Modeling and Analysis of Internet Differentiated Services Traffic

Muckai K Girish[a] and Jian-Qiang Hu[b]

[a]SBC Technology Resources, Inc.
4698 Willow Woad
Pleasanton, California 94588, USA

[b]Department of Manufacturing Engineering
Boston University
15 St. Mary's Street
Boston, Massachusetts 02215, USA

ABSTRACT

It is widely accepted that the next major requirement and milestone in the Internet is the ability to provide qualities of service to various applications. The Internet community and the standards bodies have been working on myriad schemes to achieve this objective. One of the prominent and promising approaches is being defined and developed at the Differentiated Services Working Group in IETF. We model the traffic associated with Assured Forwarding (AF) Per-Hop Behavior (PHB) in differentiated services enabled IP networks in this paper. Furthermore, we analyze the effect of such traffic on network resources with the objective of developing efficient traffic engineering methodologies. We also formulate the optimization problem relating to traffic engineering in DS networks with an MPLS core.

Keywords: Internet, Differentiated services, TCP/IP, Traffic management, Traffic engineering

1. INTRODUCTION

The Internet has grown exponentially since the commercialization and continues to do so. Though the Internet backbones have been designed with the ability to provide various levels of service guarantees, it is still impossible to obtain end to end QoS as the packets may travel over several network domains. The networking and data communication industry and researchers have been actively developing methods in which QoS can be guaranteed end to end in the Internet. Several approaches have thus come into existence as a result of intense research and standardization.

The Differentiated Services Working Group in IETF is in the process of standardizing the Per-Hop-Behavior (PHB) for IP networks. The main idea is to construct simple building blocks called PHBs that can be used to develop services that meet various QoS needs for IP networking. Various IP services can be constructed and implemented in the network via appropriate PHBs. Specific SLAs may be defined for each of these services based on the features and capabilities of the PHBs as well as the network architecture. An architecture for differentiated services is proposed in.[2] The definition of the DS field is given in.[9]

In order to implement differentiated services properly, the edge nodes must perform a number of traffic conditioning functions which include traffic policing and/or shaping, packet classification and marking. We characterize the dynamics of a token bucket policer and a three color marker scheme in this paper. All nodes in the network also implement buffer management schemes, queue scheduling policies and optional filtering and reclassification and marking. Moreover, some form of traffic engineering may

be required for efficient use of the network resources and also to satisfy the QoS requirements of the services.

Other author information: M. K. Girish: Email: mgirish@tri.sbc.com; Telephone: 925-598-1263; Fax: 925-598-1322. J. Q. Hu: Email: hu@enga.bu.edu; Telephone: 617-353-9696; Fax: 617-353-5548; The work of J. Q. Hu is supported in part by the National Science Foundation under Grant No. EEC-9527422.

Several issues and problems need to be addressed prior to implementing differentiated services for offering real QoS based IP services. Traffic engineering of DS networks is an important topic. When there are no end to end flows or connections, it is imperative that some mechanism is available to estimate or quantify the traffic at various nodes and output interfaces in the network. In this paper, we develop models to estimate the traffic at various points of the network which can be used to develop traffic engineering guidelines. The IETF has been standardizing a connection oriented protocol for IP networking called MultiProtocol Label Switching (MPLS) for efficient traffic engineering. When differentiated service building blocks are used in the edges of the network and the core network consists of connections or MPLS label switched paths, an optimization problem arises the optimal solution of which results in efficient traffic engineering. We formulate this problem in this paper.

We would like to point out to the reader that the problem that we are addressing in this paper is non-trivial in nature. Instead of developing complex exact models, we focus on simplified approaches and approximations that can be used for practical decision making and control. In conjunction with proper simulation, the results of this paper has great potential in understanding the complex dynamics in large IP networks that provide differentiated QoS and in developing simple traffic control and engineering methodologies.

The rest of this paper is organized as follows. We provide an overview of the AF PHB group in Section 2 which will be the main topic of interest in this paper. Traffic models, the corresponding constraints and guidelines on traffic engineering are developed in Section 3 and this paper is concluded in Section 4.

2. THE EF AND AF PHB GROUPS

The Expedited Forwarding (EF) PHB is proposed in[8] to address the needs For constructing end to end services through a network with low loss, delay,

jitter and assured bandwidth. For example, EF can be used to build a virtual leased line service through a DS domain. One of the requirements of an EF HB is that the departure rate of the aggregate packets from any DS node must not be less than the configurable rate. Moreover, this rate should be made available to EF traffic irrespective of the load of any other traffic. In addition, this rate should average at least the configured rate over any time interval greater than or equal to the time taken to transmit an MTU size packet at the configured rate. EF PHB can be implemented by a variety of queue scheduling mechanisms including priority queueing, weighted round robin, weighted fair queueing or

class based queueing.

The Assured Forwarding PHB group (see[6]) proposes 4 classes which have no delay bounds. Each class is independently forwarded and a node assigns its resources (bandwidth and buffer) for these four classes. Moreover, each class defines 3 drop precedence levels. The packets that are within a committed rate are forwarded with very high probability, whereas the second and third drop precedence packets are forwarded with decreasing probabilities. One of the requirements of AF is that the packets in a microflow should not be reordered, irrespective of the drop profiles associated with them in order to optimize the performance of that microflow. This means that the drop precedences do not play a role in queue scheduling and is used only in buffer management schemes. When a packet arrives at an output interface queue and if the buffer occupancy is such that this incoming packet cannot be queued, then if there are packets with lower drop preference in the buffer, sufficient number of them may be dropped (from the tail or via other buffer management mechanisms) to accommodate this packet. In the rest of this paper, we fill focus on a single AF class only.

3. TRAFFIC MODELS AND ANALYSIS

3.1. Dynamics of a Token Bucket Policer

The packets that come into the network from a customer site is often policed based on a token bucket scheme. In this section, we characterize the distributions of the packet departure from the token bucket and the token occupancy. A token bucket is characterized by two parameters, the token generation rate, r (in bytes/second) and the bucket size, B (in bytes). Let A_n be the interarrival time between packets $n-1$ and n, where we define $A_0 = 0$. Let S_n be the packet size (in bytes) of packet n. We assume that $S_n \leq B \ \forall \ n$. Tokens will be added to the bucket at a constant rate of r till the bucket is full. When a packet arrives, it is forwarded if there are sufficient tokens available in the bucket. It is clear that we only need to consider the epochs when packets arrive in the model. Let U_n be the token bucket size just before packet n arrives and let W_n be the token bucket size just after a packet is forwarded or discarded. We will consider the case of delaying a packet when sufficient tokens are not available later. Through an analysis of the sample path, we can show that:

$$\begin{aligned} U_n &= \mathrm{Min}[B, W_{n-1} + rA_n], & (1) \\ W_n &= U_n - S_n 1_{\{S_n \leq U_n\}}, & (2) \end{aligned}$$

where $U_0 = W_0 = B$ and $1_{\{Z\}}$ is the indicator function which has a value of 1, if Z is TRUE and a value of 0, otherwise. The above expressions can be exploited to derive the distributions of the token occupancy. The interarrival time distribution, in conjunction with the token occupancy distribution determine the distribution of the interdeparture time of packets from the token bucket (those that are forwarded into the network as well as those that are dropped).

3.2. Dynamics of A Three Color Marker

A three color marker scheme is proposed in[7] for marking incoming packets that belong to an AF class in a differentiated services domain by the edge routers as part of the traffic conditioning action. This scheme uses two token buckets: one for packets that are within a committed rate (which marks these packets as green) and the other for packets that are within an excess rate (which marks packets as yellow). Packets that violate both these buckets are marked as red. We denote the token bucket sizes as B^1 and B^2, respectively. The token bucket sizes just before packet n arrives are denoted by U_n^1 and U_n^2, and just after this packet is forwarded are denoted by W_n^1 and W_n^2, respectively. We can then show, by analyzing the sample path, that

$$\begin{aligned} U_n^1 &= \mathrm{Min}[B^1, W_{n-1}^1 + rA_n], & (3) \\ U_n^2 &= \mathrm{Min}[B^2, W_{n-1}^2 + rA_n], & (4) \\ W_n^1 &= U_n^1 - S_n 1_{\{S_n \leq U_n^1\}}, & (5) \\ W_n^2 &= U_n^2 - S_n 1_{\{S_n \leq U_n^2\}}(1 - 1_{\{S_n \leq U_n^1\}}), & (6) \end{aligned}$$

where $U_0^1 = W_0^1 = B^1$ and $U_0^2 = W_0^2 = B^2$. These expressions can be used to derive the distributions of the token occupancy.

3.3. Traffic Models

We model and analyze the traffic due to a single AF class in this section. We first introduce the following notation:

N : the number of nodes in the diff-serv domain
s_i : the number of customer sites/stub links in node $i, i = 1, \ldots, N$
b_{ij} : bandwidth of site j in node $i, j = 1, \ldots, s_i$
c : number of drop profiles or classes, c = 1, 2, 3
r_{ij}^c : subscription rate in site j in node i, $c = 1, 2, 3$
n_{ijkl} : number of microflows from site j of node i to site l of node k
x_{fijklm}^c : bit rate of class c in the forward direction of microflow m at node i, $m = 1, \ldots, n_{ijkl}$
x_{bijklm}^c : bit rate in the reverse direction (e.g. TCP ACK packets) of microflow m at node k
$\overline{x}_{fijklm}^c$: flow of class c in forward direction at site l of node k
$\overline{x}_{bijklm}^c$: flow of class c in reverse direction at site j of node i
R_{fijklm} : set of output interfaces or hops in the route of flow m in the forward direction
R_{bijklm} : set of output interfaces or hops in the route of flow m in the reverse direction
P_γ^c : drop probability in output interface γ
P_{fijklm}^c : total drop probability of class c packets in the forward direction
P_{bijklm}^c : total drop probability of class c packets in the reverse direction

The subscription rates must reflect the following property:

$$\tau_{ij}^1 \leq \tau_{ij}^2 \leq \tau_{ij}^3 \leq b_{ij} \; \forall \, i, j. \tag{7}$$

We note that

$$\begin{aligned}
x_{fijklm} &= \sum_{c=1}^{3} x_{fijklm}^c \\
x_{bijklm} &= \sum_{c=1}^{3} x_{bijklm}^c \\
\overline{x}_{fijklm} &= \sum_{c=1}^{3} \overline{x}_{fijklm}^c \\
\overline{x}_{bijklm} &= \sum_{c=1}^{3} \overline{x}_{bijklm}^c
\end{aligned}$$

Moreover,

$$\begin{aligned}
P_{fijklm}^c &= 1 - \prod_{\gamma \in R_{fijklm}} (1 - P_\gamma^c), \; \forall \, c, \\
P_{bijklm}^c &= 1 - \prod_{\gamma \in R_{bijklm}} (1 - P_\gamma^c), \; \forall \, c.
\end{aligned}$$

For differentiating the services, for any output interface γ, the network has to ensure that $P_\gamma^1 \leq P_\gamma^2 \leq P_\gamma^3$ and for assured delivery, $P_\gamma^1 \approx 0$. Obviously, the actual probabilities are dynamic in the sense that packet loss ratios depend on various factors such as the route through the network, the traffic load at various interfaces in the network, the buffer sizes and the buffer management policies. We can now express the following:

$$\begin{aligned}
\overline{x}_{fijklm}^c &= x_{fijklm}^c (1 - P_{fijklm}^c), \; \forall \, c, \\
\overline{x}_{bijklm}^c &= x_{bijklm}^c (1 - P_{bijklm}^c), \; \forall \, c.
\end{aligned}$$

The total traffic into node i from site j is:

$$X_{ij}^c = \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{ijkl}} x_{fijklm}^c + \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{klij}} x_{bklijm}^c . \tag{8}$$

It is clear that due to policing,

$$\sum_{g=1}^{c} X_{ij}^g \leq \tau_{ij}^c \; \forall \, c.$$

The total traffic of class c out of node i towards site j is:

$$\overline{X}_{ij}^c = \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{klij}} \overline{x}_{fklijm}^c + \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{ijkl}} \overline{x}_{bijklm}^c . \tag{9}$$

The excess packets (when traffic rate $> b_{ij} + Q_{ij}/t$) will be dropped

due to bandwidth and buffer constraints, where Q_{ij} is the buffer size in output interface j in site i in bits and t is the time duration when the rate exceeds the bandwidth.

3.4. Symmetric Traffic Distributions

In this section, we consider the case when the number of TCP sessions and traffic between any two sites are symmetric. This implies that

$$\begin{aligned} n_{ijkl} &= n_{klij}, \\ x^c_{fijklm} &= x^c_{fklijm} \ \forall\, c, \\ x^c_{bijklm} &= x^c_{bklijm} \ \forall\, c. \end{aligned}$$

We assume that the ingress node polices traffic arriving from each site on an aggregate basis based on the subscription rates of the three drop profiles. This scheme would reduce the processing requirement on the edge routers and also simplify the subscription specifications by the customer. This would lead to, for all c

$$\sum_{g=1}^{c} X^g_{ij} \leq \tau^c_{ij}. \tag{10}$$

The total traffic going out of site j in node i, from Equation (9), can then be shown to be:

$$\begin{aligned} \sum_{c=1}^{3} \overline{X}^c_{ij} &= \sum_{c=1}^{3}[\sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{klij}} \overline{x}^c_{fklijm} + \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{ijkl}} \overline{x}^c_{bijklm}] \\ &\leq \sum_{c=1}^{3}[\sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{klij}} x^c_{fklijm} + \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{ijkl}} x^c_{bijklm}] \\ &= \sum_{c=1}^{3}[\sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{ijkl}} x^c_{fijklm} + \sum_{k \neq i} \sum_{l=1}^{s_k} \sum_{m=1}^{n_{klij}} x^c_{bklijm}] \\ &= \sum_{c=1}^{3} X^c_{ij} \ \leq \ \tau^3_{ij} \ \leq \ b_{ij}. \end{aligned}$$

Therefore, it can be observed that for symmetric traffic between sites, it is sufficient to specify and control the aggregate traffic belonging to each class at the ingress stub links in order to prevent the stub links of the egress nodes from being bottlenecks. Thus, it ensures that the packets are not carried through the network before being dropped at the egress node which can make inefficient use of the network resources. This is consistent with the results in[5] where a similar model is analyzed for IP Virtual Private Networks. Furthermore, we can derive the conditions under which the above results hold for various other traffic patterns such as asymmetric, equally distributed, client-server type and multicast.

3.5. Traffic Engineering

We assume that a QoS aware routing protocol is implemented in the network so that the routing protocols can provide the required QoS to the IP packets. Nevertheless, our analysis holds true even if this is not the case. Let us assume that the routing protocol determines a route for packets from one edge node to another in the DS domain which is based on a microflow level for traffic engineering purposes. Efficient algorithms to achieve this are still under research. But, this assumption lets us analyze the system without loss of generality. Let R_{fijklm} and R_{bijklm} be the sets containing the list of output interfaces or hops along the forward and reverse paths of microflow m, respectively. If the routing protocol does not differentiate the paths for microflows, and only depends on the destination node in the DS domain, then R_{fijklm} and R_{bijklm} would be independent of j, l, m. Let g_{fh} be the h-th element (output interface) in the set R_{fijklm} and let g_{bh} be the h-th element in the set R_{bijklm}. Moreover, let $G_{f\gamma}$ be the position of the interface γ in the set R_{fijklm} and $G_{b\gamma}$ be the position of the interface γ in the set R_{bijklm}. We omit the subscripts i, j, k, l, m here for convenience and notational simplicity. The traffic of class c at an arbitrary interface γ can be expresed as:

$$\lambda^c_\gamma = \sum_i \sum_j \sum_k \sum_l \sum_m x^c_{fijklm} \prod_{h < G_{f\gamma}} (1 - P^c_{g_{fh}}) + \sum_i \sum_j \sum_k \sum_l \sum_m x^c_{bijklm} \prod_{h < G_{b\gamma}} (1 - P^c_{g_{bh}}), \tag{11}$$

where the summations are taken appropriately. The total traffic in interface γ is:

$$\lambda_\gamma = \sum_{c=1}^{3} \lambda_\gamma^c. \tag{12}$$

When Expedited Forwarding (EF) packets are also forwarded through the network, then the available bandwidth for AF would be reduced depending on the intensity and resources allocated for the EF traffic. Incorporating DS in connectionless IP backbone networks pose serious difficulties in traffic engineering. One of the ways in which the resources can be optimized is by using a QoS aware routing protocol. If a QoS aware protocol is not used, then one may have to resort to establishing some form of connections through the network. This can be either by the use of ATM or frame relay connections or via the emerging IETF standard MultiProtocol Label Switching (MPLS). One of the network architectures being considered by service providers and vendors is to use differentiated services PHBs at the edges of the network and using MPLS LSPs (Label Switched Paths) in the core of the network. This results in the ability to engineer the core network at the same time providing the ability to define SLAs based on the simple DS at the customer side and between domains. We now formulate the optimization problem associated with an MPLS network here which is also called the constraint based routing problem. Let us first define the following:

V = set of nodes in the network
E = set of links in the network. Defined as directed arcs
C = set of capacity and other constraints associated with V and E
U = set of LSRs where one or more LSPs terminate
F = set of LSPs
D = set of demands associated with F
λ_i = Effective bandwidth of LSP i
s_i = source LSR of LSP i
d_i = destination LSR of LSP i
h_i = maximum number of LSR hops through the network for LSP i
μ_l = bandwidth of link l
a_l = administrative cost of link l
K_l = maximum allocation multiplier or oversubscription factor for link l
u_l = originating LSR of link l
v_l = terminating LSR of link l

where $G = (V, E, C)$ is a graph describing the physical topology of the network and $H = (U, F, D)$ is the induced MPLS graph. LSP i has a demand or effective bandwidth of λ_i. This is derived from the subscriptions rates of the different AF drop precedence values and using an appropriate equivalent bandwidth calculation function. For EF traffic, λ_i is the configured rate. Given i, we also know the source LSR, s_i and destination LSR, d_i. The links in the network are depicted as directed arcs, with a total of L directed links in the network. For link l, we know the source node, u_l and the destination node v_l, the link capacity, μ_l and the admin cost of the link, a_l. The unknown variables that need to be determined based on optimizing a certain function and satisfying a set of constraints are the following:

$$x_{il} = \begin{cases} 1, & \text{if LSP } i \in F \text{ is routed on link } l \in E \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

We assume that all LSPs will be routed through the network. If this is not the case, then the problem will have no feasible solution. Resource based optimization would imply minimizing the sum over all links of the product of the administrative cost and the total flow in each link. We assume that the administrative cost can be applicable on a unit flow basis. This objective function is formulated as:

$$\text{Min } Z_R = \sum_{l \in E} a_l \sum_{i \in F} \lambda_i x_{il}. \tag{14}$$

For load balancing among the links in the network, the following objective function is proposed. This function minimizes the sum of the deviations from

the average load proportional to the link bandwidth:

$$\text{Min } Z_{LB} = \sum_{l \in E}[(\sum_{i \in F} \lambda_i x_{il}) - (\sum_{l \in E} \frac{\sum_{i \in F} \lambda_i x_{il} \mu_l}{\sum_{l \in E} \mu_l})]^2 \tag{15}$$

We propose a new objective function which combines the resource and load balancing based optimization. Given weights for the objectives for resources and load balancing, z_R, z_{LB}, respectively, that are chosen based on the network we have

$$\text{Min } Z = (z_R Z_R) + (z_{LB} Z_{LB}). \tag{16}$$

Note that these weights can be chosen in a way to reflect the preferences of

the service provider in optimizing the network.

The basic set of constraints for the optimization problem are:

$$\begin{aligned}
\sum_{i \in F} \lambda_i x_{il} &\leq \mu_l K_l, \;\; \forall\, l \in E && (17)\\
\sum_{l \in E} x_{il} &\leq h_i, \;\; \forall\, i \in F && (18)\\
(\sum_{\forall i \text{s.t.} s_i = n} \sum_{\forall l \text{s.t.} u_l = n} \lambda_i x_{il}) - (\sum_{\forall i \text{s.t.} d_i = n} \sum_{\forall l \text{s.t.} v_l = n} \lambda_i x_{il}) &= b_n, \;\; \forall\, n \in V && (19)\\
\sum_{\forall i \text{s.t.} s_i = n} \sum_{\forall l \text{s.t.} u_l = n} \lambda_i x_{il} &= \sum_{\forall i \text{s.t.} s_i = n} \lambda_i \;\; \forall\, n \in U && (20)\\
\sum_{\forall i \text{s.t.} d_i = n} \sum_{\forall l \text{s.t.} v_l = n} \lambda_i x_{il} &= \sum_{\forall i \text{s.t.} d_i = n} \lambda_i \;\; \forall\, n \in U && (21)\\
0 \leq x_{il} &\leq 1, \;\; \forall\, i \in F, l \in E \text{ and integer} && (22)
\end{aligned}$$

where

$$b_n = \sum_{\forall i \text{s.t.} s_i = n} \lambda_i - \sum_{\forall i \text{s.t.} d_i = n} \lambda_i \;\; \forall\, n \in V \tag{23}$$

Note that b_n specifies the total flow in or out of a node $n \in V$ in the network. If the aggregate flow is going out of a node, then b_n will be positive and negative, if the aggregate flow is into the node and finally, it will have a zero value for nodes $(n \in V - U)$ that do not have any LSPs originating or terminating in them.

Constraint (17) ensures that the link capacities are not exceeded. Note that the virtual link capacity is used here as the link may be oversubscribed or undersubscribed. Constraint (18) restricts the number of LSR hops in the path of a LSP. Constraint (19) ensures that the aggregate flow in a node via all links is equal to b_n. Constraints (20) and (21) assure that all LSPs originating and terminating, respectively, in a LSR are routed. Finally, Constraint (22) specifies that all decision variables are either 0 or 1. Additional constraints may be defined depending on specific requirements in networks (for example, refer to[1] for a comprehensive set of constraints and[4] for some mathematical formulations).

The above optimization problem can be shown to be NP-complete. The decision variables are integers. If the objective function is simply to minimize the administrative costs, this problem is still NP-complete as it is an integer linear programming problem. Incorporating the load balancing objective makes the objective function non-linear, resulting in a quadratic programming problem with integer variables. No NP-complete problem can be solved by any known polynomial time algorithm (see[10]). Some of the existing techniques are branch and bound and cutting plane methods. One of the main issues now is to develop efficient and computationally manageable techniques for solving this. Development of heuristics, approximation algorithms and exact solutions for simplified versions will be the focus of our work moving forward. We are presently working on using Lagrange multipliers to develop solution techniques for this problem.

4. CONCLUSIONS

The Internet has been growing exponentially and this trend has been projected to continue for many years into the future. A fundamental change expected to occur in the near future is the ability to construct and offer differentiated services in the Internet such that existing and new applications can be given appropriate qualities of service. One of the promising approaches is being standardized at the Differentiated Services Working Group in IETF. In order to incorporate differentiated services in large IP networks, it is imperative that we are able to model the traffic and also to analyze the impact on the resources such that these networks can be engineered optimally while meeting the service level agreements offered to customers. Being connectionless in nature, analyzing the traffic associated with IP packets are not only complex, but also highly variable. Hence stochastic approximations of queueing models is one of the practical approaches.

In this paper, we modeled the traffic associated with Assured Forwarding Per-Hop Behavior classes. We analyzed the dynamics of the token bucket policing scheme and a three color marker. The aggregate traffic at various points in the network were quantified and the traffic engineering problem was outlined. The results of this paper, in conjunction with simulation studies can be used to understand succinctly the complex dynamics in QoS based connectionless IP networks and they can be extended for efficient traffic engineering and control. We also formulated the traffic engineering problem in DS networks with an MPLS core.

Future work will focus on traffic engineering with QoS aware routing protocols, effect of various buffer management policies and congestion avoidance methods such as Random Early Detection (see[3]). It is also interesting to study the effect of EF class and multiple AF class PHBs being supported in the network.

REFERENCES

1. D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell and J. McManus, "Requirements for Traffic Engineering Over MPLS," *Internet Draft, draft-awduche-mpls-traffic-eng-00.txt,* April 1998.
2. D. Black et al., "An Architecture for Differentiated Services," *Internet RFC 2475,* December 1998.
3. S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Transactions on Networking,* **1:4,** 397-413, August 1993.
4. M. K. Girish, "Mathematical Formulation of the Constraint Based Routing Problem in MPLS Based IP Networks," *Technical Report,* SBC Technology Resources, Inc., November 1998.
5. M. K. Girish, J. Yu and T. Soon, "A QoS Proposal for IP Virtual Private Networks," *Technical Report,* SBC Technology Resources, Inc., March 1998.
6. J. Heinanen et al., "Assured Forwarding PHB Group," *Internet Draft, draft-ietf-diffserv-af-06.txt,* February 1999.
7. J. Heinanen and R. Guerin, "A Three Color Marker," *Internet Draft, draft-heinanen-diffserv-tcm-01.txt,* February 1999.
8. V. Jacobson, K. Nichols and K. Poduri, "An Expedited Forwarding PHB," *Internet Draft, draft-ietf-diffserv-phb-ef-02.txt,* February 1999.
9. K. Nichols et al, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," *Internet RFC 2474,* December 1998.
10. C. H. Papadimitriou and K. Steiglitz, "Combinatorial Optimization: Algorithms and Complexity," *Prentice Hall, Inc.,* 1982.

APPENDIX A. LIST OF ACRONYMS USED IN THIS PAPER

- AF: Assured Forwarding PHB
- DS: Differentiated Services
- EF: Explicit Forwarding PHB
- IETF: Internet Engineering Task Force
- IP: Internet Protocol

- LSP: Label Switched Path
- LSR: Label Switching Router
- MPLS: MultiProtocol Label Switching
- PHB: Per-Hop Behavior
- QoS: Quality of Service
- TCP: Transmission Control Protocol