

IDR
Internet-Draft
Intended status: Standards Track
Expires: August 2, 2018

S. Shah

K. Patel
Arrcus, Inc
S. Bajaj
Viptela
L. Tomotaki
Verizon
M. Boucadair
Orange
January 29, 2018

Inter-domain Traffic Conditioning Agreement (TCA) Exchange Attribute
draft-ietf-idr-sla-exchange-13.txt

Abstract

Network administrators typically enforce Quality of Service (QoS) policies according to Traffic Conditioning Agreement (TCA) with their providers. The enforcement of such policies often relies upon vendor-specific configuration language. Both learning of TCA, either thru TCA 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 TCA 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 <https://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 2, 2018.

Internet-Draft

Inter-domain TCA Exchange

January 2018

Copyright Notice

Copyright (c) 2018 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 (<https://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	5
3.1.	QoS Attribute SubType	6
3.2.	TCA SubType	7
3.3.	TCA Content for ADVERTISE TCA Event	9
3.3.1.	Supported IPFIX identifiers for Traffic Class Elements	12
3.3.2.	Traffic Class Service types and respective TLVs	15
4.	Originating TCA Notification	22
4.1.	TCA Contexts	23
4.1.1.	TCA Advertisement for Point-to-Point Connection	23
4.1.2.	TCA Advertisement for Destination AS Multiple Hops Away	24
5.	QoS Attribute Handling at Forwarding Nodes	24
5.1.	BGP Node Capable of Processing QoS Attribute	24
5.2.	QoS Attribute Handling at Receiver	25
6.	Error Handling	25
7.	Deployment Considerations	25
8.	IANA Considerations	27
9.	Security Considerations	28
10.	Acknowledgements	29
11.	References	29
11.1.	Normative References	29
11.2.	Informative References	30
	Authors' Addresses	31

Internet-Draft

Inter-domain TCA Exchange

January 2018

1. Introduction

Typically there is a contractual Traffic Conditioning Agreement (TCA) for Quality of Service (QoS) established between a customer and a provider or between providers [[RFC7297](#)]. This QoS TCA 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 TCA is established, QoS TCA 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 TCA to QoS policies using router (vendor) specific provisioning language. In a multi-vendor network, translating TCAs 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 TCA agreements into technical clauses and configurations and thus both the steps of out of band learning of negotiated TCA and provisioning them in a vendor specific language can be complex and error-prone.

TCA parameters may have to be exchanged through organizational boundaries, thru TCA 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 TCA policy described in this document does not require any specific method for the provider in establishing TCAs. It only requires that the provider wishes to send the QoS TCA policy via BGP UPDATE [[RFC4271](#)] messages from the provider to a set of receivers (BGP peers). In reaction to, a receiving router may translate that to relevant QoS policy definition on the device. The TCA negotiation and assurance is outside the scope of this document.

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

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 [[RFC8040](#)] can set these standardize in configuration mechanisms. For ephemeral state, the I2RS protocol is being developed to set ephemeral state. While these protocols 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", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

This document makes use of the following terms:

- o BGP Speaker: A functional component on a BGP capable device that

functions as per BGP specification.

- o BGP peers: BGP Speakers adjacent to each other.
- o QoS Attribute Speaker: A functional component on a BGP capable device that produces and/or processes content of the QoS Attribute. A device that is QoS Attribute Speaker is also always a BGP Speaker. However, a BGP Speaker not necessarily always a QoS Attribute Speaker.
- o QoS Attribute content is produced and processed outside the function of the BGP Speaker and thus content of the QoS Attribute is completely opaque to the BGP Speaker. At BGP capable device where QoS Attribute content is produced, length and value of the QoS Attribute is passed from QoS Attribute Speaker to the BGP Speaker where BGP Speaker inserts the attribute into the BGP UPDATE message with appropriate attribute flags, attribute type, and length and value passed from the QoS Attribute Speaker. Similarly, a BGP capable device when receives QoS Attribute in the

BGP UPDATE message, BGP Speaker extracts QoS Attribute value from the message and passes it to the QoS Attribute Speaker where QoS Attribute Speaker processes the content from that passed down value. How the content of the QoS Attribute is passed from the QoS Attribute Speaker to the BGP Speaker and vice versa is implementation specific.

In the context of use of QoS Attribute for TCA parameters exchange, following roles are defined further within the scope of the QoS Attribute Speaker.

- o TCA Producer: This is a QoS Attribute Speaker that produces QoS Attribute for the TCA SubType.
- o TCA Consumer: This is a QoS Attribute Speaker that is intended receiver of QoS Attribute with the TCA SubType.

[3.](#) QoS Attribute Definition

The QoS Attribute is an optional transitive attribute (TBD - attribute code to be assigned by IANA) which is applicable to the Source AS and NLRIs advertised in the BGP UPDATE message this

attribute is included in. The format of the QoS Attribute is shown in Figure 1.

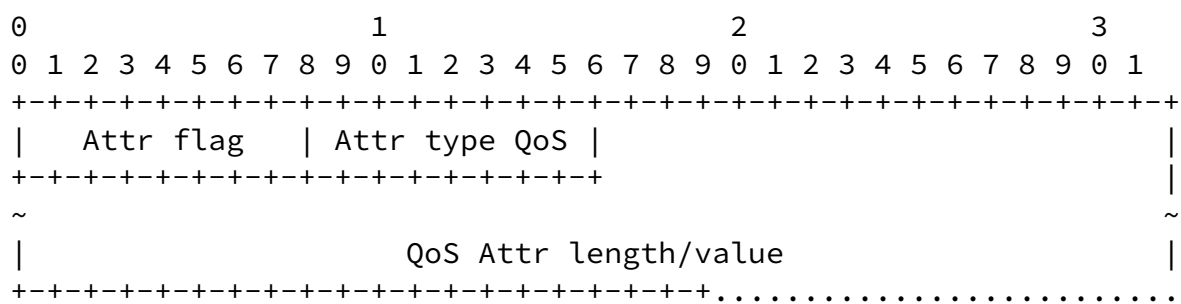


Figure 1: QoS attribute

Attribute flags - 8-bits field

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

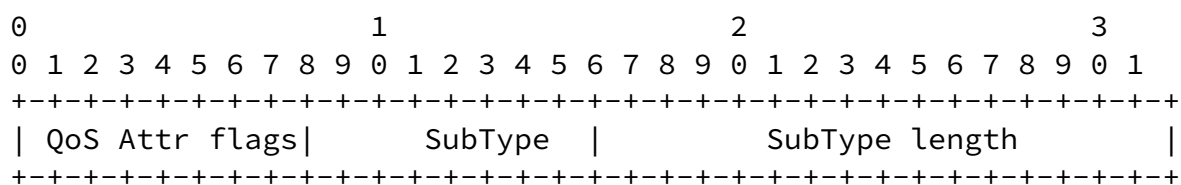
The content of the QoS Attribute is further specified with flags, applicable to QoS Attribute content, and a SubType in a TLV form.

3.1. QoS Attribute SubType

The Value field of the QoS Attribute contains the following:

QoS Attribute flags

Tuple (SubType of the QoS Attribute, SubType length, SubType value)



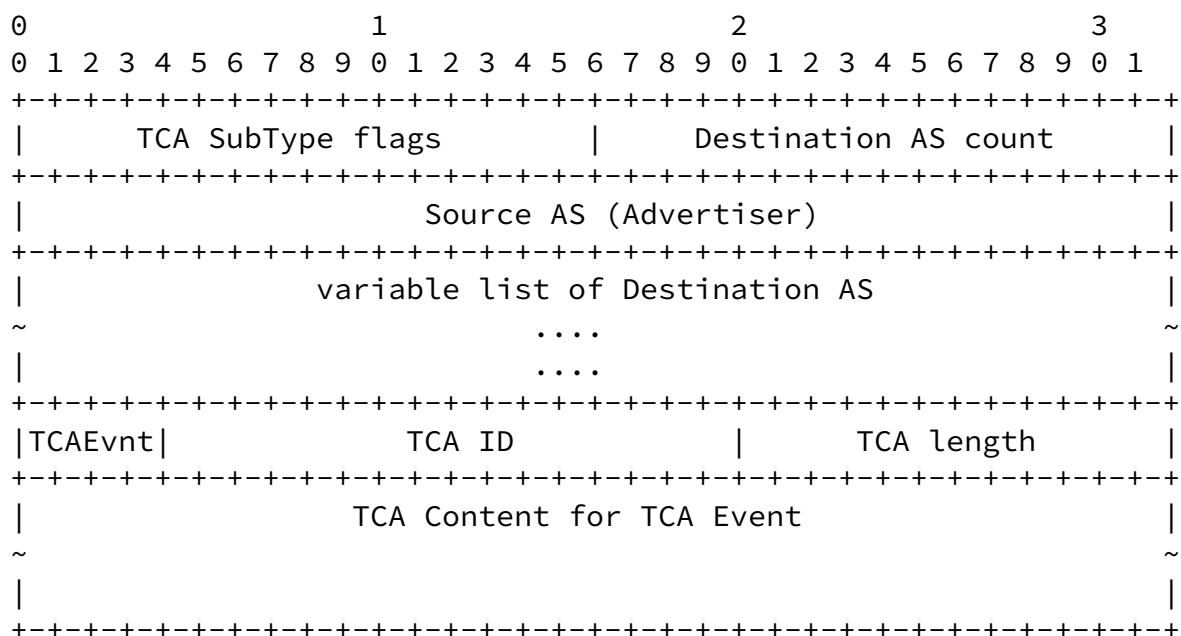


Figure 3: Format of the TCA SubType of the QoS attribute

TCA SubType flags - 16-bits field

Currently un-used. All bits in this field MUST be set to 0. The field is defined for the future use.

Destination AS count - 16-bits field that specifies count of destination ASNs present in the Destination AS list.

If this count is 0 then that is an error condition which should be handled as described in [Section 6](#).

Source AS - 32-bits field AS number space as defined in [[RFC6793](#)]

This is the AS where TCA Content is originated from. The Source AS MUST be of the same AS that is originating TCA ID and TCA Content.

The Source AS value of 0 is illegal and thus should be considered

an error which should be handled as described in [Section 6](#).

Destination AS list - variable length field that holds as many ASN. identifiers, each is 32-bits AS number space is defined in [[RFC6793](#)], as specified in the Destination AS count field.

List of ASNs for which the TCA is relevant to, each of which is a 32-bit number.

TCA Event - 4-bits field with following values:

0x0 = reserved

0x1 = ADVERTISE

0x2 to 0xf = Reserved for future use

The only TCA Event defined in this specification is ADVERTISE.

TCA ID - 16-bits field that specifies identifier which is unique in the scope of Source AS.

The significance of a TCA ID is in the context of the source that is advertising TCA Content. The TCA ID is not globally unique but it MUST be unique within the source AS.

The TCA ID applies to aggregate traffic to prefixes for a given AFI/SAFI that share the same Source AS and TCA ID.

TCA Length - 12-bits field that specifies the length of the TCA Content. The length is expressed in octets. The TCA Content is optional for an advertised TCA ID. If the TCA Content need not be there, the TCA length field MUST be set to zero in such a case.

TCA Content - A variable length field (optional field)

The TCA Content field contains TCA parameters relevant to specified TCA SubType. Since the only defined TCA SubType is ADVERTISE, this specification describes TCA Content only for the ADVERTISE TCA Event.

If TCA Content field exists in a BGP UPDATE message that contains the QoS Attribute with a TCA SubType for TCA Event ADVERTISE, format of the TCA Content is as described in [Section 3.3](#).

If the TCA Content field does not exist, then the advertised message refers to TCA Content advertised in the previous message for the same TCA ID. If there does not exist any prior TCA Content to relate to the advertised TCA ID, then receiver, TCA Consumer, can ignore the TCA advertisement and it may simply update Destination AS count and Destination AS list.

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

3.3. TCA Content for ADVERTISE TCA Event

The only TCA Event described in this specification is ADVERTISE. The format of TCA Content for the ADVERTISE Event is shown in Figure 4.

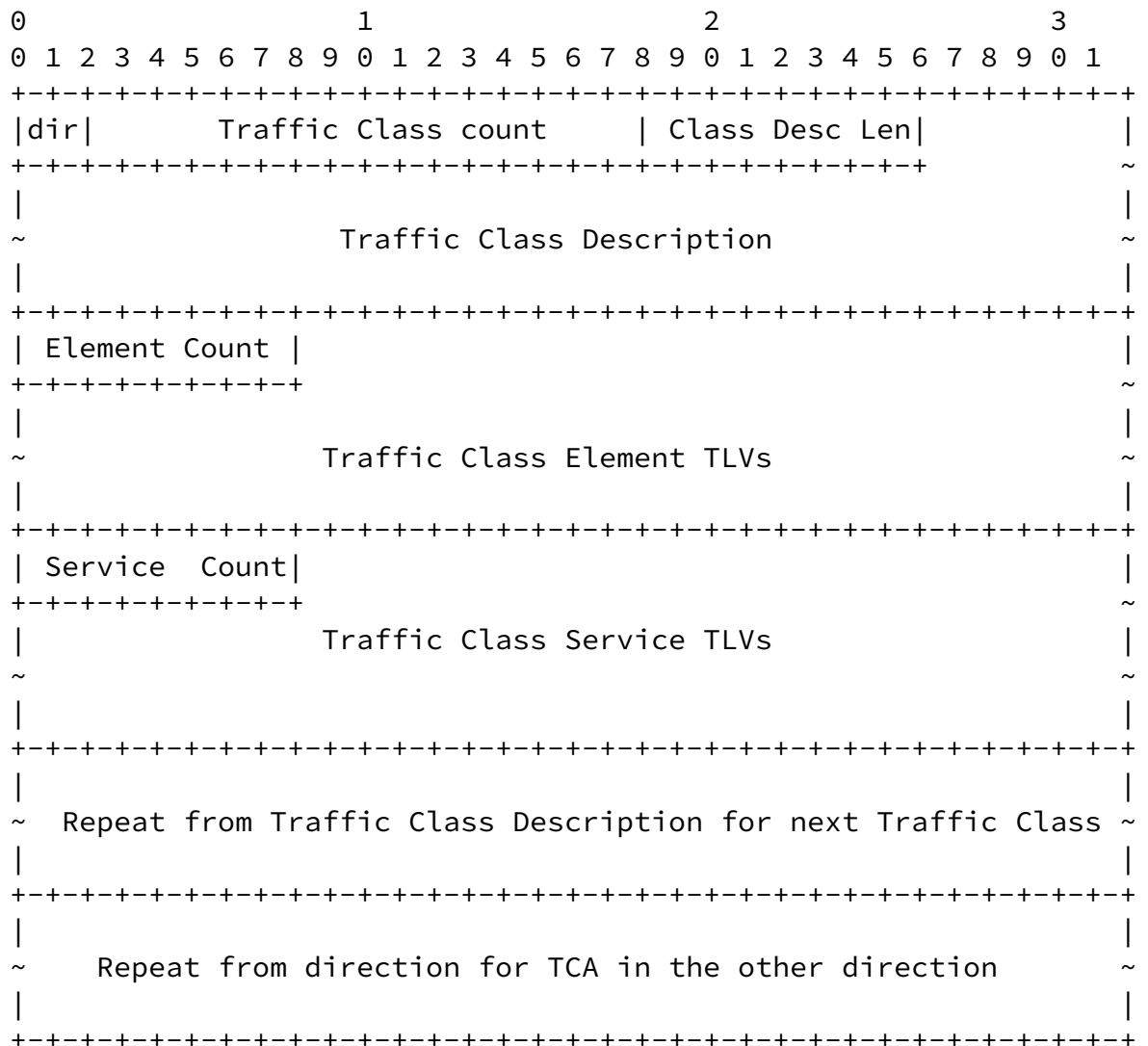


Figure 4: TCA-Content for ADVERTISE TCA Event

TCA Content contains Traffic Class TLVs that is a set of Traffic Class Elements (Classifiers) and Traffic Class Service TLVs for a list of Traffic Classes specified by Traffic Class count. This Traffic Class TLVs MUST be specified for one direction first and then optionally followed by the specification for the other direction.

dir (Direction) - 2-bits field that specifies Direction of the traffic TCA is applicable to. The following values are defined:

0x0 = reserved

0x1 = incoming, traffic to source AS from destination AS

0x2 = outgoing, traffic from source AS towards destination AS

0x3 = for future use

Traffic Class (Classifier Group) count - 16 bits field that specifies number of Traffic Classes.

The value of zero (0x00) in this field is a special value which means no TCA for the traffic in a specified direction. When Traffic Class count is 0, for a specific direction, the rest of the TCA Content fields MUST NOT be encoded, for that specific direction.

Traffic Class Description Len - 8-bits field that specifies the length of the Traffic Class Description field. The length is expressed in octets.

The value of zero in this field indicates that no Traffic Class Description field follows.

Traffic Class Description - variable length field, as expressed in The Traffic Class Description Len field, MUST carry UTF-8 encoded ([[RFC3629](#)]) description.

Traffic Class Elements (Classifier) Count - 8-bits field that specifies the count of Traffic Class Elements.

The value zero (0x00) means there are no Traffic Class 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 specified direction.

Traffic Class that has 0 elements MUST be presented last in the advertised list of Traffic Classes for a specific Direction.

Otherwise it is considered an error condition which should be handled as described in [Section 6](#).

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 be handled as described in [Section 6](#).

Traffic Class Element TLVs - (optional) variable length field holding as many TLVs specified by the Traffic Class Elements Count field. Each TLV has the following format:

IPFIX Element Identifier - 8-bits field that specifies IPFIX Identifiers listed in Table 1.

Length of Value field - 8-bits field that specifies the length, expressed in octets, of the value field.

Value - A variable field that specifies a value appropriate for the IPFIX Element Identifier. It is an error, if the value field does not contain the appropriate format, which should be handled as described 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 advertised AFI/SAFI NLRI within the BGP UPDATE message the QoS Attribute is contained in. If a receiver, TCA Consumer, receives a BGP UPDATE message with QoS Attribute for an unsupported AFI/SAFI then TCA Consumer MAY ignore advertised TCA. TCA Consumer MAY update only Destination AS count and Destination AS list, and then QoS Attribute and rest of the BGP UPDATE message

MUST be forwarded as per QoS Attribute and BGP protocol specification.

Traffic Class Service Count - 8-bits field that specifies count of Traffic Class Service TLVs.

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-bits field that specifies a service type. Each service type is detailed in [Section 3.3.2](#). The list of available service types are,

Shah, et al.

Expires August 2, 2018

[Page 11]

Internet-Draft

Inter-domain TCA Exchange

January 2018

0x00 = reserved

0x01 = COMMITTED_TSPEC

0x02 = PEAK_TSPEC

0x03 = COMMITTED_IN_PROFILE_MARKING

0x04 = COMMITTED_OUT_PROFILE_MARKING

0x05 = PEAK_OUT_PROFILE_MARKING

0x06 = DROP_THRESHOLD

0x07 = RELATIVE_PRIORITY

0x08 = EFFECTIVE_MAX_RATE

Length of Value field - 08-bits field that specifies the length of the value field. The length of the value is expressed in octets.

Value - a variable length field that specifies the value appropriate for each of the Service Types. It is an error, if this field does not contain the appropriate format, which should be handled as described in [Section 6](#). The format of the value for

each of the service types is described in [Section 3.3.2](#)

3.3.1. Supported IPFIX identifiers for Traffic Class 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/ipfix.xml#ipfix-information-elements>> ipfix.xml#ipfix-information-elements). Only the IPFIX attributes listed in Table 1 are supported. Any new attribute to be supported by TCA SubType MUST be a Standards Action as described in IANA section.

ID	Name	Context
195	ipDiffServCodePoint	Indicates the value of the marking used in the link(s) between the TCA Consumer and TCA Producer domains.
203	mplsTopLabelExp	Indicates the value of the marking used in the link(s) between the TCA Consumer and

		TCA Producer domains.
244	dot1qPriority	Indicates the value of the marking used in the link(s) between the TCA Consumer and TCA Producer domains.
8	sourceIPv4Address	Indicates the source IPv4 address of an aggregate traffic over a connection subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
27	sourceIPv6Address	Indicates the source IPv6 address of an aggregate traffic over a connection

		subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
9	sourceIPv4PrefixLength	Indicates the length of the source IPv4 prefix.
29	sourceIPv6PrefixLength	Indicates the length of the source IPv6 prefix.
44	sourceIPv4Prefix	Indicates the source IPv4 prefix of an aggregate traffic over a connection subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
170	sourceIPv6Prefix	Indicates the source IPv6 prefix of an aggregate traffic over a connection subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
12	destinationIPv4Address	Indicates the destination IPv4 address of an aggregate traffic over a connection

		subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
28	destinationIPv6Address	Indicates the destination IPv6 address of an aggregate traffic over a connection subject to a TCA; the direction is being explicitly indicated in the ADVERTISE

		Event message.
13	destinationIPv4PrefixLength	Indicates the length of the destination IPv4 prefix.
30	destinationIPv6PrefixLength	Indicates the length of the destination IPv6 prefix.
45	destinationIPv4Prefix	Indicates the destination IPv4 prefix of an aggregate traffic over a connection subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
169	destinationIPv6Prefix	Indicates the destination IPv6 prefix of an aggregate traffic over a connection subject to a TCA; the direction is being explicitly indicated in the ADVERTISE Event message.
4	protocolIdentifier	Indicates whether any or a specific protocol for the traffic class.
7	sourceTransportPort	This parameter is used only for protocols with port identifiers. It indicates the source port number for the transport protocol identified by "protocolIdentifier".
11	destinationTransportPort	This parameter is used only for protocols with port

	identifiers. It indicates the destination port number for the transport protocol identified by
--	--

		"protocolIdentifier".
+	+	+

Table 1

3.3.2. Traffic Class Service types and respective TLVs

3.3.2.1. COMMITTED TSPEC

The COMMITTED TSPEC TLV definition:

Type - 0x01

Length - 8-bits field that specifies length, expressed in octets, of the value field. The length of the value field MUST be specified to be 8 octets to hold the value defined as per format below.

Value - COMMITTED_TSPEC value consists of the (r), (b) parameters as described in Invocation Information section of [RFC2212] and shown in Figure 5. Note that inheriting the definition of TSPEC (Traffic SPECification) here does not enable RFC2212 functionality. Only the format of the Traffic Specification is used in this specification.

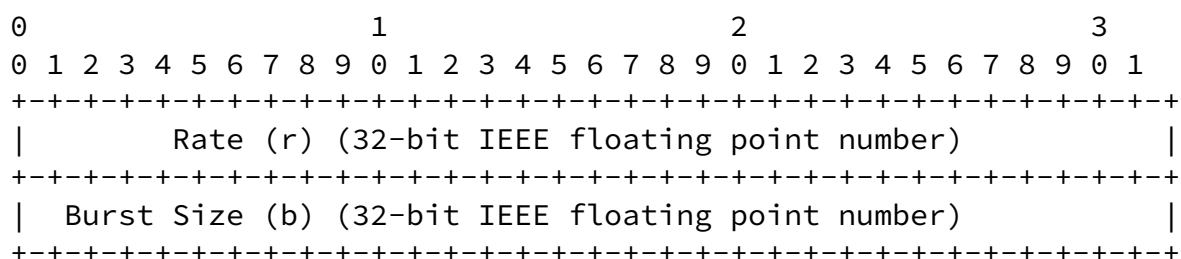


Figure 5: Traffic Class COMMITTED TSPEC

Format of Parameters (r) and (b): 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 committed-rate of the traffic class. This

rate indicates the minimum rate, measured in octets of IP datagrams per second (a.k.a, bytes 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 a TCA unless the traffic class definition clearly represents a sole receiver of a TCA.

Parameter (b): indicates maximum burst size, measured in octets of IP datagram size. Since queuing delay can be considered a function of burst size (b) and committed-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 TCA 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.

3.3.2.2. PEAK_TSPEC

The PEAK_TSPEC TLV definition:

Type - 0x01

Length - 8-bits field that specifies length, expressed in octets, of the value field. The length of the value field MUST be specified to be 8 octets to hold the value defined as per format below.

Value - PEAK_TSPEC value consists of the (r), (b) parameters as described in Invocation Information section of [\[RFC2212\]](#) and shown in Figure 5. Note that inheriting the definition of TSPEC (Traffic SPECification) here does not enable [RFC2212](#) functionality. Only the format of the Traffic Specification is used in this specification.

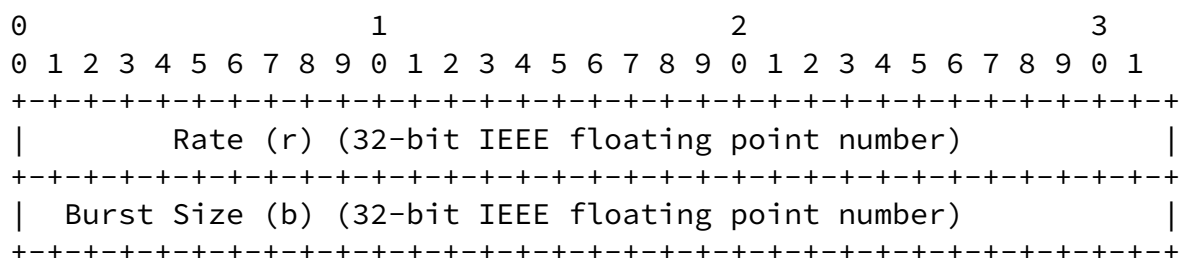


Figure 6: Traffic Class PEAK_TSPEC

Format of Parameters (r) and (b): 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 peak-rate of the traffic class. This rate indicates the maximum rate, measured in octets of IP datagrams per second (a.k.a, bytes per second), that the service advertiser is providing for a given class of traffic on advertiser's hop.

Parameter (b): indicates maximum burst size, measured in octets of IP datagram size.

When PEAK_TSPEC TLV is advertised, COMMITTED_TSPEC TLV MUST be present in the advertisement. Advertisement of PEAK_TSPEC TLV without COMMITTED_TSPEC TLV MUST be considered an error condition which should be handled as described in [Section 6](#). If committed-rate of the TCA is 0 then rate advertised in the COMMITTED_TSPEC shall be 0. Note that existence of COMMITTED_TSPEC in TCA advertisement is not mandatory nor is it a mandate that COMMITTED_TSPEC and PEAK_TSPEC must always go together. COMMITTED_TSPEC TLV is optional but only when there is no PEAK_TSPEC TLV present in the advertised TCA.

PEAK_TSPEC TLV with rate value of 0 MUST be considered an error condition which should be handled as described in [Section 6](#).

[3.3.2.3](#). COMMITTED_IN_PROFILE_MARKING

This Traffic Class Service Type defines action performed, by the TCA Producer, on packets that are compliant to the committed-rate specified in the COMMITTED_TSPEC TLV. If committed-rate specified in the COMMITTED_TSPEC TLV is 0 then TLV for this Traffic Class Service Type SHOULD NOT be advertised. COMMITTED_IN_PROFILE_MARKING TLV SHOULD be ignored by the TCA Consumer if there does not exist COMMITTED_TSPEC TLV for the specified direction, or committed-rate specified in the COMMITTED_TSPEC TLV is 0.

The COMMITTED_IN_PROFILE_MARKING TLV definition:

Type - 0x03

Length - 8-bits field that specifies length, expressed in octets, of the value field. The length of the value field MUST be specified to be 2 octets to hold the value defined as per format below.

Value - contains the Marking code-point type and value

Shah, et al.

Expires August 2, 2018

[Page 17]

Internet-Draft

Inter-domain TCA Exchange

January 2018

Marking code-point type - 8-bits IPFIX Element Identifier.

Marking code-point value - 8-bits code-point number.

The marking code-point type of 0x00 is a drop identifier. When marking code-point type value is 0x00 (that is drop), the marking code-point value in this case has no meaning and thus the value in this field should be ignored.

The following table lists the supported IPFIX Identifiers. Any value other than 0 or identifier from the following table is an error condition which should be handled as described in [Section 6](#).

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 2

[3.3.2.4](#). COMMITTED_OUT_PROFILE_MARKING

This Traffic Class Service Type defines action performed, at the TCA Producer, on packets that are not compliant to the committed-rate specified in the COMMITTED_TSPEC TLV, and compliant to rate specified in the PEAK_TSPEC TLV if PEAK_TSPEC TLV exists.

The COMMITTED_OUT_PROFILE_MARKING TLV definition:

Type - 0x04

Length - 8-bits field that specifies length, expressed in octets, of the value field. The length of the value field MUST be specified to be 2 octets to hold the value defined as per format below.

Value - contains the Marking code-point type and value

Marking code-point type - 8-bits IPFIX Element Identifier

Marking code-point value - 8-bits code-point number

The marking code-point type of 0x00 is a drop identifier. When marking code-point type value is 0x00 (that is drop), the marking code-point value in this case has no meaning and thus the value in this field should be ignored.

Table 2 lists the supported IPFIX Identifiers. Any value other than 0 or identifier from the Table 2 is an error condition which should be handled as described in [Section 6](#).

[3.3.2.5](#). PEAK_OUT_PROFILE_MARKING

This Traffic Class Service Type defines action performed, at the TCA Producer, on packets that are not compliant to the max-rate specified in the PEAK_TSPEC TLV. PEAK_OUT_PROFILE_MARKING TLV SHOULD be ignored by the TCA Consumer if there does not exist PEAK_TSPEC TLV for the specified direction.

The PEAK_OUT_PROFILE_MARKING TLV definition:

Type - 0x06

Length - 8-bits field that specifies length, expressed in octets, of the value field. The length of the value field MUST be specified to be 2 octets to hold the value defined as per format below.

Value - contains the Marking code-point type and value

Marking code-point type - 8-bits IPFIX Element Identifier

Marking code-point value - 8-bits code-point number

The marking code-point type of 0x00 is a drop identifier. When marking code-point type value is 0x00 (that is drop), the marking code-point value in this case has no meaning and thus the value in this field should be ignored.

Table 2 lists the supported IPFIX Identifiers. Any value other than 0 or identifier from the Table 2 is an error condition which should be handled as described in [Section 6](#).

[3.3.2.6](#). DROP_THRESHOLD

The DROP_THRESHOLD TLV definition:

Shah, et al.

Expires August 2, 2018

[Page 19]

Internet-Draft

Inter-domain TCA Exchange

January 2018

Type - 0x07

Length - 8-bits field that specifies length, expressed in octets, of the value field.

Value - Count of drop thresholds, followed by content for each drop threshold in the form of (code-point type, count of code-points, list of code-points, threshold value).

Count of drop thresholds - 8-bits field that specifies number of drop thresholds specified in this TLV. Content of each drop threshold is to follow following format

Code-point type - 8-bits IPFIX Element Identifier from the list shown in Table 6.

Count of code-points - 8-bits field that specifies number of code-point values to follow for a specified code-point type.

List of code-points - each code-point value is specified in size of 8 bits and thus total size for this field is 8 bits multiplied by as many number of code-points specified.

Burst value - This is a fixed size 32-bits IEEE floating point number that specifies burst value in unit of bytes.

All advertised drop thresholds, for a specific traffic class, are applicable to a single queue associated with that traffic class. A threshold for a set of code-points is a logical marker where an arrived packet is to be dropped if overall depth of a queue is beyond a threshold of a code-point set a packet is classified into. Choice of dropping discipline is implementation specific. If a packet can not be classified into any of the advertised code-point set then that means the TCA Producer is not defining any specific dropping behavior and thus dropping behavior is subject to implementation specific of the TCA Consumer.

ID	Name
195	ipDiffServCodePoint
203	mplsTopLabelExp
244	dot1qPriority

Table 3

[3.3.2.7](#). RELATIVE_PRIORITY

The RELATIVE_PRIORITY TLV definition:

Type - 0x08

Length - 8-bits field that specifies length, expressed in octets, of the value field. Given supported range of priority values in this specification, the length of the value field MUST be limited to and thus MUST be specified exactly as 1 octet.

Value - A value from range of 0 - 255. 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 classification 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 should be advertised with the same value.

A higher priority class of traffic to be served without pre-empted by lower priority class of traffic for more than a packet time at the configured rate.

For a system that implements WRR only (i.e., no priority queuing), it is possible to use a hierarchical WRR scheduling to achieve a behavior close to priority queueing where a root scheduling node has two child nodes. One child node is a queue assigned with a maximum possible value of a weight and advertised rate of highest priority Traffic Class as output bandwidth. The other child node is a scheduling node serving group of rest other advertised Traffic Classes (in the form of queues or yet another level of hierarchical WRR scheduler). Note that implementation specifics are out of the scope of this specification and this is an example to highlight how relative priority attribute can be relevant and treated by a system that implements only WRR. A system may choose to implement alternate methods to achieve a similar behavior.

[3.3.2.8](#). EFFECTIVE_MAX_RATE

The EFFECTIVE_MAX_RATE TLV definition:

Type - 0x02

Length - 8-bits field that specifies length, expressed in octets, of the value field. The length of the value field MUST be specified to be 5 octets to hold the value defined as per format below.

Value - Contains value of rate and per packet overhead

Aggregate max rate - 32-bits IEEE floating point number

Per packet overhead - 8-bits specifying value of overhead octets

Aggregate max rate indicates rate measured based on combined octets of packet's IP datagram size and advertised per packet overhead.

A packet traversing from the TCA Producer to the TCA Consumer or vice-versa may see packet overhead, additional octets on top of IP datagram size, difference between the Producer and the Consumer sent or received over a physical link. In cases, where advertised TCA is for a Consumer where total traffic between Consumer and Producer is to be capped to a specific sub-rate of a physical link, due to packet overhead differences between Producer and Consumer, sum of traffic from each TRAFFIC CLASS may overrun that total cap causing undesired behavior. In such cases, Producer can explicitly notify this TLV in advertised TCA.

[4.](#) Originating TCA Notification

The QoS Attribute for the TCA SubType MUST only be added to the BGP UPDATE message at the node that is TCA Producer. Any QoS Attribute Speaker, in the path to the TCA Consumer MUST NOT modify content of that attribute except modification of the Destination AS list.

QoS Attribute with the TCA SubType SHOULD NOT be advertised periodically just for the purpose of KEEPALIVE between TCA Producer and TCA Consumer. Some sort of TCA policy change, at the TCA Producer, may be considered as a trigger for the advertisement.

For any modified TCA policy at the TCA Producer, the TCA Producer MUST re-advertise the entire set of TCA parameters. There is no provision to advertise partial set of TCA parameters. Announcing a TCA ID different from an earlier advertised one, for the same prefix and from the same Source AS, indicates Source AS is advertising new TCA Content to replace the previous one advertised with the same TCA ID.

In order to withdraw a given TCA between TCA Producer and TCA Consumer, the TCA Produced MUST sent TCA Content with the same TCA ID, AS Source, and NLRI prefix, as were used to advertise earlier TCA parameters, and the Traffic Class count MUST be set to 0.

[4.1.](#) TCA Contexts

[4.1.1.](#) TCA Advertisement for Point-to-Point Connection

In certain cases, the advertisement of a TCA is intended to relate to aggregate traffic over a point-to-point connection between a specific destination and a specific source. A point-to-point connection may be a physical link or a virtual link (e.g. a tunnel). In such cases, a BGP UPDATE message with source AS number and NLRI prefix as an IP address of a TCA Producer can uniquely identify physical/virtual link in order to establish the context for the advertised TCA 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 the PE with the following:

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

The TCA 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 PE's IP address, establishes TCA context for the aggregate traffic through CE-to-PE link.

The TCA advertised in the QoS Attribute in the BGP UPDATE message from PE to CE, with PE's AS number and any other prefix, means TCA for that specific prefix based traffic, a subset of traffic through CE-to-PE link.

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

When BGP UPDATE message with the QoS Attribute, containing TCA SubType, is triggered for a point-to-point connection (physical or logical), the Source AS number in the TCA SubType SHOULD be set to

TCA Producer's AS number and destination AS number SHOULD be set to AS number of BGP peer's that is targeted TCA Consumer.

[4.1.2.](#) TCA Advertisement for Destination AS Multiple Hops Away

When advertised TCA is not for the BGP peer of a TCA Producer, the Source AS field, in the TCA SubType, MUST be set. The list of destination AS(es) also MUST be set, in the TCA SubType, to avoid flooding of the QoS Attribute data in the network beyond those destinations. Destination AS(es) is a list of TCA Consumers the advertised TCA is intended for.

If a new prefix is learned and traffic with this new prefix is subject to TCA parameters that have already been advertised before for other existing prefixes, then the BGP UPDATE for this new prefix MAY include QoS Attribute containing just a TCA ID that was advertised earlier. This BGP UPDATE message does not require to have the whole TCA Content. The TCA ID is sufficient to relate TCA parameters to new advertised prefixes.

[5.](#) QoS Attribute Handling at Forwarding Nodes

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

[5.1.](#) BGP Node Capable of Processing QoS Attribute

If a BGP peer is also a QoS Attribute Speaker, it MAY process the QoS Attribute. If BGP UPDATE message has a QoS Attribute with a list of destination ASes, QoS Attribute Speaker 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.

A QoS Attribute Speaker MUST drop TCA SubType from the QoS Attribute, if there are no more ASes left in the QoS Attribute's destination list. The rest of the QoS Attribute contents may be forwarded if there exist other SubTypes of QoS Attribute and forwarding rules meet other SubTypes requirements. If there is no other SubTypes in that QoS Attribute content then QoS Attribute Speaker MUST drop the entire QoS Attribute all together. BGP Speaker MAY announce further other attributes and NLRI information, if they meet rules defined by other attributes and BGP specification.

Except extracting the entire TCA SubType of the QoS Attribute and trimming the list of Destination AS list, all other content MUST NOT be modified by any QoS Attribute Speaker or BGP Speaker in the path of a BGP UPDATE message.

[5.2.](#) QoS Attribute Handling at Receiver

Once QoS Attribute with the TCA SubType is received at intended receiver (TCA Consumer) , processing of advertised TCA Content is optional for the TCA Consumer. TCA Consumer MAY just trim the Destination AS list as per rules described in this specification, without processing any other content of the Attribute. If Receiver chooses to process advertised TCA content, it may encounter errors beyond the ones described in this document, errors like unavailability of resources if Receiver chooses to implement policies for advertised TCA. In such a case Receiver MAY simply log a message. QoS attribute still MUST be forwarded as per rules defined in this document and rest of the BGP UPDATE message MUST be processed as per BGP specification. If intended receiver is not a QoS Attribute Speaker than BGP Speaker MUST forward this attribute without any change if rest of the BGP UPDATE message also meets forwarding rules as per BGP specification.

When BGP UPDATE messages are triggered only as a result of TCA policy change, propagating BGP UPDATE message beyond intended TCA Consumers is not necessary. If the TCA Consumer device implementations are capable of policy based filtering, it may implement a policy to filter such BGP UPDATE messages based on prefixes and QoS Attribute containing TCA SubType.

[6.](#) Error Handling

Error conditions, while processing of the QoS Attribute content, MUST be handled with the approach of attribute discard as described in [\[RFC7606\]](#). Processing of QoS Attribute content is done by QoS Attribute Speaker and thus in case of errors, resulting in attribute discard, QoS Attribute Speaker SHOULD convey such indication to the BGP Speaker and rest of the BGP message SHOULD be processed by the BGP Speaker as per BGP specification.

7. Deployment Considerations

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

Contracted TCA from PE to CE may be full line-rate or sub line-rate or finer granular controlled services. The advertised TCA 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).

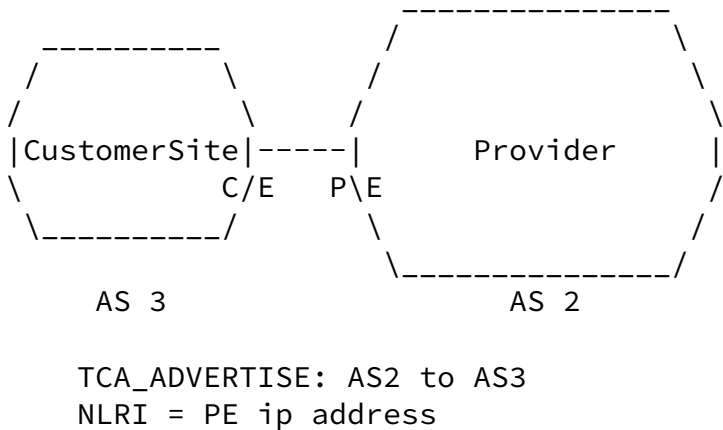
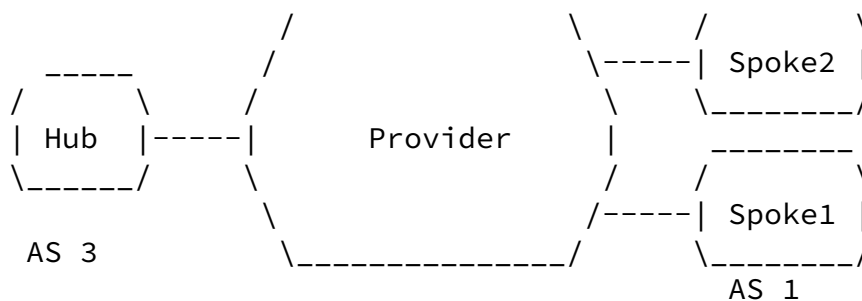


Figure 7: - Example 1

Another use case can be to advertise TCAs among different network sites within one Enterprise network. In Hub and Spoke deployments, Administrator may define TCAs at spoke and advertise QoS TCA 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 TCA to AS3 in the context of that tunnel ip address.

AS 2



TCA_ADVERTISE: AS2 to AS3
NLRI = AS2 tunnel address

TCA_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, TCA parameters may be advertised from one to the other.

8. 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, divided into two pools. One pool of numbers to be assigned on a standards action/early allocation basis. The initial assignments are as shown below. The other pool is for the private use, available range for which is as shown below.

QoS Attribute SubTypes	
=====	
Reserved	0x00
TCA	0x01
Reserved	0x02-0xf0 (Standards Action)
Private use	0xf1-0xff

IANA is requested to create a registry for QoS Attribute TCA Event Types. This is a registry of 4-bits value, divided into two pools.

One pool of numbers to be assigned on a standards action/early allocation basis. One pool of numbers to be assigned on a standards action/early allocation basis. The initial assignments are as shown below. The other pool is for the private use, available range for which is as shown below.

QoS Attribute TCA Event Types	
=====	
Reserved	0x0
ADVERTISE	0x1
Reserved	0x2 - 0xc (Standards Action)
Private use	0xd - 0xf

IANA is requested to create a registry to define QoS Attribute TCA Direction. This is the direction in forwarding path, advertised QoS TCA is applicable to. This is a 2-bit registry. Values for QoS Attribute TCA direction are:

QoS Attribute TCA Direction	
=====	
Reserved	0x0
To source AS from destination AS	0x1
From source AS to destination AS	0x2
Reserved (Standards Action)	0x3

QoS Attribute TCA Traffic Class Element Types will be referring to existing IPFIX IANA types as listed in Table 1. While IPFIX registry

is maintained by IANA out of scope of this specification, the use of IPFIX identifiers for this specification are limited to what is described in Table 1. Any new addition of IPFIX identifiers to this table should be a Standards Action.

IANA is requested to create a registry for QoS Attribute TCA 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
COMMITTED_TSPEC	0x01

PEAK_TSPEC	0x02	
COMMITTED_IN_PROFILE_MARKING	0x03	
COMMITTED_OUT_PROFILE_MARKING	0x04	
PEAK_OUT_PROFILE_MARKING	0x05	
DROP_THRESHOLD	0x06	
RELATIVE_PRIORITY	0x07	
EFFECTIVE_MAX_RATE	0x08	
Standards Action	0x09	- 0x3FFF
FCFS	0x4000	- 0x4FF0

9. Security Considerations

BGP security vulnerabilities analysis is documented in [\[RFC4272\]](#), while BGP-related security considerations are discussed in [\[RFC4271\]](#). Also, the reader may refer to [\[RFC7132\]](#) for more details about BGP path threat model. Means to prevent route hijacking SHOULD be enabled. Such means include RPKI based origin validation [\[RFC7115\]](#) and BGP Path validation (e.g., [\[I-D.ietf-sidr-bgpsec-protocol\]](#)). Rest of the content in this section discusses additional privacy and security considerations that are applicable to the attribute defined in this document.

The information conveyed in the QoS Attribute TCA SubType reveals sensitive data that should not be exposed publicly to non-authorized parties. Deployment considerations mainly target use of QoS Attribute and TCA SubType in managed networks and those where a trust relationship is in place (Customer to Provider, or Provider to Provider). Administrators MUST disable this attribute to be sent to a remote peer which whom no trust relationship is in place. Both TCA Producer and Consumer SHOULD NOT publish valid TCA IDs to non-authorized nodes.

The attribute may be advertised by a misbehaving node to communicate TCA parameters that are not aligned with the TCA agreements. The enforcement of TCA parameters is outside the scope of this document.

The attribute defined in this document may be used by a misbehaving node for denial-of-service (e.g., inadequately rate-limit or drop some critical traffic). As a mitigation, a BGP peer MUST accept this

attribute only from trusted BGP peers. For example, ACLs may be configured to identify the trusted ASes that are allowed to send the attribute. Further, administrators of a TCA Consumer's domain are RECOMMENDED to generate TCA ID using pseudo-random schemes [RFC4086]. Using robust TCA IDs make it hard to guess a valid TCA.

10. Acknowledgements

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

11. References

11.1. 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, <<https://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, <<https://www.rfc-editor.org/info/rfc2212>>.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, [RFC 3629](#), DOI 10.17487/RFC3629, November 2003, <<https://www.rfc-editor.org/info/rfc3629>>.
- [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, <<https://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, <<https://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, <<https://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, <<https://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, <<https://www.rfc-editor.org/info/rfc7606>>.

11.2. Informative References

- [I-D.ietf-sidr-bgpsec-protocol] Lepinski, M. and K. Sriram, "BGPsec Protocol Specification", [draft-ietf-sidr-bgpsec-protocol-22](#) (work in progress), January 2017.
- [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, <<https://www.rfc-editor.org/info/rfc2475>>.
- [RFC4086] Eastlake 3rd, D., Schiller, J., and S. Crocker, "Randomness Requirements for Security", [BCP 106](#), [RFC 4086](#), DOI 10.17487/RFC4086, June 2005, <<https://www.rfc-editor.org/info/rfc4086>>.
- [RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", [RFC 4272](#), DOI 10.17487/RFC4272, January 2006, <<https://www.rfc-editor.org/info/rfc4272>>.
- [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, <<https://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, <<https://www.rfc-editor.org/info/rfc6020>>.

Internet-Draft

Inter-domain TCA Exchange

January 2018

- [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, <<https://www.rfc-editor.org/info/rfc6241>>.
- [RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", [RFC 6793](#), DOI 10.17487/RFC6793, December 2012, <<https://www.rfc-editor.org/info/rfc6793>>.
- [RFC7132] Kent, S. and A. Chi, "Threat Model for BGP Path Security", [RFC 7132](#), DOI 10.17487/RFC7132, February 2014, <<https://www.rfc-editor.org/info/rfc7132>>.
- [RFC7297] Boucadair, M., Jacquenet, C., and N. Wang, "IP Connectivity Provisioning Profile (CPP)", [RFC 7297](#), DOI 10.17487/RFC7297, July 2014, <<https://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, <<https://www.rfc-editor.org/info/rfc7674>>.
- [RFC8040] Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", [RFC 8040](#), DOI 10.17487/RFC8040, January 2017, <<https://www.rfc-editor.org/info/rfc8040>>.

Authors' Addresses

Shitanshu Shah

Email: shitanshu_shah@hotmail.com

Keyur Patel
Arrcus, Inc

Email: keyur@arrcus.com

Sandeep Bajaj
Viptela

Internet-Draft

Inter-domain TCA Exchange

January 2018

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

