the CE device may be a different person, or even a different department, from the ones negotiating SLA contracts with the Provider and thus an administrator at the CE first requires to manually learn negotiated SLA, thru SLA documents or via some other off-band method. In a subsequent step an administrator requires to translate SLA to QoS policies using router (vendor) specific provisioning language. In a multi-vendor environment, translating the SLA into technology-specific configuration and then enforcing that 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. For an example for voice service, the Provider may negotiate service for such traffic thru EF code-point in Diffserv networks. Administrator at the CE not only will have to know that Provider's service for voice traffic is EF based but will also have to implement DSCP EF classification rule along with Low Latency Service rule as per vendor's provisioning language.

Given the Provider also maintains established contracts, which very well may even be enforced at the PE, an in-band method of signaling it from the PE to the CE can help eliminate manual administrative process described above. Provider may have SLA negotiated with the Customer via some defined off-band method. Once negotiated, the Provider may translate that SLA in networking language on the PE (this process remains same as is done today). This SLA instance then can be signaled to the CE via some in-band protocol exchange. In reaction to that message, receiver CE router may automatically translate that to relevant QoS policy definition on the box. This in-band signaling method helps eliminate manual complex process required by administrator at the CE. Taking same voice service as an

example, given Provider already may provision definition of EF code-point for such, signaling this code-point traffic class from PE to CE along with low latency service definition, omits administrator at the CE to worry about such translation.

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

- The most common use-case of SLA exchange is across Autonomous Systems. And BGP is the most suitable protocol for any inter-domain exchange
- There is no other suitable 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 a definition of a new BGP attribute to advertise/ learn SLA details in-band. The BGP attribute proposed, in this document, is intended to advertise SLA from one AS to a list of interested AS. QoS services advertised could be for the incoming traffic to the AS community, 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.

The aim with the signaling of this attribute, across administrative boundaries, is to help network administrators speed up and simplify QoS provisioning with automatic learning of SLAs and thus avoiding complexities and possible errors with manual learning.

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 QoS attribute open for extensions, in future, for other SLA specific requirements or even beyond SLA specific needs. For example, SLA Negotiation and Assurance is out of scope of this document which 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, in BGP, 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   | QoS Attr type |                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               |
    ~                                                               ~
    |                      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
```

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 data-base, when this is the last BGP receiver from the list of AS 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

3.1. SLA, QoS attribute sub-type, Definition

The value field of the QoS Attribute contains further TLVs, following 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                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.........................
```

subType - 8 bits

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

SLA sub-type specific value field details 1) sender and receiver(s) and 2) SLA parameters. SLA Parameters include SLA event type (such as Advertise, Request) and content 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. SLA sub-type specifics, from the QoS attribute,
    MUST be removed by the receiver in such case.

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 = broadcast

Destination AS list
32-bit destination AS number, this field is omitted if broadcast
....
.... [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. SLA identifier is not globally unique but it MUST be unique in the context of 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 SLA attribute 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 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 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 was 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 to have 0 elements Traffic Class definition last in the ordered list.If Advertised SLA does not have this Traffic Class last in the advertised list, receivers MUST re-order it, for the forwarding purpose, as the last Traffic Class, in the ordered list, from the source AS. It is MUST that advertisement from a specific source does not have more than one Traffic classes with element count 0. If there are more than one such Traffic Classes then advertised SLA MUST be ignored. It is okay for SLA message though to have none Traffic Class with element count 0.

```
   Classifier Element values in a Traffic Class (optional),

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

        Element Types (08-bit)
        0x00 = Invalid
        0x01 = Reserved
        0x02 = IP_DSCP,   (length = 06-bits, value = 0..63)
        0x03 = MPLS_TC,   (length = 03-bits, value = 0..7)
        0x04 = 802_1Q_COS,(length = 03-bits, value = 0..7)
        0x05 = 802_1Q_DEI,(length = 01-bit, value = 0..1)
        0x06 = PHB_ID,    (length = 12-bits, value = 0..4095)
        0x07 to 0xff = for future use




   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 = MINRATE
      04-bit, unit type
          0x00 = reserved
          0x04 = PERCENT
          0x05 = KBPS
          0x06 to 0x0f = for future use
      32-bit, value in unit kbps


    - 0x02 = MINRATE_BURST
      32-bit, value in bytes
```

```
- 0x03 = MINRATE_IN_PROFILE_MARKING
  04-bit, re-mark type
          0x00 = Invalid
          0x01 = Reserved
          0x02 = IP_DSCP
          0x03 = MPLS_TC
          0x04 = 802_1Q_COS
          0x05 = 802_1Q_DEI
          0x06 to 0x0f = for future use
  08-bit, value


- 0x04 = MINRATE_OUT_PROFILE_MARKING
  04-bit, re-mark type
          0x00 = Invalid
          0x01 = Reserved
          0x02 = IP_DSCP
          0x03 = MPLS_TC
          0x04 = 802_1Q_COS
          0x05 = 802_1Q_DEI
          0x06 to 0x0f = for future use
  08-bit, value


- 0x05 = MAXRATE
  04-bit, unit type
      0x00 = reserved
      0x04 = PERCENT
      0x05 = KBPS
      0x06 to 0x0f = for future use
  32-bit, value


- 0x06 = MAXRATE_BURST
  32-bit, value in bytes


- 0x07 = MAXRATE_IN_PROFILE_MARKING
  04-bit, re-mark type
          0x00 = Invalid
          0x01 = Reserved
          0x02 = IP_DSCP
          0x03 = MPLS_TC
          0x04 = 802_1Q_COS
          0x05 = 802_1Q_DEI
          0x06 to 0x0f = for future use
  08-bit, value
```

```
    - 0x08 = MAXRATE_OUT_PROFILE_MARKING
      04-bit, re-mark type
              0x00 = Invalid
              0x01 = DROP
              0x02 = IP_DSCP
              0x03 = MPLS_TC
              0x04 = 802_1Q_COS
              0x05 = 802_1Q_DEI
              0x06 to 0x0f = for future use
      08-bit, value

      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



    - 0x09  = DROP_THRESHOLD
      03-bit count of drop-priority fields to follow with
               (type,value, unit,value) tuple

      04-bit, drop priority type
              0x00 = Invalid
              0x01 = None
              0x02 = IP_DSCP
              0x03 = MPLS_EXP
              0x04 = 802_1Q_COS
              0x05 = 802_1Q_DEI
              0x06 to 0x0f = for future use
      08-bit, drop priority type value

      04-bit, unit type
          0x00 = reserved
          0x01 = TIME_US
          0x02 = PERCENT
          0x03 to 0x0f = for future use
      08-bit, drop threshold value as per unit type
```

- 0x0A = RELATIVE_PRIORITY
  04-bit, priority value
  lower the value, higher the priority

  Relative priority indicates scheduling priority. For example voice traffic, that requires lowest latency compare to any other traffic, will 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 scheduling perspective do not require to be distinguish with different priority. Relative priority for those classification groups may be advertised with the same value.

- 0x0B = 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 differentiated 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

QoS attribute to advertise SLA MUST be added by the originator of a BGP UPDATE message. Any BGP speaker in the forwarding path of a message MUST NOT insert QoS attribute for the same prefix.

SLA messages SHOULD NOT be sent periodically just for the purpose of keep alive. Since SLA changes are in-frequent, some sort of SLA policy change can be considered as a trigger for the advertisement.

For any SLA modification, originator MUST re-advertise entire SLA. There is no provision to advertise partial SLA. To invalidate previously advertised SLA, a message MUST be sent with new SLA advertisement with Traffic Class count as 0.

4.1. SLA Contexts

In certain cases, the advertisement may be to establish SLA for aggregate traffic on a point to point connection between a specific destination and a specific source. A point to point connection may be a physical link, connecting BGP peers, or may be a virtual link (like 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 aggregate traffic for that point to point link.

In the simplest case where PE and CE are directly connected via a physical link and have only single link between them, CE can uniquely identify forwarding link to PE with AS number of the PE and NLRI prefix being an address of PE, to CE (that is 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 that is subset of traffic under CE to PE link.

Even though this example is in the context of IP prefix, SLA exchange does not have to be limited to IPv4 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, originator MAY not encode source/destination AS numbers (that is source AS set to 0 and destination AS count 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 specific list of AS(es) as receiver then list of destination AS MUST be populated 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 has already been advertised before for other existing prefixes, then to advertise that new prefix to be part of earlier advertised SLA, a trigger of new BGP update message with QoS attribute containing SLA id is sufficient. Update message does not require to have whole SLA content.

When BGP update messages are triggered as a result of SLA policy change and so for the purpose of SLA exchange only, 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 list of destination of AS then 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 list of destination AS, 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. Rest of the QoS attributes message 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 existing in the QoS attribute message then node MUST drop QoS attribute all together. Rest other attributes and NLRI may be announced further if it meets rules defined by other attributes and BGP protocol.

If most significant bit in the QoS attribute flag is set to 1 then entire BGP update message MUST be dropped if there are no destination left in the list to advertise to. However, If SLA message is meant to be broadcast then message MUST not be dropped/trimmed.

Except extracting entire SLA sub-type of the QoS attribute, trimming the list of destination AS list and inserting NLRI at Aggregator

node, rest 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 BGP node is not capable of processing QoS attribute, it MUST forward attribute message as it is received.

5.3. Aggregator

It is RECOMMENDED to not aggregate prefixes from BGP update messages that contain QoS SLA attribute. If Aggregator MUST aggregate prefixes then it MUST copy QoS SLA attribute in new aggregated BGP update message. At the same time, it MUST also insert NLRI, from the original update message, as an optional advertiser id to go along with source AS in the QoS attribute.

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

6. SLA attribute handling at Receiver

Reception of and reaction to advertised messages are optional for the receiver.

As described in earlier section, while reacting to SLA advertisement

- receiver SHOULD invalidate previous advertised SLA and then if one exists for advertised NLRI. 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 is with flag set to indicate to drop this message, receiver MUST drop message if it is the last receiver, in the update path, this message is advertised to.

If advertised SLA is from the next hop, in reverse path, the receiver can establish 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 to accept advertised SLA and for which ones to not.

For cases where if earlier message has not yet reached to the intended receiver, a re-signaling is required. A signaling event

REQUEST is required, for this purpose, to be triggered by intended receiver. Since BGP messages are considered reliable, discussion of REQUEST, 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 an event if a 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 common that switching/routing technologies used in 2 different AS could be different. For example, Provider may tunnel Customer's IP traffic thru MPLS cloud. In such cases traffic class definition for QoS services is also different in both AS. For the meaningful use of advertised SLA in such cases, receiver is required to map traffic class from one type to another.

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 from PE, CE 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 Consideration

Typical use-case aimed with this proposal is for Provider to advertise contracted SLA to Customer Edge. SLA established between customer and Provider is provisioned by the provider on the PE device (facing Customer Edge). This provisioning, in a form supported by Provider, is advertised thru proposed BGP QoS attribute to the Customer Edge. Customer may read thru advertised SLA to provision one on the Customer Edge link facing towards PE.

Contracted SLA from PE to CE may be full line-rate or sub-rate of a link or finer granular controlled services. SLA is not required to be advertised if the SLA contract is simply a physical link. SLA advertise can be useful when contracted service is sub-rate of a link

and/or if for finer granular traffic classes that are controlled. Like 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 SLA among different network sites within one Enterprise network. In Hub and Spoke deployments, Hub may define SLA for individual spokes and advertise this SLA thru BGP updates.

```
                                                       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 = AS2 tunnel address
```

It very well could be possible that AS2 may first learn its SLA with Provider from Provider Edge it is connected to and then advertises

same or subset of the SLA to AS3 with AS2 to AS3 tunnel's ip address as NLRI.

Deployment options are not limited to involving CEs only. For any contract between Provider to Provider, SLA may be advertised from one PE to another PE also.

8. Acknowledgements

Thanks to Fred Baker for his suggestions and to Ken Briley, Rahul Patel, Fred Yip, Lou Berger and Brian Carpenter for the review. Thanks to Bertrand Duvivier for his valuable contributions to help make subsequent revision better.

9. IANA Considerations

This document defines a new BGP attribute. IANA maintains the list of existing BGP attribute types. Proposal is to define a new attribute type code for the QoS attribute.

With the proposal, there is a list defined for Traffic Class Elements type and associated Service types. IANA will be required to maintain list of both new types.

```
Proposed definition of Traffic Class Element Types
     0x00 = Invalid
     0x01 = Reserved
     0x02 = IP_DSCP,    (length = 06-bits, value = 0..63)
     0x03 = MPLS_TC,    (length = 03-bits, value = 0..7)
     0x04 = 802_1Q_COS,(length = 03-bits, value = 0..7)
     0x05 = 802_1Q_DEI,(length = 01-bit, value = 0..1)
     0x06 = PHB_ID,     (length = 12-bits, value = 0..4095)

Proposed definition of Traffic Class Service Types
    0x00 = reserved
    0x01 = MINRATE
    0x02 = MINRATE_BURST
    0x03 = MINRATE_IN_PROFILE_MARKING
    0x04 = MINRATE_OUT_PROFILE_MARKING
    0x05 = MAXRATE
    0x06 = MAXRATE_BURST
    0x07 = MAXRATE_IN_PROFILE_MARKING
    0x08 = MAXRATE_OUT_PROFILE_MARKING
    0x09 = DROP_THRESHOLD
    0x0A = RELATIVE_PRIORITY
    0x0B = SUB_TRAFFIC_CLASSES


Proposed definition of Unit Types
    0x00 = reserved
    0x01 = TIME_US
    0x02 = PERCENT
    0x03 = KBPS
```

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

[RFC1771] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

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

[RFC3140] Black, D., Brim, S., Carpenter, B., and F. Le Faucheur, "Per Hop Behavior Identification Codes", RFC 3140, June 2001.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.

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

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[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.

11.2. Informative References

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

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