Internet Engineering Task Force (IETF)
Request for Comments: 6601
Category: Experimental
ISSN: 2070-1721

G. Ash, Ed.
AT&T
D. McDysan
Verizon

April 2012

# Generic Connection Admission Control (GCAC) Algorithm Specification for IP/MPLS Networks

## Abstract

This document presents a generic connection admission control (GCAC) reference model and algorithm for IP-/MPLS-based networks. Service provider (SP) IP/MPLS networks need an MPLS GCAC mechanism, as one motivational example, to reject Voice over IP (VoIP) calls when additional calls would adversely affect calls already in progress. Without MPLS GCAC, connections on congested links will suffer degraded quality. The MPLS GCAC algorithm can be optionally implemented in vendor equipment and deployed by service providers. MPLS GCAC interoperates between vendor equipment and across multiple service provider domains. The MPLS GCAC algorithm uses available standard mechanisms for MPLS-based networks, such as RSVP, Diffserv-aware MPLS Traffic Engineering (DS-TE), Path Computation Element (PCE), Next Steps in Signaling (NSIS), Diffserv, and OSPF. The MPLS GCAC algorithm does not include aspects of CAC that might be considered vendor proprietary implementations, such as detailed path selection mechanisms. MPLS GCAC functions are implemented in a distributed manner to deliver the objective Quality of Service (QoS) for specified QoS constraints. The objective is that the source is able to compute a source route with high likelihood that via-elements along the selected path will in fact admit the request. In some cases (e.g., multiple Autonomous Systems (ASes)), this objective cannot always be met, but this document summarizes methods that partially meet this objective. MPLS GCAC is applicable to any service or flow that must meet an objective QoS (delay, jitter, packet loss rate) for a specified quantity of traffic.

## Status of This Memo

This document is not an Internet Standards Track specification; it is published for examination, experimental implementation, and evaluation.

This document defines an Experimental Protocol for the Internet community. This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6601.

## Copyright Notice

Table of Contents

## 1. Introduction

This document presents a generic connection admission control (GCAC) reference model and algorithm for IP-/MPLS-based networks. Service provider (SP) IP/MPLS networks need an MPLS GCAC mechanism, as one motivational example, to reject Voice over IP (VoIP) calls when additional calls would adversely affect calls already in progress. Without MPLS GCAC, connections on congested links will suffer degraded quality. Given the capital constraints in some SP networks, over-provisioning is not acceptable. MPLS GCAC supports all access technologies, protocols, and services while meeting performance objectives with a cost-effective solution and operates across routing areas, autonomous systems, and service provider boundaries.

This document defines an MPLS GCAC reference model, algorithm, and functions implemented in one or more types of network elements in different domains that operate together in a distributed manner to deliver the objective QoS for specified QoS constraints, such as bandwidth. With MPLS GCAC, the source router/server is able to compute a source route with high likelihood that via-elements along the selected path will in fact admit the request. MPLS GCAC includes nested CAC actions, such as RSVP aggregation, nested RSVP - Traffic Engineering (RSVP-TE) for scaling between provider edge (PE) routers, and pseudowire (PW) CAC within traffic-engineered tunnels. MPLS GCAC focuses on MPLS and PW-level CAC functions, rather than application-level CAC functions.

MPLS GCAC is applicable to any service or flow that must meet an objective QoS (latency, delay variation, loss) for a specified quantity of traffic. This would include, for example, most real-time/RTP services (voice, video, etc.) as well as some non-real-time services. Real-time/RTP services are typically interactive, relatively persistent traffic flows. Other services subject to MPLS GCAC could include, for example, manually provisioned label switched paths (LSPs) or PWs and automatic bandwidth assignment for applications that automatically build LSP meshes among PE routers. MPLS GCAC is applicable to both access and backbone networks, for example, to slow-speed access networks and to broadband DSL, cable, and fiber access networks.

This document is Experimental. It is intended that service providers and vendors experiment with the GCAC concept and the algorithm described in this document in a controlled manner to determine the benefits of such a mechanism. That is, they should first experiment with the GCAC algorithm in their laboratories and test networks. When testing in live networks, they should install the GCAC algorithm on selected routers in only part of their network, and they should

carefully monitor the effects. The installation should be managed such that the routers can quickly be switched back to normal operation if any problem is seen.

Since application of GCAC is most likely in Enterprise VPNs and/or internal TE infrastructure, it is RECOMMENDED that the experiment be conducted in such applications, and it is NOT RECOMMENDED that the experiment be conducted in the Internet. If possible, the experimental configuration will address interoperability issues, such as, for example, the use of different constraint models across different traffic domains.

The experiment can monitor various measures of quality of service before and after deployment of GCAC, particularly when the experimental network is under stress during an overload or failure condition. These quality-of-service measures might include, for example, dropped packet rate and end-to-end packet delay. The results of such experiments may be fed back to the IETF community to refine this document and to move it to the Standards Track (probably within the MPLS working group) if the experimental results are positive.

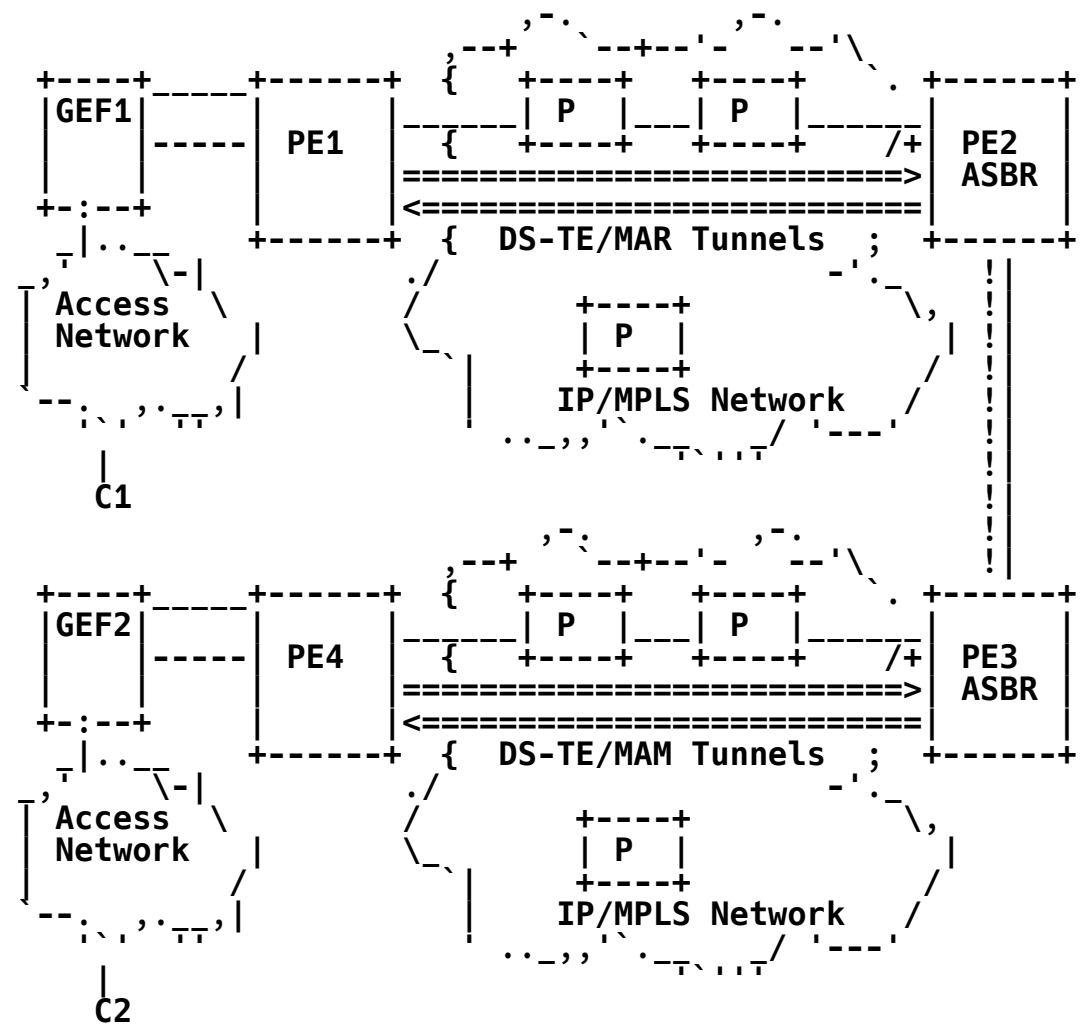
It should be noted that the algorithm might have negative effects on live deployments if the experiment is a failure. Effects might include blockage of traffic that would normally be handled or congestion caused by allowing excessive traffic on a link. For these reasons, experimentation in production networks needs to be treated with caution as described above and should only be carried out after successful simulation and experimentation in test environments. In Section 2, we describe the MPLS GCAC reference model, and in Section 3, we specify the MPLS GCAC algorithm based on the principles in the reference model and requirements. Appendix A gives an example of MPLS GCAC implementation including path selection, bandwidth management, QoS signaling, and queuing implementation.

## 1.1. Conventions Used in This Document

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 RFC 2119 [RFC2119].

## 2. MPLS GCAC Reference Model and Algorithm Summary

Figure 1 illustrates the reference model for the MPLS GCAC algorithm:




CUSTOMER I/F PARAMETERS: BW, QoS, CoS, priority

NOTE: PE, P, ASBR, GEF elements all support GCFs

```
LEGEND:
---------
ASBR:  Autonomous System Border Router
BW:    bandwidth
CoS:   class of service
DS-TE: Diffserv-aware MPLS Traffic Engineering
GCAC:  generic connection admission control
GCF:   GCAC core function
GEF:   GCAC edge function
I/F:   interface
MAM:   Maximum Allocation Model
MAR:   Maximum Allocation with Reservation
P:     provider router
PE:    provider edge router
---    connection signaling
___    bearer/media flows
```

Figure 1: MPLS GCAC Reference Model

MPLS GCAC is applicable to any service or flow for which MPLS GCAC is required to meet a given QoS. As such, the reference model applies to most real-time/RTP services (voice, video, etc.) as well as some non-real-time services. Real-time/RTP services are typically interactive, relatively persistent traffic flows. Non-real-time applications subject to MPLS GCAC could include, for example, manually provisioned LSPs or PWs and automatic bandwidth assignment for applications that automatically build LSP meshes among PE routers. The reference model also applies to MPLS GCAC when MPLS is used in access networks, which include, for example, slow-speed access networks and broadband DSL, cable, and fiber access networks. Endpoints will be IP/PBXs (Private Branch Exchanges) and individual-usage SIP/RTP end devices (hard and soft SIP phones, Integrated Access Devices (IADs)). This traffic will enter and leave the core via possibly bandwidth-constrained access networks, which may also be MPLS aware but may use some other admission control technology.

The basic elements considered in the reference model are the MPLS GCAC edge function (GEF), GCAC core functions (GCFs), the PE routers, Autonomous System Border Routers (ASBRs), and provider (P) routers. As illustrated in Figure 1, the GEF interfaces to the application at the source and destination PE, and the GCF exists at the PE, P, and ASBR routers. GEF has an end-to-end focus and deals with whether individual connection requests fit within an MPLS tunnel, and GCF has a hop-by-hop focus and deals with whether an MPLS tunnel can be established across specific core network elements on a path. The GEF functionality may be implemented in the PE, ASBR, or a stand-alone network element. The source/destination routers (or external devices

through a router interface) support both GEF and GCF, while internal routers (or external devices through a router interface) support GCF. In Figure 1, the GEF handles both signaling and bearer control.

## 2.1. Inputs to MPLS GCAC

Inputs to the GEF and GCF include the following, where most are inputs to both GEF and GCF, except as noted. Most of the parameters apply to the specific flow/LSP being calculated, while some parameters, such as request type, apply to the calculation method. Required inputs are marked with (*); all other inputs are optional:

Traffic Description
- * Bandwidth per DS-TE class type [RFC4124] (GEF, GCF)
- * Bandwidth for LSP from [RFC3270] (GEF, GCF)
- * Aggregated RSVP bandwidth requirements from [RFC4804] (GEF)
- Variance Factor (GEF, GCF)

Class of Service (CoS) and Quality of Service (QoS)
- * Class Type (CT) from [RFC4124] (GEF, GCF)
- Signaled or configured Traffic Class (TC) [RFC5462] to Per Hop Behavior (PHB) mapping from [RFC3270] (GEF, GCF)
- Signaled or configured PHB from [RFC3270] (GEF, GCF)
- QoS requirements from NSIS/Y.1541 [RFC5971][RFC5974][RFC5975] [RFC5976] (GEF)

Priority
- Admission priority (high, normal, best effort) from NSIS/Y.1541 [RFC5971][RFC5974][RFC5975][RFC5976] (GEF, GCF)
- Preemption priority from [RFC4124] (GEF, GCF)

Request type
- Primary tunnel (GEF, GCF)
- Backup tunnel and fraction of capacity reserved for backup (GEF, GCF)

Oversubscription method (see [RFC3270])
- Over/undersubscribe requested capacity (GEF, GCF)
- Over/undersubscribe available bandwidth (GEF, GCF)

These inputs can be received by the GEF and GCF from a signaling interface (such as SIP or H.323), RSVP, or an NMS. They can also be derived from measured traffic levels or from elsewhere.

## 2.2. MPLS GCAC Algorithm Summary

Figure 1 is a reference model for MPLS GCAC and illustrates the GEF to GEF MPLS GCAC algorithm to determine whether there is sufficient bandwidth to complete a connection. The originating GEF receives a connection request including the above input parameters over the input interface, for example, via an RSVP bandwidth request as specified in [RFC4804]. The GEF a) determines whether there is enough bandwidth on the route between the originating and terminating GEFs via routing and signaling communication with the GCFs at the P, PE, and ASBR network elements along the path to accommodate the connection, b) communicates the accept/reject decision on the input interface for the connection request, and c) keeps account of network resource allocations by tracking bandwidth use and allocations per CoS. Optionally, the GEF may dynamically adjust the tunnel size by signaling communication with the GCFs at nodes along the candidate paths. For example, the GEF could a) maintain per-CoS tunnel capacity based on aggregated connection requests and respond on a connection-by-connection basis based on the available capacity, b) periodically adjust the tunnel capacity upward, when needed, and downward when spare capacity exists in the tunnel, and c) use a 'make before break' mechanism to adjust tunnel capacity in order to minimize disruption to the bearer traffic.

In the reference model, DS-TE [RFC4124] tunnels are configured between the GEFs based on the traffic forecast and current network utilization. These guaranteed bandwidth DS-TE tunnels are created using RSVP-TE [RFC3209]. DS-TE bandwidth constraints models are applied uniformly within each domain, such as the Maximum Allocation with Reservation (MAR) Bandwidth Constraints Model [RFC4126], the Maximum Allocation Model (MAM) [RFC4125], and the Russian Dolls Model (RDM) [RFC4127]. An IGP such as OSPF or IS-IS is used to advertise bandwidth availability by CT for use by the GCF to determine MPLS tunnel bandwidth allocation and admission on core (backbone) links. These DS-TE tunnels are configured based on the forecasted traffic load, and when needed, LSPs for different CTs can take different paths.

As described in Section 3, the unreserved link bandwidth on CTc on link k (ULBCck) is the only bandwidth allocation parameter that must be available to the MPLS GCAC algorithm. In the case that a connection is set up across multiple service provider networks, i.e., across multiple routing domains/autonomous systems (ASes), there are several options to enable MPLS GCAC to be implemented:

1. Use [OIF-E-NNI] to advertise ULBCck parameters to the originating GEF, for the full topology of adjacent domains/areas/ASes, as described in Section 3.3.2.1.2 of [OIF-E-NNI]. Note that the

option of abstract node summarization described in [OIF-E-NNI] will not suffice since the process of summarization results in loss of topology and capacity usage information. In this manner, the originating GEF can implement the MPLS GCAC algorithm described in Section 3 across multiple domains/areas/ASes.

2. Use [BGP-TE] to advertise ULBCck parameters via BGP to the originating GEF for the full topology of adjacent domains/areas/ASes. In this manner, the originating GEF can implement the MPLS GCAC algorithm described in Section 3 across multiple domains/areas/ASes. However, network providers may be reluctant to divulge full topology and capacity usage information to other providers. Furthermore, [BGP-TE] was never intended to provide full TE topology distribution across ASBRs. Such a mechanism would be neither stable nor scalable.

3. Use individual AS control and MPLS crankback [RFC4920] to retain originating GEF control. For example, in Figure 1, if a connection crosses the two ASes shown (call them AS1 and AS2), the source GEF1 applies the GCAC algorithm described in Section 3 to the links in AS1, that is, between PE1 and PE2/ASBR in Figure 1. Then, in AS2, the GCF in PE3/ASBR applies the MPLS GCAC algorithm to the links in AS2, that is, between PE3 and PE4 in Figure 1. If the flow is rejected in AS2, crankback signaling is used to inform GEF1. In routing a connection across multiple ASes, e.g., across AS1-->AS2-->AS3, if the flow is rejected, say in AS2, the originating GEF1 can seek an alternate route, perhaps AS1-->AS4-->AS3. This option does not achieve full originating GEF control with the desired full topology visibility across ASes but avoids possible issues with obtaining full topology visibility across ASes.

4. Use Path Computation Elements (PCEs) [RFC4655] across multiple ASes. PCEs could potentially execute the GCAC algorithm within each AS and communicate/interwork across domains to determine which high-level path can supply the requested bandwidth.

In the reference model, the GEFs implement RSVP aggregation [RFC4804] for scalability. The GEF RSVP aggregator keeps a running total of bandwidth usage on the DS-TE tunnel, adding the bandwidth requirements during connection setup and subtracting during connection teardown. The aggregator determines whether or not there is sufficient bandwidth for the connection from that originating GEF to the destination GEF. The destination GEF also checks whether there is enough bandwidth on the DS-TE tunnel from the destination GEF to the originating GEF. The aggregate bandwidth usage on the DS-TE tunnel is also available to the DS-TE bandwidth constraints model. If the available bandwidth is insufficient, then the GEF sends a PATH

message through the tunnel to the other end, requesting bandwidth using GCFs, and if successful, the source would then complete a new explicit route with a PATH message along the path with increased bandwidth, again invoking GCFs on the path. If the size of the DS-TE tunnel cannot be increased on the primary and alternate LSPs, then when the DS-TE tunnel bandwidth is exhausted, the GEF aggregator sends a message to the endpoint denying the reservation. If the DS-TE tunnels are underutilized, the tunnel bandwidth may be reduced periodically to an appropriate level. In the case of a basic single class TE scenario, there is a single TE tunnel rather than multiple-CT DS-TE tunnels; otherwise, the above GCAC functions remain the same.

Optionally, the reference model implements separate queues with Diffserv based on Traffic Class (TC) bits [RFC5462]. For example, these queues may include two Expedited Forwarding (EF) priority queues, with the highest priority assigned to Emergency Telecommunications Service (ETS) traffic and the second priority assigned to normal-priority real-time traffic (alternatively, there could be a single EF queue with dual policers [RFC5865]). Several Assured Forwarding (AF) queues may be used for various data traffic, for example, premium private data traffic and premium public data traffic. A separate best-effort queue may be used for the best-effort traffic. Several DS-TE tunnels may share the same physical link and therefore share the same queue.

The MPLS GCAC algorithm increases the likelihood that the route selected by the GEF will succeed, even when the LSP traverses multiple service provider networks.

Path computation is not part of the GCAC algorithm; rather, it is considered as a vendor proprietary function, although standard IP/MPLS functions may be included in path computation, such as the following:

a) Path Computation Element (PCE) [RFC4655][RFC4657][RFC5440] to implement inter-area/inter-AS/inter-SP path selection algorithms, including alternate path selection, path reoptimization, backup path computation to protect DS-TE tunnels, and inter-area/inter-AS/inter-SP traffic engineering.

b) Backward-Recursive PCE-Based Computation (BRPC) [RFC5441].

c) Per-Domain Path Computation [RFC5152].

d) MPLS fast reroute [RFC4090] to protect DS-TE LSPs against failure.

e) MPLS crankback [RFC4920] to trigger alternate path selection and enable explicit source routing.

## 3. MPLS GCAC Algorithm

MPLS GCAC is performed at the GEF during the connection setup attempt phase to determine if a connection request can be accepted without violating existing connections' QoS and throughput requirements. To enable routing to produce paths that will likely be accepted, it is necessary for nodes to advertise some information about their internal CAC states. Such advertisements should not require nodes to expose detailed and up-to-date CAC information, which may result in an unacceptably high rate of routing updates. MPLS GCAC advertises CAC information that is generic (e.g., independent of the actual path selection algorithms used) and rich enough to support any CAC.

MPLS GCAC defines a set of parameters to be advertised and a common admission interpretation of these parameters. This common interpretation is in the form of an MPLS GCAC algorithm to be performed during MPLS LSP path selection to determine if a link or node can be included for consideration. The algorithm uses the advertised MPLS GCAC parameters (available from the topology database) and the characteristics of the connection being requested (available from QoS signaling) to determine if a link/node will likely accept or reject the connection. A link/node is included if the MPLS GCAC algorithm determines that it will likely accept the connection and excluded otherwise.

### 3.1. Bandwidth Allocation Parameters

MPLS GCAC bandwidth allocation parameters for each DS-TE CT are as defined within DS-TE [RFC4126], OSPF-TE extensions [RFC4203], and IS-IS-TE extensions [RFC5307]. The following parameters are available from DS-TE/TE extensions, advertised by the IGP, and available to the GEF and GCF [RFC4124]. Note that the approach presented in this section is adapted from [PNNI], Appendix B.

MRBk Maximum reservable bandwidth on link k specifies the maximum bandwidth that may be reserved; this may be greater than the maximum link bandwidth, in which case the link may be oversubscribed.

BWCck Bandwidth constraint for CTc on link k = allocated (minimum guaranteed) bandwidth for CTc on link k.

ULBCck Unreserved link bandwidth on CTc on link k specifies the amount of bandwidth not yet reserved for CTc.

Note that BWCck and ULBCck are the only DS-TE parameters flooded by the IGP [RFC4124][RFC4203][RFC5307]. For example, when bandwidth reservation is used [RFC4126], ULBCck is calculated and flooded by the IGP as follows:

RBTk Reservation bandwidth threshold for link k.

ULBCck Unreserved link bandwidth on CTc on link k specifies the amount of bandwidth not yet reserved for CTc, taking RBTk into account,

ULBCck = ULBk - delta0/1(CTck) * RBTk
where
delta0/1(CTck) = 0 if RBWck < BWCck
delta0/1(CTck) = 1 if RBWck >= BWCck

Also derivable at the GEF and GCF is MRBCck, the maximum reservable link bandwidth for CTc. For example, when bandwidth reservation is used [RFC4126], MRBCck is calculated as follows:

MRBCck Maximum reservable link bandwidth for CTc on link k specifies the amount of bandwidth not yet reserved for CTc.

MRBCck = MRBk - delta0/1(CTck) * RBTk
where
delta0/1(CTck) = 0 if RBWck < BWCck
delta0/1(CTck) = 1 if RBWck >= BWCck

Note that these bandwidth parameters must be configured in a consistent way within domains and across domains. GEF routing of LSPs is based on ULBCck, where ULBk is available and RBTk can be accounted for by configuration, e.g., RBTk typically = .05 * MRBk.

Also available are administrative weight (denoted as "link cost" in [RFC2328]), TE metric [RFC3630], and administrative group (also called color) 4-octet mask [RFC3630].

The following quantities can be derived from information advertised by the IGP and otherwise available to the GEF and GCF:

RBWck Reserved bandwidth on CTc on link k (0 <= c <= MaxCT-1).

RBWck = total amount of bandwidth reserved by all established LSPs that belong to CTc
RBWck = BWCck - ULBCck.

ULBk Unreserved link bandwidth on link k specifies the amount of bandwidth not yet reserved for any CT.

ULBk = MRBk - sum [RBWck (0 <= c <= MaxCT-1)].

The GCAC algorithm assumes that a DS-TE bandwidth constraints model is used uniformly within each domain (e.g., MAR [RFC4126], MAM [RFC4125], or RDM [RFC4127]). European Advanced Networking Test Center (EANTC) testing [EANTC] has shown that interoperability is problematic when different DS-TE bandwidth constraints models are used by different network elements within a domain. Specific testing of MAM and RDM across different vendor equipment showed the incompatibility. However, while the characteristics of the 3 DS-TE bandwidth constraints models are quite different, it is necessary to specify interworking between them even though it could be complex.

The following parameters are also defined and available to GCF and are assumed to be locally configured to be a consistent value across all nodes and domain(s):

SBWck Sustained bandwidth for CTc on link k (aggregate of existing connections).

SBWck = factor * RBWck where factor is configured based on standard 'demand overbooking' factors.

VFck Variance factor for CTc on link k; VFck is BWMck normalized by variance of SBWck. VFck is configured based on typical traffic variability statistics.

In many implementations of the Private Network-Network Interface (PNNI) GCAC algorithm, the variance factor is not included, or equivalently, VFck is assumed to be zero. A simplified MPLS GCAC algorithm is also derived assuming VFck = 0.

Note that different demand overbooking factors can be specified for each CT, e.g., no overbooking might be used for constant bitrate services, while a large overbooking factor might be used for bursty variable bitrate services. We specify demand overbooking rather than link overbooking for the GCAC algorithm; one advantage is the demand overbooking is compatible with source routing used by the GCAC algorithm.

Also defined is

BWMck bandwidth margin for CTc on link k; BWMck = RBWck - SBWck

GEF uses BWCck, RBWck, ULBCck, SBWck, BWMck, and VFck for LSP/IGP routing. GEF also needs to track per-CT LSP bandwidth allocation and reserved bandwidth parameters, which are defined as follows:

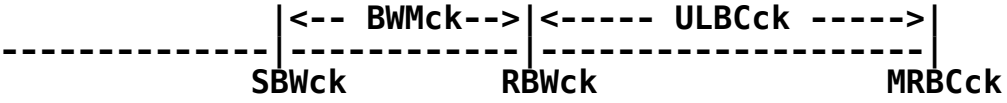
RBWLcl reserved bandwidth for CTc on LSP l

UBWLcl unreserved bandwidth for CTc on LSP l

## 3.2. GCAC Algorithm

The assumption behind the MPLS GCAC is that the ratio between the bandwidth margin that the node is putting above the sustained bandwidth and the standard deviation of the sustained bandwidth does not change significantly as one new aggregate flow is added on the link. Any ingress node doing path selection can then compute the new standard deviation of the aggregate rate (from the old value and the aggregate flow's traffic descriptors) and an estimate of the new BWMck. From this, the increase in bandwidth required to carry the new aggregate flow can be computed and compared to BWCck.

To expand on the discussion above, let RBWck denote the reserved bandwidth capacity, i.e., the amount of bandwidth that has been allocated to existing aggregate flows for CTc on link k by the actual CAC used in the node. BWMck is the difference between RBWck and the aggregate sustained bandwidth (SBWck) of the existing aggregate flows. SBWck can be either the sum of existing aggregate flows' declared sustainable bandwidth (SBWi for aggregate flow i) or a smaller (possibly measured or estimated) value. Let MRBCck denote the maximum reservable bandwidth that is usable by aggregate flows for CTc on link k. The following diagram illustrates the relationship among MRBCck, RBWck, BWMck, SBWck, and ULBCck:




The assumption is that BWMck is proportional to some measure of the burstiness of the traffic generated by the existing aggregate flows, this measure being the standard deviation of the aggregate traffic rate defined as the square root of the sum of SBWi(PBWi - SBWi) over all existing aggregate flows, where SBWi and PBWi are declared sustainable and peak bandwidth for aggregate flow i, respectively. This assumption is based on the simple argument that RBWck needs to be some multiple of the standard deviation above the mean aggregate

traffic rate to guarantee some level of packet loss ratio and packet queuing time. Depending on the actual CAC used, the BWMck-to-standard-deviation ratio may vary as aggregate flows are established and taken down. It is reasonable to assume, however, that with a sufficiently large value of RBWck, this ratio will not vary significantly. What this means is a link can advertise its current BWMck-to-standard-deviation ratio (actually in the form of VF, which is the square of this number), and the MPLS GCAC algorithm can use this number to estimate how much bandwidth is required to carry an additional aggregate flow.

Following the derivation given in [PNNI], Appendix B, the MPLS GCAC algorithm is derived as follows. Consider an aggregate flow bandwidth change request DBWi with peak bandwidth PBWi and sustainable bandwidth SBWi and a link with the following MPLS GCAC parameters: ULBCck, BWMck, and VFck for CTc and link k. Denote the variance (i.e., square of standard deviation) of the aggregate traffic rate by VARk (not advertised). Denote other unadvertised MPLS GCAC quantities by RBWck and SBWck. Then,

$$VARk = \sum_{\text{over existing aggregate flows } i} SBWi*(PBWi-SBWi) \quad (1)$$

and

$$VFck = \frac{BWMck**2}{VARk} \quad (2)$$

Using the above equation, VARk can be computed from the advertised VFck and BWMck as:

VARk = (BWMck**2)/VFck.

Let DBWi be the additional bandwidth capacity needed to carry the flow within aggregate sustainable bandwidth SBWi. The MPLS GCAC algorithm basically computes DBWi from the advertised MPLS GCAC parameters and the new aggregate flow's traffic descriptors, and compares it with ULBCck. If ULBCck >= DBWi, then the link is included for path selection consideration; otherwise, it is excluded, i.e.,

If (ULBCck >= DBWi), then include link k; else exclude link k (3)

Let BWMcknew denote the bandwidth margin if the new aggregate flow were accepted. Denote other 'new' quantities by RBWcknew, SBWcknew, and VARnew. Then,

DBWi = BWMcknew - BWMck + SBWi (4)

since BWMcknew = RBWcknew - SBWcknew, BWMck = RBWck - SBWck, and SBWcknew - SBWck = SBWi. Substituting (4) into (3), rearranging terms, and squaring both sides yield:

If ((ULBCck+BWMck-SBWck)**2 >= BWMcknew**2), then include link k;
else exclude link k (5)

Using the MPLS GCAC assumption made earlier, BWMcknew**2 can be computed as:

BWMcknew**2 = VFck * VARnew, (6)

Where

VARnew = VARk + SBWck * (PBWi-SBWi). (7)

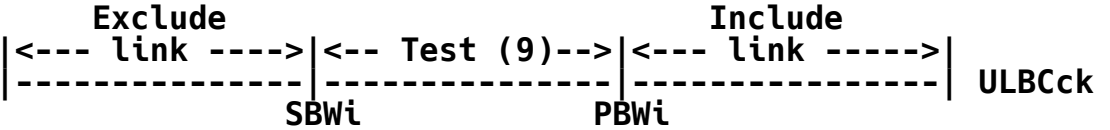
Substituting (2), (6) and (7) into (5) yields:

If ((ULBCck+BWMck-SBWi)**2 >= BWMck**2 + VFck*SBWi(PBWi-SBWi)),
then include link k;
else exclude link k (8)

and moving BWMck**2 to the left-hand side and rearranging terms yield

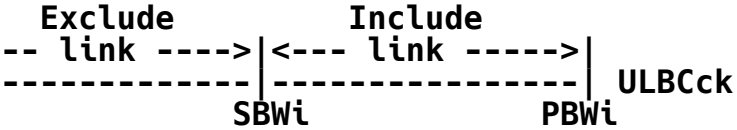
If ((ULBCck-SBWi) * (ULBCck-SBWi+2*BWMck) >= VFck*SBWi(PBWi-SBWi),
then include link k;
else exclude link k (9)

Equation (9) represents the Constrained Shortest Path First (CSPF) method implemented by most vendors and deployed by most service providers in MPLS networks. In general, DBWi is between SBWi and PBWi. So, the above test is not necessary for the cases ULBCck >= PBWi and ULBCck < SBWi. In the former case, the link is included; in the latter case, the link is excluded.



Note that VF and BWM are frequently not implemented; equivalently, these quantities are assumed to be zero, in which case Equation (9) becomes

If (ULBCck >= SBWi), then include link k; else exclude link k (10)




PNNI GCAC implementations often do not incorporate the variance factor VF, in which case Equation (10) is used.

MPLS GCAC must not reject a best-effort (BE, unassigned bandwidth) aggregate flow request based on bandwidth availability, but it may reject based on other reasons such as the number of BE flows exceeding a chosen threshold. MPLS GCAC defines only one parameter for the BE service category -- maximum bandwidth (MBW) -- to advertise how much capacity is usable for BE flows. The purpose of advertising this parameter is twofold: MBW can be used for path optimization, and MBW = 0 is used to indicate that a link is not accepting any (additional) BE flows.

Demand overbooking of LSP bandwidth is employed and must be compliant with [RFC4124] and [RFC3270] to over-/undersubscribe requested capacity. It is simplest to use only one oversubscription method, i.e., the GCAC algorithm assumes oversubscription of demands per CT, both within domains and for interworking between domains. The motivation is that interworking may be infeasible between domains if different overbooking models are used. Note that the same assumption was made for DS-TE bandwidth constraints models, in that the GCAC algorithm assumes a consistent DS-TE bandwidth constraints model is used within each domain and interoperability of bandwidth constraints models across domains.

## 4. Security Considerations

It needs to be clearly understood that all routers contain local and implementation-specific rules (or algorithms) to help them determine what to do with traffic that exceeds capacity and how to admit new flows. If these rules are poorly designed or implemented with defects, then problems may be observed in the network. Furthermore, the implementation of such algorithms provides a mechanism for attacking the delivery of traffic within the network. In view of this, routers and their software are usually extensively tested before deployment, router vendors are extended a degree of trust, and a "compromised router" (i.e., one on which an attacker has installed

their own code) is considered a weak spot in the system. Note that if a router is compromised, it can be made to do substantially more problematic things than simply modifying the admission control algorithm. Implementers are RECOMMENDED to ensure that software modifications to routers are fully secured, and operators are RECOMMENDED to apply security measures (that are outside the scope of this document) to prevent unauthorized updates to router software. Nothing in this document suggests any change to normal software security practices.

The use of a GCAC priority parameter raises possibilities for theft-of-service attacks because users could claim an emergency priority for their flows without real need, thereby effectively preventing serious emergency calls to get through. Several options exist for countering such user attacks at the interface to the user, for example:

- Only some user groups (e.g., police) are authorized to set the emergency priority bit using a policy applied to RSVP-TE signaling.

- Any user is authorized to employ the emergency priority bit for particular destination addresses (e.g., police) using a policy applied to RSVP-TE signaling.

- If an attack occurs, the user/group and actions taken should be logged to trace the attack.

- [RFC5069] identifies a number of security threats against emergency call marking and mapping. Section 6 of [RFC5069] lists security requirements to counter these threats, and those requirements should be followed by implementations of this document.

- The security requirements listed in Section 11 of [RFC4412] should be followed. These requirements apply to use of the Communications Resource Priority Header for the Session Initiation Protocol (SIP) and concern aspects of authentication and authorization, confidentiality and privacy requirements, protection against denial-of-service attacks, and anonymity.

Within the network, the policy and integrity mechanisms already present in RSVP-TE [RFC3209] can be used to ensure that the MPLS LSP has the right policy and security credentials to assume the signaled priority and bandwidth. Further discussion of this topic for the signaling of priority levels using RSVP can be found in [RFC6401]. Some similarities may also be drawn to the security issues

surrounding the placement of emergency calls in Internet multimedia systems [RFC5069] although the concepts are only comparable at the highest levels.

Like any algorithm, the algorithm specified in this document operates on data that is supplied as input parameters. That data is assumed to be collected and stored locally (i.e., on the router performing the algorithm). It is a fundamental assumption of the secure operation of any router that the data stored on that router cannot be externally modified. In this particular case, it is important that the input parameters to the algorithm cannot be influenced by an outside party. Thus, as with all configuration parameters on a router, the implementer MUST supply and the operator is RECOMMENDED to use security mechanisms to protect writing of the configuration parameters for this algorithm.

## 5. Acknowledgements

The authors greatly appreciate Adrian Farrel's support in serving as the sponsoring Area Director for this document and for his valuable comments and suggestions on the document. The authors also greatly appreciate Young Lee serving as the document shepherd and his valuable comments and suggestions. Finally, Robert Sparks' thorough review and helpful suggestions are sincerely appreciated.

## 6. Normative References

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

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3270] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC4124] Le Faucheur, F., Ed., "Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering", RFC 4124, June 2005.

[RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[RFC4804] Le Faucheur, F., Ed., "Aggregation of Resource ReSerVation Protocol (RSVP) Reservations over MPLS TE/DS-TE Tunnels", RFC 4804, February 2007.

[RFC4920] Farrel, A., Ed., Satyanarayana, A., Iwata, A., Fujita, N., and G. Ash, "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", RFC 4920, July 2007.

[RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

## 7. Informative References

[BGP-TE] Gredler, H., Farrel, A., Medved, J., and S. Previdi, "North-Bound Distribution of Link-State and TE Information using BGP", Work in Progress, March 2012.

[EANTC] "Multi-vendor Carrier Ethernet Interoperability Event", Carrier Ethernet World Congress 2006, Madrid Spain, September 2006.

[FEEDBACK] Ashwood-Smith, P., Jamoussi, B., Fedyk, D., and D. Skalecki, "Improving Topology Data Base Accuracy with Label Switched Path Feedback in Constraint Based Label Distribution Protocol", Work in Progress, June 2003.

[OIF-E-NNI] Optical Interworking Forum (OIF), "External Network-Network Interface (E-NNI) OSPFv2-based Routing - 2.0 (Intra-Carrier) Implementation Agreement", IA # OIF-ENNI-OSPF-02.0, July 13, 2011.

[PNNI] ATM Forum Technical Committee, "Private Network-Network Interface Specification Version 1.1 (PNNI 1.1)", af-pnni-0055.002, April 2002.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[RFC3246] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, March 2002.

[RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[RFC4125] Le Faucheur, F. and W. Lai, "Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering", RFC 4125, June 2005.

[RFC4126] Ash, J., "Max Allocation with Reservation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering & Performance Comparisons", RFC 4126, June 2005.

[RFC4127] Le Faucheur, F., Ed., "Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering", RFC 4127, June 2005.

[RFC4412] Schulzrinne, H. and J. Polk, "Communications Resource Priority for the Session Initiation Protocol (SIP)", RFC 4412, February 2006.

[RFC4655] Farrel, A., Vasseur, JP., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC4657] Ash, J., Ed., and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.

[RFC5069] Taylor, T., Ed., Tschofenig, H., Schulzrinne, H., and M. Shanmugam, "Security Threats and Requirements for Emergency Call Marking and Mapping", RFC 5069, January 2008.

[RFC5152] Vasseur, JP., Ed., Ayyangar, A., Ed., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.

[RFC5440] Vasseur, JP., Ed., and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

[RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, February 2009.

[RFC5865] Baker, F., Polk, J., and M. Dolly, "A Differentiated Services Code Point (DSCP) for Capacity-Admitted Traffic", RFC 5865, May 2010.

[RFC5971] Schulzrinne, H. and R. Hancock, "GIST: General Internet Signalling Transport", RFC 5971, October 2010.

[RFC5974] Manner, J., Karagiannis, G., and A. McDonald, "NSIS Signaling Layer Protocol (NSLP) for Quality-of-Service Signaling", RFC 5974, October 2010.

[RFC5975] Ash, G., Ed., Bader, A., Ed., Kappler, C., Ed., and D. Oran, Ed., "QSPEC Template for the Quality-of-Service NSIS Signaling Layer Protocol (NSLP)", RFC 5975, October 2010.

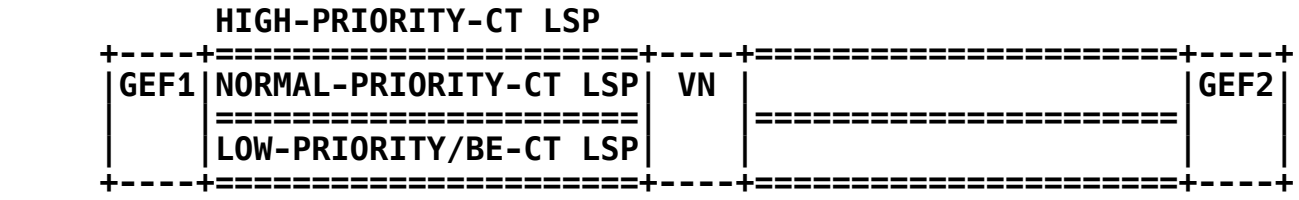
[RFC5976] Ash, G., Morton, A., Dolly, M., Tarapore, P., Dvorak, C., and Y. El Mghazli, "Y.1541-QOSM: Model for Networks Using Y.1541 Quality-of-Service Classes", RFC 5976, October 2010.

[RFC6401] Le Faucheur, F., Polk, J., and K. Carlberg, "RSVP Extensions for Admission Priority", RFC 6401, October 2011.

[TQO] Ash, G., "Traffic Engineering and QoS Optimization of Integrated Voice and Data Networks", Elsevier, 2006.

## Appendix A: Example MPLS GCAC Implementation Including Path Selection, Bandwidth Management, QoS Signaling, and Queuing

Figure 2 illustrates an example of the integrated voice/data MPLS GCAC method in which bandwidth is allocated on an aggregated basis to the individual DS-TE CTs. In the example method, CTs have different priorities including high-priority, normal-priority, and best-effort-priority services CTs. Bandwidth allocated to each CT is protected by bandwidth reservation methods, as needed, but bandwidth is otherwise shared among CTs. Each originating GEF monitors CT bandwidth use on each MPLS LSP [RFC3031] for each CT, and determines when CT LSP bandwidth needs to be increased or decreased. In Figure 2, changes in CT bandwidth capacity are determined by GEFs based on an overall aggregated bandwidth demand for CT capacity (not on a per-connection/per-flow demand basis). Based on the aggregated bandwidth demand, GEFs make periodic discrete changes in bandwidth allocation, that is, they either increase or decrease bandwidth on the LSPs constituting the CT bandwidth capacity. For example, if aggregate flow requests are made for CT LSP bandwidth that exceeds the current DS-TE tunnel bandwidth allocation, the GEF initiates a bandwidth modification request on the appropriate LSP(s). This may entail increasing the current LSP bandwidth allocation by a discrete increment of bandwidth denoted here as DBW, where DBW is the additional amount needed by the current aggregate flow request. The bandwidth admission control for each link in the path is performed by the GCF based on the status of the link using the bandwidth allocation procedure described below, where we further describe the role of the different parameters (such as the reserved bandwidth threshold RBT shown in Figure 2) in the admission control procedure. Also, the GEF periodically monitors LSP bandwidth use, and if bandwidth use falls below the current LSP allocation, the GEF initiates a bandwidth modification request to decrease the LSP bandwidth allocation to the current level of bandwidth utilization.




LEGEND
---
BE - Best Effort
CT - Class Type
GEF - GCAC Edge Function
LSP - Label Switched Path
VN - Via Node

- Distributed bandwidth allocation method applied on a per-class-type (CT) LSP basis

- GEF allocates bandwidth to each CTc LSP based on demand

  - GEF decides CTc LSP bandwidth increase based on

    + aggregate flow sustained bandwidth (SBWi) and variance factor VFck

    + routing priority (high, normal, best effort)

    + CTc reserved bandwidth (RBWck) and bandwidth constraint (BWCck)

    + link reserved bandwidth threshold (RBTk) and unreserved bandwidth (ULBk)

  - GEF periodically decreases CTc LSP bandwidth allocation based on bandwidth use

- VNs send crankback messages to GEF if DS-TE/MAR bandwidth allocation rules not met

- Link(s) not meeting request excluded from TE topology database before attempting another explicit route computation

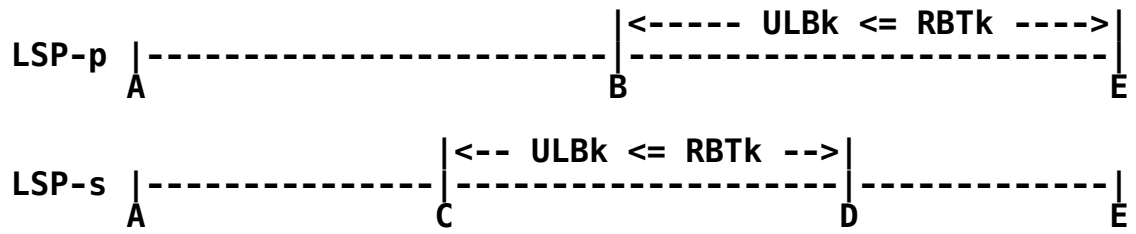
Figure 2: Per-Class-Type (CT) LSP Bandwidth Management

GEF uses SBWi, VFck, RBWck, BWCck, RBTk, and ULBk for LSP bandwidth allocation decisions and IGP routing and uses RBWcl and UBWcl to track per-CT LSP bandwidth allocation and reserved bandwidth. In making a CT bandwidth allocation modification, the GEF determines the CT priority (high, normal, or best effort), CT bandwidth-in-use, and CT bandwidth allocation thresholds. These parameters are used to determine whether network capacity can be allocated for the CT bandwidth modification request.

## A.1. Example of Path Selection and Bandwidth Management Implementation

In OSPF, link-state flooding is used to make status updates. This is a state-dependent routing (SDR) method where CSPF is typically used to alter LSP routing according to the state of the network. In general, SDR methods calculate a path cost for each connection request based on various factors such as the load state or congestion state of the links in the network. In contrast, the example MPLS GCAC algorithm uses event-dependent routing (EDR), where LSP routing is updated locally on the basis of whether connections succeed or fail on a given path choice. In the EDR learning approaches, the path that was last tried successfully is tried again until congested, at which time another path is selected at random and tried on the next connection request. EDR path choices can also be changed with time in accordance with changes in traffic load patterns. Success-to-the-top (STT) EDR path selection, illustrated in Figure 3, uses a simplified decentralized learning method to achieve flexible adaptive routing. The primary path (path-p) is used first if available, and a currently successful alternate path (path-s) is used until it is congested. In the case that path-s is congested (e.g., bandwidth is not available on one or more links), a new alternate path (path-n) is selected at random as the alternate path choice for the next connection request overflow from the primary path. Bandwidth reservation is used under congestion conditions to protect traffic on the primary path. STT-EDR uses crankback when an alternate path is congested at a via node, and the connection request advances to a new random path choice. In STT-EDR, many path choices can be tried by a given connection request before the request is rejected.

Figure 3 illustrates the example MPLS GCAC operation of STT-EDR path selection and admission control combined with per-CT bandwidth allocation. GEF A monitors CT bandwidth use on each CT LSP and determines when CT LSP bandwidth needs to be increased or decreased. Based on the bandwidth demand, GEF A makes periodic discrete changes in bandwidth allocation, that is, either increases or decreases bandwidth on the LSPs constituting the CT bandwidth capacity. If aggregate flow requests are made for CT LSP bandwidth that exceeds the current LSP bandwidth allocation, GEF A initiates a bandwidth modification request on the appropriate LSP(s).




Example of STT-EDR routing method:

1. If node A to node E bandwidth needs to be modified (say increased by DBW), primary LSP-p (e.g., LSP A-B-E) is tried first.

2. Available bandwidth is tested locally on each link in LSP-p. If bandwidth not available (e.g., unreserved bandwidth on link BE is less than the reserved bandwidth threshold and this CT is above its bandwidth allocation), crankback to node A.

3. If DBW is not available on one or more links of LSP-p, then the currently successful LSP-s (e.g., LSP A-C-D-E) is tried next.

4. If DBW is not available on one or more links of LSP-s, then a new LSP is searched by trying additional candidate paths until a new successful LSP-n is found or the candidate paths are exhausted.

5. LSP-n is then marked as the currently successful path for the next time bandwidth needs to be modified.

Figure 3: STT-EDR Path Selection and Per-CT Bandwidth Allocation

For example, in Figure 3, if the LSR-A to LSR-E bandwidth needs to be modified, say increased by DBW, the primary LSP-p (A-B-E) is tried first. The bandwidth admission control for each link in the path is performed based on the status of the link using the bandwidth allocation procedure described below, where we further describe the role of the reserved bandwidth RBWck shown in Figure 3 in the admission control procedure. If the first choice LSP cannot admit the bandwidth change, node A may then try an alternate LSP. If DBW is not available on one or more links of LSP-p, then the currently successful LSP-s A-C-D-E (the 'STT path') is tried next. If DBW is not available on one or more links of LSP-s, then a new LSP is searched by trying additional candidate paths (not shown) until a new successful LSP-n is found or all of the candidate paths held in the cache are exhausted. LSP-n is then marked as the currently

successful path for the next time bandwidth needs to be modified. DBW is set to the additional amount of bandwidth required by the aggregate flow request.

If all cached candidate paths are tried without success, the search then generates a new CSPF path. If a new CSPF calculation succeeds in finding a new path, that path is made the stored path, and the bottom cached path falls off the list. If all cached paths fail and a new CSPF path cannot be found, then the original stored LSP is retained. New requests go through the same routing algorithm again, since available bandwidth, etc., has changed and new requests might be admitted. Also, GEF A periodically monitors LSP bandwidth use (e.g., once each 2-minute interval), and if bandwidth use falls below the current LSP allocation, the GEF initiates a bandwidth modification request to decrease the LSP bandwidth allocation to the currently used bandwidth level. Bandwidth reservation occurs in STT-EDR with PATH/RESV messages per application of [RFC4804].

In the STT-EDR computation, most of the time the primary path and stored path will succeed, and no CSPF calculation needs to be done. Therefore, the STT-EDR algorithm achieves good throughput performance while significantly reducing link-state flooding control load [TQO]. An analogous method was proposed in the MPLS working group [FEEDBACK], where feedback based on failed path routing attempts is kept by the TE database and used when running CSPF.

In the example GCAC method, bandwidth allocation to the primary and alternate LSPs uses the MAR bandwidth allocation procedure, as described below. Path selection uses a topology database that includes available bandwidth on each link. From the topology database pruned of links that do not meet the bandwidth constraint, the GEF determines a list of shortest paths by using a shortest path algorithm (e.g., Bellman-Ford or Dijkstra methods). This path list is determined based on administrative weights of each link, which are communicated to all nodes within the routing domain (e.g., administrative weight = 1 + e x distance, where e is a factor giving a relatively smaller weight to the distance in comparison to the hop count). Analysis and simulation studies of a large national network model show that 6 or more primary and alternate cached paths provide the best overall performance.

PCE [RFC4655][RFC4657][RFC5440] is used to implement inter-area/inter-AS/ inter-SP path selection algorithms, including alternate path selection, path reoptimization, backup path computation to protect DS-TE tunnels, and inter-area/inter-AS/inter-SP traffic engineering. The DS-TE tunnels are protected against

failure by using MPLS Fast Reroute [RFC4090]. OSPF TE extensions [RFC4203] are used to support the TE database (TED) required for implementation of the above PCE path selection methods.

The example MPLS GCAC method incorporates the MAR bandwidth constraint model [RFC4126] incorporated within DS-TE [RFC4124]. In DS-TE/MAR, a small amount of reserved bandwidth RBTk governs the admission control on link k. Associated with each CTc on link k are the allocated bandwidth constraints BWCck to govern bandwidth allocation and protection. The reservation bandwidth on a link, RBTk, can be accessed when a given CTc has reserved bandwidth RBWck below its allocated bandwidth constraint BWCck. However, if RBWck exceeds its allocated bandwidth constraint BWCck, then the reservation bandwidth threshold RBTk cannot be accessed. In this way, bandwidth can be fully shared among CTs if available but is otherwise protected by bandwidth reservation methods. Therefore, bandwidth can be accessed for a bandwidth request = DBW for CTc on a given link k based on the following rules:

For an LSP on a high-priority or normal-priority CTc:

If RBWck = BWCc, admit if DBW = ULBk
If RBWck > BWCc, admit if DBW = ULBk - RBTk;

or, equivalently:

If DBW = ULBCck, admit the LSP.

where

ULBCck = idle link bandwidth on link k for CTc = ULBk -
delta0/1(CTck) x RBWk
delta0/1(CTck) = 0 if RBWck < BWCck
delta0/1(CTck) = 1 if RBWck = BWCck

For an LSP on a best-effort-priority CTc:

allocated bandwidth BWCc = 0;
Diffserv queuing serves best-effort packets only if there is available link bandwidth.

In setting the bandwidth constraints for CTck, for a normal-priority CTc, the bandwidth constraints (BWCck) on link k are set by allocating the maximum reservable link bandwidth (MRBk) in proportion to the forecast or measured traffic load bandwidth TRAF_LOAD_BWck for CTc on link k. That is:

PROPORTIONAL_ BWck =
TRAF_LOAD_ BWck/[S (c) {TRAF_LOAD_ BWck, c=0, MaxCT-1}] x MRBk

For a normal-priority CTc:
BWCck = PROPORTIONAL_ BWck

For a high-priority CT, the bandwidth constraint BWCck is set to a multiple of the proportional bandwidth. That is:

For high-priority CTc:
BWCck = FACTOR * PROPORTIONAL_ BWck

where FACTOR is set to a multiple of the proportional bandwidth (e.g., FACTOR = 2 or 3 is typical). This results in some over-allocation ('overbooking') of the link bandwidth and gives priority to the high-priority CTs. Normally, the bandwidth allocated to high-priority CTs should be a relatively small fraction of the total link bandwidth, a maximum of 10-15 percent being a reasonable guideline.

As stated above, the bandwidth allocated to a best-effort-priority CTc is set to zero. That is:

For a best-effort-priority CTc:
BWCck = 0

Analysis and simulation studies show that the level of reserved capacity RBTk in the range of 3-5% of link capacity provides the best overall performance.

We give a simple example of the MAR bandwidth allocation method. Assume that there are two class types, CT0 and CT1, and a particular link with

MRB = 100

with the allocated bandwidth constraints set as follows:

BWC0 = 30
BWC1 = 50

These bandwidth constraints are based on the forecasted traffic loads, as discussed above. Either CT is allowed to exceed its bandwidth constraint BWCc as long a there is at least RBW units of spare bandwidth remaining. Assume RBT = 10. So under overload, if

RBW0 = 20
RBW1 = 70

Then, for this loading

UBW = 100 - 20 - 70 = 10

If a bandwidth increase request = 5 = DBW arrives for Class Type 0 (CT0), then accept for CT0 since RBW0 < BWC0 and DBW (= 5) < ILBW (= 10).

If a bandwidth increase request = 5 = DBW arrives for Class Type 1 (CT1), then reject for CT1 since RBW1 > BC1 and DBW (= 5) > ILBW - RBT = 10 - 10 = 0.

Therefore, CT0 can take the additional bandwidth (up to 10 units) if the demand arrives, since it is below its BWC value. CT1, however, can no longer increase its bandwidth on the link, since it is above its BWC value and there is only RBT=10 units of idle bandwidth left on the link. If best effort traffic is present, it can always seize whatever idle bandwidth is available on the link at the moment but is subject to being lost at the queues in favor of the higher-priority traffic.

On the other hand, if a request arrives to increase bandwidth for CT1 by 5 units of bandwidth (i.e., DBW = 5), we need to decide whether or not to admit this request. Since for CT1,

RBW1 > BWC1 (70 > 50), and
DBW > UBW - RBT (i.e., 5 > 10 - 10)

the bandwidth request is rejected by the bandwidth allocation rules given above. Now let's say a request arrives to increase bandwidth for CT0 by 5 units of bandwidth (i.e., DBW = 5). We need to decide whether or not to admit this request. Since for CT0

RBW0 < BWC0 (20 < 30), and
DBW < UBW (i.e., 5 < 10)

The example illustrates that with the current state of the link and the current CT loading, CT1 can no longer increase its bandwidth on the link, since it is above its BWC1 value and there is only RBW=10 units of spare bandwidth left on the link. But CT0 can take the additional bandwidth (up to 10 units) if the demand arrives, since it is below its BWC0 value.

For the example GCAC, the method for bandwidth additions and deletions to LSPs in is as follows. The bandwidth constraint parameters defined in the MAR method [RFC4126] do not change based on traffic conditions. In particular, these parameters defined in [RFC4126], as described above, are configured and do not change until

reconfigured: MRBk, BWCck, and RBTk. However, the reserved bandwidth variables change based on traffic: RBWck, ULBk, and ULBCck. The RBWck and bandwidth allocated to each DS-TE/MAR tunnel is dynamically changed based on traffic: it is increased when the traffic demand increases (using RSVP aggregation), and it is periodically decreased when the traffic demand decreases. Furthermore, if tunnel bandwidth cannot be increased on the primary path, an alternate LSP path is tried. When LSP tunnel bandwidth needs to be increased to accommodate a given aggregate flow request, the bandwidth is increased by the amount of the needed additional bandwidth, if possible. The tunnel bandwidth quickly rises to the currently needed maximum bandwidth level, wherein no further requests are made to increase bandwidth, since departing flows leave a constant amount of available or spare bandwidth in the tunnel to use for new requests. Tunnel bandwidth is reduced every 120 seconds by 0.5 times the difference between the allocated tunnel bandwidth and the current level of the actually utilized bandwidth (i.e., the current level of spare bandwidth). Analysis and simulation modeling results show that these parameters provide the best performance across a number of overload and failure scenarios.

## A.2. Example of QoS Signaling Implementation

The example GCAC method uses Next Steps in Signaling (NSIS) algorithms for signaling MPLS GCAC QoS requirements of individual flows. NSIS QoS signaling has been specified by the IETF NSIS working group and extends RSVP signaling by defining a two-layer QoS signaling model:

- NSIS Transport Layer Protocol (NTLP) [RFC5971]
- NSIS Signaling Layer Protocol (NSLP) for Quality-of-Service Signaling [RFC5974]

[RFC5975] defines a QoS specification (QSPEC) object, which contains the QoS parameters required by a QoS model (QOSM) [RFC5976]. A QOSM specifies the QoS parameters and procedures that govern the resource management functions in a QoS-aware router. Multiple QOSMs can be supported by the QoS-NSLP, and the QoS-NSLP allows stacking of QSPEC parameters to accommodate different QOSMs being used in different domains. As such, NSIS provides a mechanism for interdomain QoS signaling and interworking.

The QSPEC parameters defined in [RFC5975] include, among others:

TRAFFIC DESCRIPTION Parameters:

- <Traffic Model> Parameters

CONSTRAINTS Parameters:

- <Path Latency>, <Path Jitter>, <Path PLR>, and <Path PER> Parameters
- <PHB Class> Parameter
- <DSTE Class Type> Parameter
- <Y.1541 QoS Class> Parameter
- <Reservation Priority> Parameter
- <Preemption Priority> and <Defending Priority> Parameters

The ability to achieve end-to-end QoS through multiple Internet domains is also an important requirement. MPLS GCAC end-to-end QoS signaling ensures that end-to-end QoS is met by applying the Y.1541-QOSM [RFC5976], as now illustrated.

The QoS GEF initiates an end-to-end, inter-domain QoS RESERVE message containing the QoS parameters, including for example, <Traffic Model>, <Y.1541 QoS Class>, <Reservation Priority>, and perhaps other parameters for the flow. The RESERVE message may cross multiple domains; each node on the data path checks the availability of resources and accumulating the delay, delay variation, and loss ratio parameters, as described below. If an intermediate node cannot accommodate the new request, the reservation is denied. If no intermediate node has denied the reservation, the RESERVE message is forwarded to the next domain. If any node cannot meet the requirements designated by the RESERVE message to support a QoS parameter, for example, it cannot support the accumulation of end-to-end delay with the <Path Latency> parameter, the node sets a flag that will deny the reservation. Also, parameter negotiation can be done, for example, by setting the <Y.1541 QoS Class> to a lower class than specified in the RESERVE message. When the available <Y.1541 QoS Class> must be reduced from the desired <Y.1541 QoS Class>, say because the delay objective has been exceeded, then there is an incentive to respond to the GEF with an available value for delay in the <Path Latency> parameter. For example, if the available <Path Latency> is 150 ms (still useful for many applications) and the desired QoS is 100 ms (according to the desired <Y.1541 QoS Class>

Class 0), then the response would be that Class 0 cannot be achieved and Class 1 is available (with its 400 ms objective). In addition, the response includes an available <Path Latency> = 150 ms, making acceptance of the available <Y.1541 QoS Class> more likely.

### A.3. Example of Queuing Implementation

In this MPLS GCAC example, queuing behaviors for the CT traffic priorities incorporates Diffserv mechanisms and assumes separate queues based on Traffic Class (TC)/CoS bits. The queuing implementation assumes 3 levels of priority: high, normal, and best effort. These queues include two EF priority queues [RFC3246][RFC5865], with the highest priority assigned to emergency traffic (GETS/ETS/E911) and the second priority assigned to normal-priority real-time (e.g., VoIP) traffic. Separate AF queues [RFC2597] are used for data services, such as premium private data and premium public data traffic, and a separate best-effort queue is assumed for the best-effort traffic. All queues have static bandwidth allocation limits applied based on the level of forecast traffic on each link, such that the bandwidth limits will not be exceeded under normal conditions, allowing for some traffic overload. In the MPLS GCAC method, high-priority, normal-priority, and best-effort traffic share the same network; under congestion, the Diffserv priority-queuing mechanisms push out the best-effort-priority traffic at the queues so that the normal-priority and high-priority traffic can get through on the MPLS-allocated LSP bandwidth.

## Authors' Addresses

Gerald Ash (editor)
AT&T

EMail: gash5107@yahoo.com

Dave McDysan
Verizon
22001 Loudoun County Pkwy
Ashburn, VA 20147

EMail: dave.mcdysan@verizon.com