Internet Draft

Gerald R. Ash
AT&T
Bilel Jamoussi
Nortel Networks
Young Lee
AT&T
Osama S. Aboul-Magd
Nortel Networks

February 1999

Expires: October 1999

QoS Resource Management in MPLS-Based Networks

<draft-ash-qos-routing-00.txt>

STATUS OF THIS MEMO:

COPYRIGHT NOTICE:

ABSTRACT:

Efficient QoS resource management is needed for a host of existing and ever-increasing new services. For service performance, flexibility, and reduced cost it is preferable to provide integration of these services on a shared network. Such integration and sharing is facilitated by QoS resource management techniques described in the draft which are applicable to MPLS-based networks. Such QoS resource management techniques are used in PSTNs to standardize service classification, bandwidth allocation, bandwidth protection, and priority routing treatment for all network services. In the draft we illustrate the principles of QoS resource management and describe

their application to MPLS-based networks. In the proposed QoS resource management method, bandwidth is allocated in discrete changes to each of three virtual networks (VNs) corresponding to high-priority key services, normal priority services, and best-effort lower priority services. Bandwidth changes in VN bandwidth capacity are determined by edge switch/routers based on an overall aggregated bandwidth demand for VN capacity (not on a per-connection demand basis). Based on the aggregated bandwidth demand, these edge switch/routers make periodic discrete changes in bandwidth allocation, that is, either increase or decrease bandwidth on the constraint-based routing label switched paths (CRLSPs) constituting the VN bandwidth capacity. We propose to add optional parameters in the constraint-based routing label distribution protocol (CRLDP) for QoS

resource management. In particular, we propose an optional depth-of-search (DoS) type/length/value (TLV) parameter in the CRLDP label request message to control the bandwidth allocation on individual links in a CRLSP. In addition, we propose an optional modify-TLV parameter in the CRLDP label request message to allow modification of the assigned traffic parameters (such as peak data rate, committed data rate, etc.) of an already existing CRLSP. Finally, we propose a crankback-TLV parameter in the CRLDP notification message to allow an edge switch/router to search out additional alternate CRLSPs when a given CRLSP cannot accommodate a bandwidth request. This draft addresses point-to-point QoS resource management, multipoint QoS resource management is left for future study.

****
NOTE: A MICROSOFT WORD VERSION OF THIS DRAFT (WITH THE FIGURES) IS
       AVAILABLE ON REQUEST
****

Table of Contents

QoS RESOURCE MANAGEMENT IN MPLS-BASED NETWORKS

1.0 Introduction

QoS resource management methods have been applied successfully in PSTNs over the past decade [A98], and are recommended in this draft for QoS resource management in IP multiprotocol label switching (MPLS)-based networks [RCV99]. In the proposed QoS resource management method, bandwidth is allocated in discrete changes to each of three virtual networks (VNs) corresponding to high-priority key services, normal priority services, and best-effort lower priority services. Examples of services within these VN categories include a) key priority services such as defense voice communication, b) normal priority services such as constant rate, interactive, delay-sensitive voice; variable rate, interactive, delay-sensitive IP-telephony; and variable rate, non-interactive, non-delay-sensitive WWW file transfer, and c) lower priority best effort services such as variable rate, non-interactive, non-delay-sensitive voice mail, email, and file transfer. Bandwidth changes in VN bandwidth capacity are determined by edge switch/routers based on an overall aggregated bandwidth demand for VN capacity (not on a per-connection demand basis). Based on the aggregated bandwidth demand, these edge switch/routers make periodic discrete changes in bandwidth allocation, that is, either increase or decrease bandwidth on the constraint-based routing label switched paths (CRLSPs) constituting the VN bandwidth capacity.

In the draft we propose that the bandwidth allocation control for each VN CRLSP be based on estimated bandwidth needs, bandwidth use, and status of links in the CRLSP. The edge switch/router, or originating switch/router (OSR), determines when VN bandwidth needs to be increased or decreased on a CRLSP, and uses a proposed MPLS CRLSP bandwidth modification procedure to execute needed bandwidth allocation changes on VN CRLSPs. In the bandwidth allocation procedure the constraint-based routing label distribution protocol (CRLDP) [J99, ADFF98] is used by specifying appropriate parameters in the label request message to request bandwidth allocation changes on each link in the CRLSP and to determined if link bandwidth can be allocated. If a link bandwidth allocation is not allowed, a proposed CRLDP notification message with crankback parameter allows the OSR to search out possible bandwidth allocation on another CRLSP. In particular, we propose an optional depth-of-search (DoS) type/length/value (TLV) parameter in the CRLDP label request message to control the bandwidth allocation on individual links in a CRLSP. In addition, we propose an optional modify-TLV parameter in the CRLDP label request message to allow modification of the assigned traffic parameters (such as peak data rate, committed data rate, etc.) of an already existing CRLSP. Finally, we propose a crankback-TLV parameter in the CRLDP notification message to allow an edge switch/router to search out additional alternate CRLSPs when a given CRLSP cannot accommodate a bandwidth request. This draft addresses point-to-point QoS resource management, multipoint QoS resource management is left for future study.

2.0 Definitions

Link: a bandwidth transmission medium between switches that is engineered as a unit;
Switch/router: a switching center or aggregation of switching centers representing a network;
O-D pair: an originating switch/router to destination switch/router pair for a given connection request;
Originating switch/router: originating point within a given network;
Path: a concatenation of links providing a connection between an O-D pair;
Route: a set of paths connecting the same O-D pair;
Routing table: describes the route choices and selection rules to select one path out of the set for a connection request;
Terminating switch/router: terminating point within a given network;
Traffic stream: a class of connections with the same traffic characteristics;
Via switch/router: a via point within a given network;

3.0 QoS Resource Management

Through the use of bandwidth allocation, reservation, and congestion control techniques, QoS resource management can provide good network performance under normal and abnormal operating conditions for all services sharing the integrated network. In the multi-service, QoS resource management network, bandwidth is allocated to the three individual VNs (high-priority key services VN, normal priority services VN, and best-effort lower priority services VN). This allocated bandwidth is protected as needed but otherwise shared. Each OSR monitors VN bandwidth use on each VN CRLSP, and determines when VN CRLSP bandwidth needs to be increased or decreased. Bandwidth changes in VN bandwidth capacity are determined by OSRs based on an overall aggregated bandwidth demand for VN capacity (not on a per-connection demand basis). Based on the aggregated bandwidth demand, these OSRs make periodic discrete changes in bandwidth allocation, that is, either increase or decrease bandwidth on the CRLSPs constituting the VN bandwidth capacity. For example, if connection requests are made for VN CRLSP bandwidth that exceeds the current CRLSP bandwidth allocation, the OSR initiates a bandwidth modification request on the appropriate CRLSP(s). This bandwidth modification request may entail increasing the current CRLSP bandwidth allocation by a discrete increment of bandwidth denoted here as delta-bandwidth (DBW). DBW is a large enough bandwidth change so that modification requests are made relatively infrequently. Also, the OSR periodically monitors CRLSP bandwidth use, such as once each minute, and if bandwidth use falls below the current CRLSP allocation the OSR initiates a bandwidth modification request to decrease the CRLSP bandwidth allocation by a unit of bandwidth such as DBW. In making a VN bandwidth allocation modification, the OSR determines the QoS resource management parameters including the VN priority (key, normal, or best-effort), VN bandwidth-in-use, VN bandwidth allocation thresholds, and whether the CRLSP is a first choice CRLSP or alternate CRLSP. A VN depth-of-search (DoS) table determines a DoS load state threshold, or the "depth" to which network capacity can be allocated, based on the QoS resource management parameters for the VN bandwidth modification request.

In using the DoS threshold to allocate VN bandwidth capacity, the OSR selects a first choice CRLSP based on the routing table selection rules. Path selection in the IP network may used open shortest path first (OSPF) [M98, S95] for intra-domain routing. In OSPF-based layer 3 routing, as illustrated in Figure 1, OSR A determines a list of shortest paths by using, for example, Dijsktra's algorithm. This path list could be determined based on administrative weights of each link, which are communicated to all switches/routers within the autonomous system (AS) domain. These administrative weights may be set, for example, to 1 + epsilon x distance, where epsilon is a factor giving a relatively smaller weight to the distance in comparison to the hop count. The OSR selects a path from the list based on, for example, fixed routing (FR), time dependent routing (TDR), state dependent routing (SDR), or event dependent routing (EDR) path selection [A98]. For example, in using the first CRLSP A-B-E, OSR A sends CRLDP label request message to via switch/router (VSR) B, which in turn forwards the CRLDP label request message to terminating switch/router (TSR) E. VSR B and TSR E are passed in the explicit routing CR/TLV parameter contained in the CRLDP label request message. Each switch/router in the CRLSP reads the CR/TLV information, and passes the CRLDP label request message to the next switch/router listed in the CRLSP. If the first path is blocked at any of the links in the path, a CRLDP notification message with a constraint-based routing type/length/value (CR/TLV) crankback parameter is returned to OSR A which can then attempt the next path. If FR is used, then this path is the next path in the shortest path list, for example path A-C-D-E. If TDR is used, then the next path is the next path in the routing table for the current time period. If SDR is used, OSPF implements a distributed method of flooding link status information, which is triggered either periodically and/or by crossing load state threshold values. This method of distributing link status information can be resource intensive and may not be any more efficient than simpler path selection methods such as EDR. If EDR is used, then the next path is the last successful path, and if that path is unsuccessful another alternate path is searched out according to the EDR path selection method.

Hence in using the selected CRLSP, the OSR sends the explicit routing, the requested traffic parameters (peak data rate, committed data rate, etc.), an optional DoS-TLV parameter, and an optional modify-TLV parameter in the CRLDP label request message to each VSR and the TSR in the selected CRLSP. Whether or not bandwidth can be allocated to the bandwidth modification request on the first choice CRLSP is determined by each VSR applying the QoS resource management rules. These rules entail that the VSR determine the CRLSP link states (lightly loaded, heavily loaded, reserved, or busy), based on bandwidth use and bandwidth available, and compare the link load state to the DoS threshold sent in the CRLDP TLV parameters. If the first choice CRLSP cannot be accessed, a VSR or TSR returns control to the OSR through the use of a proposed crankback-TLV parameter in the CRLDP notification message. At that point the OSR may then try an alternate CRLSP. Whether or not bandwidth can be allocated to the bandwidth modification request on the alternate path again is determined by the use of the DoS threshold compared to the CRLSP link load state at each VSR. Priority queuing is used during the time the connection is established, and at each link the queuing discipline is maintained such that the packets are given priority according to the VN traffic priority.

In the proposed method of QoS resource management, the admission control for bandwidth modification on each VN CRLSP is based on the status of the links in the CRLSP. The OSR may select any CRLSP for which the first CRLSP link is allowed according to QoS resource management criteria. If a subsequent CRLSP link is not allowed, then a proposed CRLDP notification message with a crankback-TLV parameter is used to return to the OSR and select an alternate CRLSP. Determination of the CRLSP link load states is necessary for QoS resource management to select network capacity on either the first choice CRLSP or alternate CRLSPs. Four link load states are distinguished: lightly loaded (LL), heavily loaded (HL), reserved (R), and busy (B). Selection of CRLSP capacity uses a link state model and a depth-of-search (DoS) model to determine if a bandwidth modification request can be admitted on a given CRLSP. The allowed DoS load state threshold determines if a bandwidth modification request can be admitted on a given link to an available bandwidth "depth." In setting up the bandwidth modification request, the OSR encodes the DoS load state threshold allowed on each link in the proposed DoS-TLV parameter in the CRLDP label request. If a CRLSP link is encountered at a VSR in which the idle link bandwidth and link load state are below the allowed DoS load state threshold, then the VSR sends a CRLDP notification message with a proposed crankback-TLV parameter to the OSR, which can then route the bandwidth modification request to an alternate CRLSP choice. For example, in Figure 1, CRLSP A-B-E may be the first path tried where link A-B is in the LL state and link B-E is in the R state. If the DoS load state allowed is HL or better, then the CRLSP bandwidth modification request in the CRLDP label request message is routed on link A-B but will not be admitted on link B-E, wherein the CRLSP bandwidth modification request will be cranked back in the CRLDP notification message to the originating switch A to try alternate CRLSP A-C-D-E. Here the CRLSP bandwidth modification request succeeds since all links have a state of HL or better.

The DoS load state threshold is a function of bandwidth-in-progress, service priority, and bandwidth allocation thresholds, as follows:

Table 1. Determination of Depth-of-Search (DoS) Load State Threshold

Load State Allowed	Key Priority VN	Normal Priority VN: First Choice CRLSP	Normal Priority VN: Alternate CRLSP	Best Effort Priority VN
R	if BWIPi <= 2 * BWmaxi	if BWIPi <= BWavgi	Not Allowed	Note 1
HL	if BWIPi <= 2 * BWmaxi	if BWIPi <= BWmaxi	if BWIPi <= BWavgi	Note 1
LL	All BWIPi	All BWIPi	All BWIPi	Note 1

where

```
    BWIPi          =       bandwidth-in-progress on VN i
    BWavgi         =       minimum guaranteed bandwidth required for VN
                           i to carry the average offered bandwidth load
```

BWmaxi = the bandwidth required for VN i to meet the blocking probability grade-of-service objective for CRLSP bandwidth allocation requests
= 1.1 x BWavgi

Note 1 = CRLSPs for the best effort priority VN are allocated zero bandwidth; Diffserv queuing admits best effort packets only if there is available bandwidth on a link

Note that BWIP, BWavg, and BWmax are specified per OSR-TSR pair, and that the QoS resource management method provides for a key priority VN, a normal priority VN, and a best effort VN. Key services admitted by an OSR on the key VN are given higher priority routing treatment by allowing greater path selection DoS than normal services admitted on the normal VN. Best effort services admitted on the best effort VN are given lower priority routing treatment by allowing lesser path selection DoS than normal. The quantities BWavgi are computed periodically, such as every week w, and can be exponentially averaged over a several week period, as follows:

BWavgi(w) = .5 x BWavgi(w-1) + .5 x [BWIPavgi(w) + BWOVavgi(w)]

BWIPavgi = average bandwidth-in-progress across a load set period on VN i

BWOVavgi = average bandwidth allocation request overflow across a load set period on VN i

where all variables are specified per OSR-TSR pair, and where BWIPi and BWOVi are averaged across various load set periods, such as morning, afternoon, and evening averages for weekday, Saturday, and Sunday, to obtain BWIPavgi and BWOVavgi.

Illustrative values of the thresholds to determine link load states are as follows:

Table 2. Determination of Link Load State

Name of State	Condition
Busy B	ILBWk < DBW
Reserved R	ILBWk * Rthrk
Heavily Loaded HL	Rthrk < ILBWk * HLthrk
Lightly Loaded LL	HLthrk < ILBWk

where

```
ILBWk           =       idle link bandwidth on link k
DBW             =       delta bandwidth requirement for a bandwidth
                        allocation request
Rthrk           =       reservation bandwidth threshold for link k
                =       N x .05 x TBWk for bandwidth reservation
                        level N
HLthrk          =       heavily loaded bandwidth threshold for link
                        k
                =       Rthrk + .05 x TBWk
```

TBWk	=	the total bandwidth required on link k to meet the blocking probability grade-of-service objective for bandwidth allocation requests on their first choice CRLSP.

QoS resource management implements bandwidth reservation logic to favor connections routed on the first choice CRLSP in situations of link congestion. If link blocking is detected, bandwidth reservation is immediately triggered and the reservation level N is set for the link according to the level of link congestion. In this manner bandwidth allocation requests attempting to alternate-route over a congested link are subject to bandwidth reservation, and the first choice CRLSP requests are favored for that link. At the same time, the LL and HL link state thresholds are raised accordingly in order to accommodate the reserved bandwidth capacity N for the VN. Figure 2 illustrates bandwidth allocation and the mechanisms by which bandwidth is protected through bandwidth reservation. Under normal bandwidth allocation demands bandwidth is fully shared, but under overloaded bandwidth allocation demands, bandwidth is protected through the reservation mechanisms wherein each VN can use its allocated bandwidth. Under failure, however, the reservation mechanisms operate to give the key VN its allocated bandwidth before the normal priority VN gets its bandwidth allocation. As noted on Table 1, the best effort lower priority VN is not allocated bandwidth nor is bandwidth reserved for the best effort VN. Illustrations are given in [A98] of the robustness of dynamic bandwidth reservation in protecting the preferred bandwidth requests across wide variations in traffic conditions.

The reservation level N (for example, N may have 1 of 4 levels), is calculated for each link k based on the link blocking level of bandwidth allocation requests. The link blocking level is equal to the total requested but rejected (or overflow) link bandwidth allocation (measured in total bandwidth), divided by the total requested link bandwidth allocation, over the last periodic update interval, which is typically three minutes. That is

BWOVk	=	total requested bandwidth allocation overflow on link k
BWOFk	=	total requested or offered bandwidth allocation on link k
LBLk	=	link blocking level on link k
	=	BWOVk/BWOFk

If LBLk exceeds a threshold value, the reservation level N is calculated accordingly. The reserved bandwidth and link states are calculated based on the total link bandwidth required on link k, TBWk, which is computed on-line, for example every 1-minute interval m, and approximated as follows:

```
TBWk(m)                  =       .5 x  TBWk(m-1) +
                                 .5 x [ 1.1 x  TBWIPk(m) +  TBWOVk(m) ]
TBWIPk                   =       sum of the bandwidth in progress
                                 (BWIPi) for all VNs i
                                 for bandwidth requests on their
                                 first choice CRLSP over link k
```

TBWOVk = sum of bandwidth overflow (BWOVi) for all VNs i for bandwidth requests on their first choice CRLSP over link k

Therefore the reservation level and load state boundary thresholds are proportional to the estimated required bandwidth load, which means that the bandwidth reserved and the bandwidth required to constitute a lightly loaded link rise and fall with the bandwidth load, as, intuitively, they should.

In addition to the QoS bandwidth management procedure for bandwidth allocation requests, a QoS priority of service queuing capability is used during the time connections are established on each of the three VNs. At each link, a queuing discipline is maintained such that the packets being served are given priority in the following order: key VN services, normal VN services, and best effort VN services. Following the MPLS CRLSP bandwidth allocation setup and the application of QoS resource management rules, the priority of service parameter and label parameter need to be sent in each IP packet, as illustrated in Figure 3. The priority of service parameter may be included in the type of service (ToS), or differentiated services (diffserv) [B98, ST98], parameter already in the IP packet header. Another possible alternative is that the priority of service parameter might be included in the MPLS label or "shim" appended to the IP packet (this is a matter for further study). In either case, from the priority of service parameter, the IP switch/router can determine the QoS treatment based on the QoS resource management (priority queuing) rules for key VN packets, normal VN packets, and best effort VN packets. From the label parameter, the IP switch/router can determine the next switch/router to route the IP packet to as defined by the MPLS protocol. In this way, the backbone switches/routers can have a very simple per-packet processing implementation to implement QoS resource management and MPLS routing.

4.0 Summary of CRLDP Proposals

In summary we make these proposals regarding CRLDP use in MPLS:

a) Edge switch/routers, or OSRs, monitor VN bandwidth use and decide when to make CRLSP bandwidth modification requests. OSRs keep track of VN priority, bandwidth-in-use, and bandwidth allocation thresholds and apply DoS rules to determine the DoS threshold to apply for a bandwidth modification request.
b) Backbone switch/routers, or VSRs, keep track of link state and compare DoS threshold parameters to link state (as do OSRs).
c) OSRs formulate the CRLDP label request message, which carries the explicit routing parameters specifying the VSRs and TSR in the selected CRLSP, the optional DoS-TLV parameter specifying the allowed bandwidth allocation threshold on each link in the CRLSP, and the optional modify-TLV parameter to allow modification of the assigned traffic parameters (such as

peak data rate, committed data rate, etc.) of an already existing CRLSP.
d) VSRs or TSRs formulate the optional crankback-TLV parameter in the CRLDP notification message, which specifies return of control of the link bandwidth allocation request to the OSR, for possible further alternate routing to search out additional alternate CRLSPs when a given CRLSP cannot accommodate a bandwidth request.

This draft addresses point-to-point QoS resource management, multipoint QoS resource management is left for future study.

5.0 References

[A98] Ash, G. R., Dynamic Routing in Telecommunications Networks, McGraw-Hill, 1998.

[ADFFT98] Anderson, L., Doolan, P., Feldman, N., Fredette, A., Thomas, B., LDP Specification, IETF draft-ietf-mpls-ldp-01.txt, August 1998.

[B98] Bernet, Y., et. al., A Framework for Differentiated Services, IETF draft-ietf-diffserv-framework-01.txt, October 1998.

[J99] Jamoussi, B., Editor, Constraint-Based LSP Setup using LDP, IETF draft-ietf-mpls-cr-ldp-01.txt, February 1999.

[M98] Moy, John, OSPF Version 2, IETF RFC 2328, April 1998.

[RVC99] Rosen, E., Viswanathan, A., Callon, R., Multiprotocol Label Switching Architecture, IETF draft-ietf-mpls-arch-04.txt, February 1999.

[S95] Steenstrup, M., Editor, Routing in Communications Networks, Prentice-Hall, 1995.

[ST98] Sikora, J., Teitelbaum, B., Differentiated Services for Internet 2, Internet 2: Join Applications/Engineering QoS Workshop, Santa Clara, CA, May 1998.

6.0 Abbreviations

B	Busy
BGP	Border Gateway Protocol
BW	Bandwidth
BWIP	Bandwidth in Progress
BWOF	Bandwidth Offered
BWOV	Bandwidth Overflow
CRLDP	Constraint-Based Routing Label Distribution Protocol
CRLSP	Constraint-Based Routing Label Switched Path
DIFFSERV	Differentiated Services
DoS	Depth-of-Search
HL	Heavily Loaded
IETF	Internet Engineering Task Force
ILBW	Idle Link Bandwidth
IP	Internet Protocol
LBL	Link Blocking Level
LDP	Label Distribution Protocol
LL	Lightly Loader
LSP	Label Switched Path
MPLS	Multiprotocol Label Switching
OSR	Originating Switch/Router
OSPF	Open Shortest Path First
PSTN	Public Switched Telephone Network

QoS	Quality of Service
R	Reserved
TBW	Total Bandwidth
TBWIP	Total Bandwidth In Progress
TLV	Type/Length/Value
ToS	Type of Service
TSR	Terminating Switch/Router

VSR   Via Switch/Router
VN   Virtual Network

7.0 Authors' Addresses

Gerald R. Ash
AT&T
Room MT E3-3C37
200 Laurel Avenue
Middletown, NJ 07748
Phone: 732-420-4578
Fax: 732-440-6687
Email: gash@att.com

Bilel Jamoussi
Nortel Networks
P O Box 3511 Station C
Ottawa, ON K1Y 4H7
Canada
phone: +1 613 765-4814
Email: jamoussi@NortelNetworks.com

Young Lee
AT&T
Room MT E3-3A04
200 Laurel Avenue
Middletown, NJ 07748
Phone: 732-420-4477
Fax: 732-440-6697
Email: younglee@att.com

Osama S. Aboul-Magd
Nortel Networks
P O Box 3511 Station C
Ottawa, ON K1Y 4H7
Canada
phone: +1 613 763-5827
Email: osama@NortelNetworks.com

****
NOTE: A MICROSOFT WORD VERSION OF THIS DRAFT (WITH THE FIGURES) IS

AVAILABLE ON REQUEST
****