Internet Engineering Task Force
INTERNET DRAFT
<draft-bonaventure-bgp-qos-00.txt>

Olivier Bonaventure
FUNDP

February, 2001
Expires August, 2001

Using BGP to distribute flexible QoS information

Status of this Memo

Abstract

This document proposes a flexible QoS attribute that can be used to distribute QoS information with BGP. The proposed attribute allows to associate a set of supported PHB, transit delay and bandwidth information to an UPDATE message. The flexibility of the proposed attribute allows each AS to decide independently which QoS information to redistribute to its peers.

1 Introduction

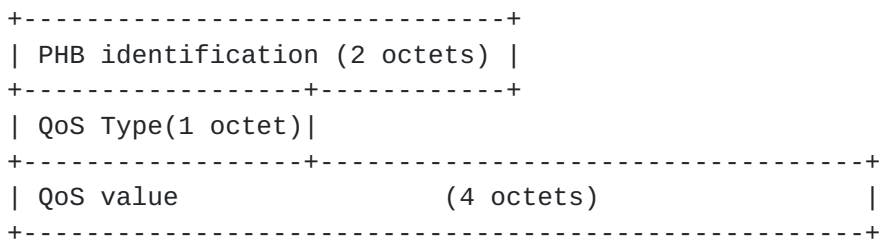
In this document, we propose a mechanism to associate QoS information to prefixes that are announced by BGP. Inside a single autonomous system, recent work has focussed on the definition of QoS information that can be distributed by link state routing protocols. In the case of Interior Gateway Protocols (IGP), the focus was the definition of the minimum set of QoS attributes that can be

associated to a link to support the QoS or traffic engineering features without overloading the flooding protocol or the route computation [SL99, KY00, KRB^+00]. At the interdomain level, the issue to be considered is different. There are many different autonomous systems on the global Internet with very different requirements and needs in terms of Quality of Service. For example, the autonomous systems that are part of a confederation could want to distribute detailed QoS information inside the confederation while announcing far fewer QoS information on the global Internet. An ISP could also want to provide different types of QoS information to its clients than to other ISPs at public interconnection points. An ISP could want to distribute different QoS information to private peers than to public peers. When dealing with differentiated services, an ISP could want to announce a bandwidth limit on routes associated with the EF PHB while no limit for routes associated with the AF PHB.Many other situations are possible. To support these various requirements, a flexible method to distribute QoS information is necessary for BGP. The solution proposed in this document is equally applicable to the global Internet as well as to BGP-based VPN solutions [RR99, KLV^+00, CTS00].Many complex routing policies, including some related to QoS, can be implemented by using the communities [CTL96] or the extended communities attribute [RTR01]. However, those communities have a local semantics and their utilization must be agreed between each pair of routers. When considering the support of QoS across interdomain boundaries, it would be more useful to have a flexible QoS attribute that can be used for this purpose instead of letting each AS define its private set of communities for almost the same purpose.

2 The QoS attribute

This document defines a new QoS attribute that can be used to associate QoS information to an UPDATE message. This QoS attribute applies to all NLRI information contained inside the UPDATE message. The QoS attribute is a variable length non-transitive optional attribute. It is encoded as follows :

- The attribute flags shall indicate that the QoS attribute is optional, non-transitive and the extended length bit is set to one since the QoS attribute may be longer than 256 bytes.
- The attribute type code is to be assigned by IANA
- The length of the entire attribute is encoded in two octets
- The value of the QoS attribute is encoded as a list of triples :



The QoS type code allows the definition of 256 different types of QoS values. Out of these 256 possible values, value 0 is reserved for future utilization, values 1-127 are to be defined by IANA while values 128-255 are reserved for vendor specific QoS attributes. This document defines QoS types 1-6.

2.1 PHB identification

The PHB identification is used for two purposes. First, it allows a border router to announce the PHB that it supports. Second, by using the the various QoS values, it is impossible to associate a specific QoS metric to each PHB. The PHB shall be encoded as specified in [BBCF01].

2.2 Empty QoS value

This special QoS value shall be used by a border router wishing to announce the support of a specific PHB towards the associated prefix without associating detailed QoS information. The QoS type for this value is 1. In this case, the QoS value field shall be equal to 0x00000000.

2.3 Maximum Bandwidth

The maximum bandwidth QoS value shall be used by a border router to associate a maximum bandwidth to a given PHB. This attribute shall be used by a border router to announce, with eBGP or iBGP, the maximum bandwidth along the path towards the associated prefix. This QoS value differs from the maximum bandwidth community defined in [RTR01] that is only used for iBGP.The maximum bandwidth QoS value is reported in bytes per second and encoded as an IEEE single precision floating point number by using the same format as proposed in [RTR01]. The QoS type field for the maximum bandwidth QoS value is 2.

2.4 Available Bandwidth

The available bandwidth QoS value shall be used by a border router to announce the available bandwidth associated with an announced prefix. This information shall correspond to the amount of available bandwidth on the path towards the associated prefix.The available bandwidth QoS value is reported in bytes per second and encoded as an IEEE single precision floating point number by using the same format as proposed in [RTR01]. The QoS type field for the maximum bandwidth QoS value is 3.We expect that an information that potentially varies frequently such as the Available Bandwidth will not be distributed through the whole Internet and that filters will be used to limit its distribution. It could for example be used within a confederation or for specific VPN purposes without being re-exported.

2.5 Maximum Transit delay

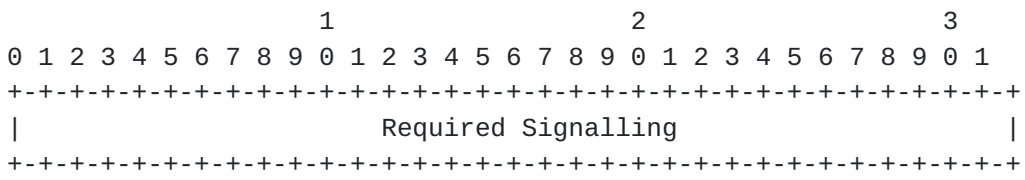
This QoS value is used by a border router to associate a maximum transit delay to an announced prefix. This QoS value is an indication of the maximum (one-way) transit delay required to reach the farthest IP address of the associated prefix.The maximum transit delay is reported in units of one microsecond and encoded as an unsigned 32 bits number. The QoS type field of the maximum transit delay QoS value is 4.

2.6 Minimum Transit delay

This QoS value is used by a border router to associate a minimum transit delay to an announced prefix. This QoS value is an indication of the minimum (one-way) transit delay required to reach the closest IP address of the associated prefix.The minimum transit delay is reported in units of one microsecond and encoded as an unsigned 32 bits number. The QoS type field of the minimum transit delay QoS value is 5.

2.7 Required signalling

This QoS value may be used by a border router to indicate whether or not an explicit signalling operation is required to reach the NLRI included in the UPDATE message. A border router may wish to enforce an explicit signalling operation to be able to perform admission control for example. This QoS value is encoded as a 32 bits field that contains a list of bit flags. Each flag indicates the type of signalling operation required to utilize the route.



Bit0 : No explicit signalling required
Bit1 : Supports RSVP for Integrated services
Bit2 : Supports RSVP for MPLS LSP
Bit3 : Supports CR-LDP for MPLS LSP

When Bit0 is set (reset), this indicates that the border router sending the UPDATE message is able (refuses) to accept normal IP packet towards the associated NLRI. When Bit1 is set (reset), this indicates that the border router is willing (refuses) to process the establishment of Integrated Services flows with RSVP towards the associated NLRI. When Bit2 is set (reset), this indicates that the border router is willing (refuses) to process MPLS LSP establishment request with RSVP towards the associated NLRI. When Bit3 is set (reset), this indicates that the border router is willing (refuses) to process MPLS LSP establishment request with CR-LDP. The QoS type field of the required signalling QoS value is 6. Additional bit flags may be defined to cover other signalling methods [BPSA01]

2.8 Routes without a QoS attribute

When a BGP router receives an UPDATE message that does not contain a QoS attribute, it may assume the following defaults :

- The set of supported PHB should be set to BE.
- The Maximum and Available bandwidth should be set to infinite or the bandwidth of the link from which the UPDATE message is received (if known).
- The required signalling flags should indicate that no explicit signalling is required
- The Minimum Transit Delay should be set to 0.
- The Maximum Transit Delay should be set to infinite.

Similar rules apply when the received QoS attribute that not contain a value for each QoS type.

3 Aggregation and QoS attributes

A key issue for the scalability of the interdomain routing system is the possibility to aggregate routes in a single announcement [CS99]. When QoS information is associated with a prefix, the accuracy of the QoS information may conflict with the need for aggregation. For example, consider the network shown in figure 1.

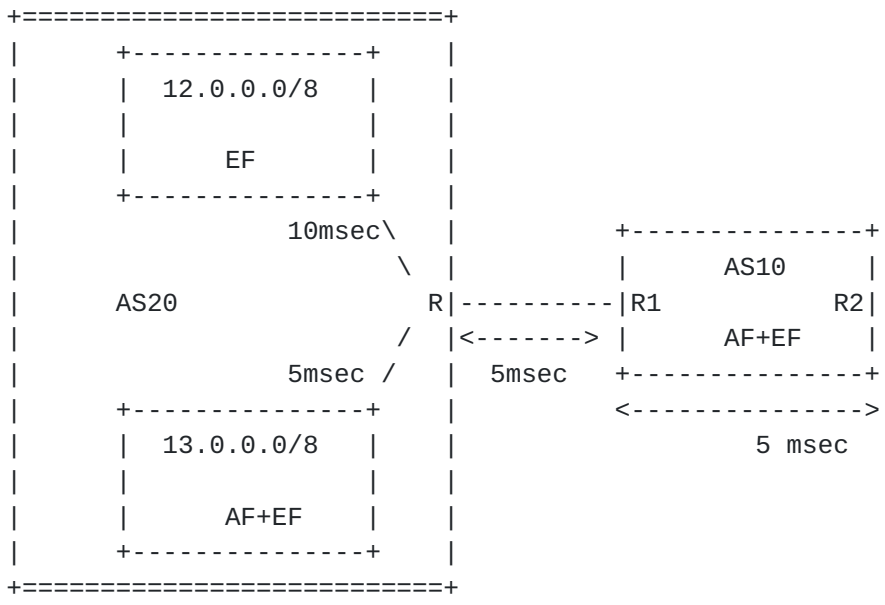


Figure 1: Simple interdomain topology

Assume that in this network, AS20 wants to advertise detailed information about networks 12.0.0.0/8 and 13.0.0.0/8 to AS10. It this case, the border router of AS20 would associate a maximum delay of 10 msec and the EF PHB to prefix 12.0.0.0/8. Network 13.0.0.0/8 would be announced with the AF and EF PHB and a maximum delay of 5 msec. Based on this information, AS10 would have to decide how to announce these two prefixes through router R2. If AS10 wants to provide detailed QoS information, then it should announce prefix 12.0.0.0/8 with the EF PHB and a maximum transit delay of 20 msec and prefix 13.0.0.0/8 with both the AF and EF PHB and a maximum transit delay of 15 msec.On the other hand, if AS10 wants to reduce the amount of prefix announced, then it should aggregate 12.0.0.0/8 and 13.0.0.0/8 in a single prefix. In this case, AS10 would announce that network 12.0.0.0/7 supports the EF PHB and that the maximum transit delay to reach this network is 20 msec.In many cases, a border router will need to aggregate several specific routes into a single less specific route. This aggregation will inevitably introduce approximation in the route announced. The only way to avoid loosing QoS information is to avoid aggregating routes. However, this solution suffers from scalability problems.Border routers should have the highest flexibility to decide

which QoS information should be announced to each peer, possibly on a prefix-by-prefix basis. We believe that it is not possible in such a document to impose strict conditions on how a border router propagates QoS information received from a peer. There are however some guidelines that should be followed in the mechanisms used by a border router to manipulate QoS information :

- When redistributing a route, a border router should try, depending on its policy, and whenever possible, to reduce the number of prefixes announced.
- When a border router needs to redistribute a route, it may modify the QoS attribute. The border router may add any QoS value associated with one of the PHB identifications explicitly indicated in the received QoS attribute. It cannot add a new PHB to the set of PHB included in the received QoS attribute. The only exception to this rule is when a border router receives an UPDATE message that does not contain the QoS attribute. In this case, this route is implicitly valid for the Best-Effort PHB and thus the border route may add QoS values provided that they are associated with the Best-Effort PHB.
- A border router that receives routes with QoS information from a peer and needs to redistribute it is in principle allowed to update or remove any received QoS information. The only exception to this rule is that if a router receives a QoS attribute that does not contain the Best-Effort PHB. In this case, the border router cannot entirely remove the QoS attribute. This implies that in this case the route should not be distributed towards a BGP router that does not support the QoS attribute defined in this document.
- When a border router needs to update the maximum (resp. available) bandwidth included in a QoS attribute, it may decrease this maximum (resp. available) bandwidth, but cannot increase it. When updating the bandwidth values, it should ensure that, for each PHB, the maximum bandwidth remains larger than the available bandwidth.
- When a border router needs to update the maximum (resp. minimum) transit delay included in a QoS attribute, it may increase this maximum (resp. minimum) transit delay, but cannot increase it. When updating the delay values, it should ensure that, for each PHB,

the
maximum transit delay remains larger than the minimum transit delay.

4 Capability negotiation

To advertise the QoS capability to a peer, a BGP speaker uses the BGP Capabilities Advertisement [CS00] This capability is advertised using the Capability code TBD_IANA (to be defined by IANA).The fields in the Capabilities Optional Parameter are set as follows. The Capability Code field is set to TBD_IANA (to be defined by IANA). The Capability Length field is set to 1. The Capability Value contains the version of the QoS attribute. This document defines version 1 of the QoS attribute.To obtain a bi-directional exchange of QoS attributes between a pair of border routers, each border router must advertise to the other the support of the QoS attribute.

5 Related work

Two extensions to BGP have been proposed recently to advertise QoS information with BGP [AV00, Jac00].In [Jac00], the QoS information is associated to a prefix by defining a new QOS_NLRI attribute. This optional transitive attribute is used to associate QoS information to an announced prefix. This document defines several types of QoS information (reserved bandwidth, available bandwidth, minimum, maximum and average transit delay) and the support of the AF PHB. However, it only allows to associate a single QoS information to each prefix. In contrast, our proposal is to define a flexible QoS attributes that can be manipulated by BGP speakers to support different QoS policies. We expect for example than inside a confederation or for VPN applications, the QoS information will be richer than on the global Internet. The flexibility of our solution allows to easily support these requirements.

[AV00] proposes on the other hand to define new optional attributes used to compute TE weights. This document also proposes methods to aggregate these traffic engineering weights. The intended application of [AV00] is to provide a mechanism similar to the BGP Multi Exit Discriminator without being limited to adjacent AS.

6 Conclusion

In this document, we have a flexible QoS attribute that can be used to distribute QoS information with BGP. The proposed attribute can be

used by a BGP speaker to announce the set of PHB that it supports and to optionally associate specific QoS values (minimum and maximum transit delay, available and maximum bandwidth) to each supported PHB.

Acknowledgements

We would like to thank Christian Jacquenet, Guy Leduc, Stefaan De Cnodder and Steve Uhlig for their very useful comments on this document. This work was partially funded by the European Commission, within the ATRIUM IST project.

References

[AV00] B. Abarbanel and S. Venkatachalam. Bgp-4 support for traffic engineering. Internet draft, draft-abarbanel-idr-bgp4-te-00.txt, work in progress, May 2000.

[BBCF01] D. Black, S. Brim, B. Carpenter, and F. Le Faucheur. Per hop behavior identification codes. Internet draft, draft-ietf-diffserv-2836bis-00.txt, work in progress, January 2001.

[BPSA01] M. Blanchet, F. Parent, and B. St-Arnaud. Optical bgp (obgp): Interas lightpath provisioning. Internet draft, draft-parent-obgp-00.txt, work in progress, January 2001.

[CS99] E. Chen and J. Stewart. A framework for inter-domain route aggregation. Internet RFC 2519, February 1999.

[CS00] R. Chandra and J. Scudder. Capabilities negotiation with BGP-4. Internet RFC2842, May 2000.

[CTL96] R. Chandra, P. Traina, and T. Li. BGP communities attribute. Internet RFC 1997, August 1996.

[CTS00] J. De Clercq, Y. T'Joens, and P. De Schrijver. BGP/IPsec VPN. Internet draft, draft-declercq-bgp-ipsec-vpn-00.txt, work in progress, July 2000.

[Jac00] C. Jacquenet. Providing quality of service indication by the BGP-4 protocol : the QoS_NLRI attribute. Internet draft, draft-jacquenet-qos-nlri-01.txt, work in progress, November 2000.

[KLV^+00] K. Kompella, M. Leelanivas, Q. Vohra, J. Achirica, R. Bonica, C. Liljenstolpe, E. Metz, C. Sargor, and V. Srinivasan. MPLS-based Layer 2 VPNs. Internet draft, draft-kompella-mpls-

l2vpn-02.txt, work in progress, November 2000.

[KRB^+00] K. Kompella, Y. Rekther, A. Banerjee, J. Drake, G. Bernstein, D. Fedyk, E. Mannie, D. Saha, and V. Sharma. IS-IS extensions in support of MPL(ambda)S. Internet draft, draft-kompella-isis-ompls-extensions-00.txt, work in progress, July 2000.

[KY00] D. Katz and D. Yeung. Traffic engineering extensions to ospf. Internet draft, draft-katz-yeung-ospf-traffic-02.txt, work in progress, August 2000.

[RR99] E. Rosen and Y. Rekhter. BGP/MPLS VPNs. Internet RFC2547, March 1999.

[RTR01] S. Ramachandra, D. Tappan, and Y. Rekhter. Bgp extended communities attribute. Internet draft,draft-ramachandra-bgp-ext-communities-08.txt, work in progress, January 2001.

[SL99] H. Smit and T. Li. Is-is extensions for traffic engineering. Internet draft, draft-ietf-isis-traffic-01.txt, work in progress, February 1999.

A Examples

In this appendix, we provide a few examples of the utilization of the QoS attribute. Assume first that AS1 wishes to announce to its peers its ability to support three different PHB : Best-Effort, EF and AF. Assume also that AS1 wishes to announce a maximum bandwidth of 100 Mbps for BE traffic, 10 Mbps for AF traffic and 1 Mbps for EF. In this case, it will send a set QoS attribute composed of :

```
{
PHB=BE;QoS-Type=2 [Maximum Bandwidth];MaxBW=100,
PHB=AF;QoS-Type=2 [Maximum Bandwidth];MaxBW=10,
PHB=EF;QoS-Type=2 [Maximum Bandwidth];MaxBW=1
}
```

Assume now that a border router wants to indicate that it supports the BE PHB without associating any QoS information to this PHB. In this case, the QoS attribute would be composed of :

```
{
PHB=BE;QoS-Type=1 [Empty]
}
```

Assume now a special border router (e.g. a router controlling a label switch router) that is not able to forward IP packets at line rate but is able to support the establishment of label switched paths with RSVP and CR-LDP. Also assume that this router supports AF and BE. In this case, this router should associate the following QoS attribute to each route it is sending to a peer :

```
{
PHB=BE;QoS-Type=6 [Required
Signalling];[Bit0:Reset,Bit1:Reset,Bit2:Set,Bit3:Set],
PHB=AF;QoS-Type=6 [Required
Signalling];[Bit0:Reset,Bit1:Reset,Bit2:Set,Bit3:Set]
}
```

Another possibility is a border router that provides best-effort service to normal IP packets but is able to provide QoS to label switched paths established by RSVP only. In this case, it could associate the following QoS attribute to routes announced to a peer :

```
{
PHB=BE;QoS-Type=6 [Required
Signalling];[Bit0:Set,Bit1:Reset,Bit2:Reset,Bit3:Reset]
PHB=AF;QoS-Type=6 [Packet switching
capability];[Bit0:Set,Bit1:Reset,Bit2:Set,Bit3:Reset]
PHB=AF;Qos-Type=2 [Maximum Bandwidth];MaxBW=100
PHB=EF;QoS-Type=6 [Packet switching
capability];[Bit0:Set,Bit1:Reset,Bit2:Reset,Bit3:Reset]
PHB=EF;Qos-Type=2 [Maximum Bandwidth];MaxBW=10
PHB=EF;Qos-Type=4 [Minimum Delay];MinTD=10msec
PHB=EF;Qos-Type=5 [Maximum Delay];MaxTD=100msec
}
```

This QoS information indicates that the border router supports the BE, AF and EF PHB. For the BE PHB, there are no specified QoS values. For the AF PHB, a LSP must be established before actual traffic can flow and the maximum reservable bandwidth is 100 Mbps. For the EF PHB, a LSP must also be established before actual traffic can flow and the transit delay is expected to be between 10 and 100 msec. The maximum bandwidth reservable for EF is set to 10 Mbps.

Author's Address

Olivier Bonaventure
Infonet group
Institut d'Informatique
Facultes Universitaires Notre-Dame de la Paix
Rue Grandgagnage 21, B-5000 Namur, Belgium.
E-mail: Olivier.Bonaventure@info.fundp.ac.be
URL : http://www.infonet.fundp.ac.be