

Network Working Group
INTERNET-DRAFT
[draft-liang-bgp-qos-00.txt](#)
Expires: December 21, 2006

L. Benmohamed
Johns Hopkins University
C. Liang
Johns Hopkins University
E. Naber
Johns Hopkins University
A. Terzis
Johns Hopkins University
June 19, 2006

QoS Enhancements to BGP in Support of Multiple Classes of Service

Status of This Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

Abstract

This document specifies the requirements posed on multi-domain Quality of Service (QoS) routing protocols that provide multiple classes of service with multi-dimensional QoS requirements. In particular, it discusses enhancements to Border Gateway Protocol version 4 (BGPv4) that allow nodes to discover multiple paths with associated QoS attributes. In addition, this documents discusses the details of several new algorithms, such as the dominant path selection algorithm (DPSA).

1. Introduction

BGPv4 is the dominating inter-domain routing protocol in the Internet. It was designed with minimal overhead traffic for scalability, and it offers extensive policy support for business needs. Information sharing among External Border Gateway Protocol

(eBGP) peers is limited. The only information passed from an AS to its neighbors is the set of destination network prefixes reachable from itself and the sequence of ASs to each destination network prefix.

Many emerging needs in commercial and military networking have exposed limitations of the current eBGP. These future IP networks will support a very diverse mix of applications with very diverse QoS requirements. In addition, some of these networks may consist of a diverse set of component networks (wireless and wireline, fixed and mobile with different degrees of mobility, long-term and short-term ad-hoc), and these component networks may have dynamic service capabilities. Therefore, different paths between the same endpoints may have very different QoS characteristics, and these QoS characteristics may be time-variant. Having the capability to specify QoS requirements in routing, session admission, packet scheduling, buffer management, service restoration priority, degree of security protection, and similar other decisions will become an important need in future IP networks.

This document specifies BGPv4 enhancements that allow multi-topology (exposing multiple routes) and QoS-aware routing, using several QoS metrics of importance to different applications. Unlike some other similar work, our enhancements support multiple QoS metrics and do not require any changes in the inter-domain routing architecture that BGPv4 is based upon. Since the enhanced BGP does not limit routers to advertise only one route for each destination prefix, we define the notion of dominant paths to keep the amount of information transfer to the minimum needed to be scalable. The enhanced BGP also incorporates mechanisms, such as Multiprotocol Label Switching (MPLS), and defines network-wide traffic classes to pin the route taken by packets with the same QoS requirements.

[2.](#) BGPv4 Enhancements

There are five enhancements to BGPv4 that enable multi-paths and QoS-aware routing:

- (1) Exchanging potentially multiple paths per destination prefix.
- (2) Maintaining multiple QoS metrics for each path.
- (3) Pruning the set of known paths to a dominant set while maintaining optimality.
- (4) Choosing a particular path from this dominant set for the unique QoS requirements of a traffic class.
- (5) Enforcing the selected route.

BGPv4 restricts each router to advertise to its neighbors only one

route per destination prefix. This information hiding behavior can prevent a router from learning the particular path that most appropriately addresses the QoS requirements for a given traffic class. Enhancement (1) removes this limitation, allowing each BGP router to advertise a set of dominant paths. Enhancement (2) associates a list of QoS metrics with each link, which are then used to make route decision. Details on how path QoS metrics are embedded in BGP_UPDATE messages and maintained will be discussed in more detail in later sections. The notion of dominant paths is enhancement (3), and it prevents each BGP router from advertising every path it knows. Dominant paths are selected by dominant path selection algorithm (DPSA), which is discussed in more detail in later sections. Exchanging all of these additional paths and QoS attributes is pointless unless the QoS-aware routing path decision is actually enforced. Enhancement (5) can be implemented with various mechanisms, which are discussed in our previous work [1]. We have chosen to run MPLS on route assignment by enhancement (4) to pin the path for each application's data flow. Class-assignment algorithm of enhancement (4) will be discussed in more detail in later sections.

Incorporating these enhancements requires updating BGPv4 path selection process accordingly.

- BGPv4 first manipulates path attributes, such as Local_Pref and AS_Path, so that routes from one or some neighbors are enabled. However, in order for QoS routing to be effective by being able to choose among a larger set of routes, routes from all neighbors should ideally be enabled.

- In BGPv4 path selection process, each border router (BR) that terminates an enabled route selects itself among all enabled routes as an AS egress point. Then, non-egress nodes selects one of these egress points. Instead, under QoS routing, all enabled routes to a destination D received at a BR (from both iBGP and eBGP peers) undergo DPSA where a set of dominant paths is selected. This set of dominant paths is maintained for routing and advertisement.

2.1 Maintaining Path QoS Metrics

The BGPv4 UPDATE message includes the AS_PATH attribute, which stores the sequence of ASs to reach the associated destination prefixes. In order to retain the structure of the BGPv4 UPDATE message, path QoS metrics are embedded in the AS_PATH attribute. This extension of the AS_PATH attribute has been modeled after TLV (Type-Length-Value) model. The TLV model provides a flexible architecture that will support the specific metrics that we are currently interested in, and it also provides support for any additional metrics in the future. Figure 1 and 2 show the old format and the new format of AS-PATH attribute:

```

-----
| Length of | Path to |
| Attribute | Prefix  |
-----

```

Figure 1: BGPv4 AS_PATH attribute format

```

-----
| Length of | # Metrics |   Type   |   Length   |   Value   |
| Attribute |   = N    | (Metric 1) | (Metric 1) | (Metric 1) | ...
-----

```

```

-----
... |   Type   |   Length   |   Value   | Path to |
    | (Metric N) | (Metric N) | (Metric N) | Prefix  |
-----

```

Figure 2: Extended AS_PATH attribute format

Each path QoS metric value is the accumulation of the QoS metric value of all the links that the path consists of. Therefore, maintaining path QoS metrics is a combined effort of each node along

that path. The general accumulation rules are:

- When advertising a path to another AS, the advertising node combines its intra-AS metric values with the metric values accumulated for the path thus far.
- When receiving a path from another AS, the receiving node combines the metric values on the link that connects the receiving node to the advertising node with the metric values accumulated for the path thus far.

2.2 Dominant Path Select Algorithm (DPSA)

For QoS requirements with a single metric, such as bandwidth, it is sufficient to know a path with the largest available bandwidth. If the best path is not feasible, then no other paths are. However, for QoS requirements with multiple metrics, such as bandwidth and delay, there might not be a single best path. The notion of dominant path is defined as a path where there is no other path with all QoS metrics being better. If a path P dominates a set S of paths, then paths in S do not need to be advertised as path P is the only one needed to make QoS routing decisions. Indeed, if a connection request cannot be supported by P then no path in S can satisfy the request.

To illustrate DPSA, we will look at a small network topology with six paths from the source to the destination node. In addition, this network topology has two QoS metrics of interest: delay and bandwidth.

^

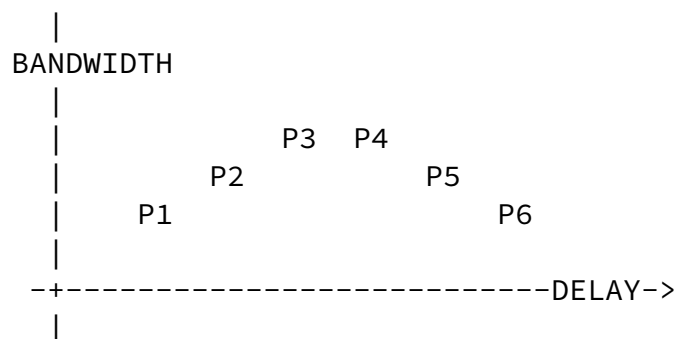


Figure 3: Delay and bandwidth QoS characteristics of the six paths

Figure 3 shows the delay and bandwidth QoS characteristics of the six paths in the network. DPSA selects P1, P2, and P3 to be the dominant paths because they are not covered by any other paths. P4, P5, and P6 are covered by P3 because P3 is equal or better in all QoS metrics.

For each destination prefix in the Loc-Routing Information Base (Loc-RIB), we have to select a set of dominant paths for advertisement. Therefore, DPSA should be logically placed between Loc-RIB and BGPv4 Output Policy Engine. If there are more than one potentially dominant paths with all QoS metrics being equal, then the tie-breaking strategies described below narrows down to only one path. The first strategy is to prefer paths with the lowest AS hop counts, and the second strategy is to prefer paths with the lowest next-hop IP address.

[2.3](#) Route Assignment For Each Traffic Class

There are various mechanisms that can be used to pin the selected path for a particular application's data flow. We have chosen to run MPLS on the notion of traffic classes. A set of network-wide traffic classes are pre-defined on every node, and the task of class-assignment algorithm is to assign at most dominant one path to each traffic class under every destination prefix. For each traffic class, the algorithm starts by selecting a set of paths that can satisfy the QoS requirements of the traffic class. Then, from this set of paths, the best path is assigned to the traffic class. For QoS requirements with bandwidth and delay, the best path can be defined as the path with the largest Euclidean distance from the traffic class requirement.

For each destination prefix in the Loc-Routing Information Base (Loc-RIB), we have to assign at most one dominant path to each traffic class. Therefore, class-assignment assignment should be logically placed between Loc-RIB and Forwarding Information Base (FIB). If there are more than one suitable paths for a traffic class, then the tie-breaking strategies described below narrows down to only one path. The first strategy is to prefer paths with the lowest AS hop counts, and the second strategy is to prefer paths with the lowest

2.4 Forwarding Decision

Forwarding decision at data plane depends on two information: packet destination address and packet traffic class. Both IPv4 Type of Service (TOS) field and IPv6 Priority field can store the identifier of the traffic class the packet belongs to. Longest-prefix-matching algorithm is first performed on the packet destination address to find the most specific destination address prefix in FIB. Since there might be multiple routes under a given address prefix in FIB, packet traffic class identifier is then used to match the route with the same traffic class identifier. The packet is dropped either if longest-prefix-matching algorithm does not return a destination address prefix from FIB, or if there is no route matching packet traffic class.

3. Relevant Results

Simulation results on the dynamic behavior and the scalability of enhanced BGP are discussed in our previous work [2]. In addition, a proof of DPDA convergence under the assumption of synchronous operation is also presented in our previous work [2].

4. Security Considerations

This document does not directly concern security. It is believed that this extension inherits security issues in BGPv4.

5. IANA Considerations

This document has no actions for IANA.

6. Informative References

[1] L. Benmohamed, B. Doshi, T. DeSimone, R. Cole, "Inter-Domain Routing with Multi-Dimensional QoS Requirements", IEEE Milcom 2005

[2] L. Benmohamed, C. Liang, E. Naber, A. Terzis, "QoS Enhancements to BGP in Support of Multiple Classes of Service", June 2006

7. Authors' Addresses

Lotfi Benmohamed
Johns Hopkins University - Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel MD, 20723
US

EMail: lotfi.benmohamed@jhuapl.edu

Internet-Draft

BGP QoS Enhancements

June 2006

Chieh-Jan Mike Liang
Johns Hopkins University - Computer Science Department
3400 North Charles Street, 224 NEB
Baltimore MD, 21218
US

EMail: cliang4@cs.jhu.edu

Eric Naber
Johns Hopkins University - Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel MD, 20723
US

EMail: eric.naber@jhuapl.edu

Andreas Terzis
Johns Hopkins University - Computer Science Department
3400 North Charles Street, 224 NEB
Baltimore MD, 21218
US

Phone: +1 410 516 5847
EMail: terzis@cs.jhu.edu

Copyright Notice

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

