

Prioritized Resource Allocation at Differentiated Services Ingress Routers

Xianghong Zeng and Cory Beard

Department of Electrical and Computer Engineering

University of Missouri-Columbia/Kansas City

5100 Rockhill Road, Kansas City, MO 64110, USA

Contact: Cory Beard, (816) 235-1550, Fax: (816) 235-1260, beardc@umkc.edu.

Abstract-Ingress routers into IETF Differentiated Services domains have become the focal point for controlling quality of service and traffic aggregation. It is necessary at ingress routers to control access to Differentiated Services traffic classes based user priorities and the numbers of connections that can be supported by various scheduling policies. Using deterministic analysis methods, we derive definitions for loss network state spaces for deadline-ordered and rate-based schedulers, discuss how the state spaces change with varying traffic parameters, and show in what cases deadline-ordered schedulers are better than rate-based schedulers. Then we propose upper limit policies, so that network managers can control loss network blocking probabilities fairly between users based on user priority and quality of service requirements.

Keywords-Network design, deterministic scheduling, connection admission control

I. INTRODUCTION

Recent developments in broadband integrated networks have addressed the problem of scalability of network control algorithms as networks increase in size. Significant developments in this arena have occurred in the Differentiated Services (Diffserv) Working Group of the Internet Engineering Task Force (IETF) [1]. This work has proposed use of aggregated traffic streams instead managing individual connections end-to-end. Simple forwarding behaviors have been defined for Diffserv core networks, like the Expedited Forwarding (EF) Per Hop Behavior [2], to provide a high throughput, low-delay service. The key to providing this service, however, is not the behavior itself, but rather mechanisms for limiting the amount of this type of traffic that enters the network.

Ingress routers into Diffserv domains must conform their traffic to a traffic conditioning specification (TCS) which limits traffic rates and burstiness. Once this traffic enters a Diffserv domain, the EF per hop behavior provides low delay through each Diffserv router. The particular point of interest for Diffserv is the ingress router; which performs functions of classification, metering, marking, dropping, and shaping. Even though delay in Diffserv core domains has been reduced, the delay at these ingress routers now becomes a concern. Traffic rates out of these routers will be limited and demand for this resource will periodically exceed demand. Controls must be implemented to fairly regulate access

to resources. Connection-oriented mechanisms and resource reservation protocols may be used for this function. In addition, individual packets may experience significant delays at these ingress points because of traffic shaping functions. Sophisticated scheduling mechanisms will need to be implemented to differentiate between classes to limit the delays experienced at ingress nodes.

The objective of this work is to enable resource managers at ingress routers to control delay (at the packet level) and blocking (due to insufficient resources to meet delay requirements). We first present methods for computing blocking probabilities when attempting to bound delay through deadline-ordered and rate-based schedulers. We then investigate differences in schedulers based on input traffic parameters. We then finish by going beyond simply being able to compute blocking probabilities to proposing the use of upper limit policies to fairly distribute blocking across user classes with different traffic demands and priorities of activities.

II. STATE SPACES

The study of the numbers of connections and users that can be supported in a network is the study of *loss networks* [3]. In a loss network, requests for resources are either accepted or rejected and do not return. The key metric of interest for such networks, therefore, is the probability that such requests will be denied (i.e., blocked). Requests are blocked when insufficient resources are available. If types of requests can be grouped into classes, then a state space can be derived that defines the numbers of connections per each class that can be supported by the network. A state space, Ω , can then be defined from the vector, \mathbf{n} , of the number of concurrent users from each class as

$$\begin{aligned} \Omega &= \{\mathbf{n} \in Z^+ : \text{sufficient resources are available to support } \mathbf{n} \text{ connections}\} \\ Z^+ &= \text{set of non - negative integers.} \end{aligned} \tag{1}$$

We assume we have K classes and that connections arrive according to a Markov process, service times are distributed generally, λ_k is the average arrival rate for class k , and μ_k is the average service rate for class k . With Ω_k as the set of states from which a new connection from class k would not be blocked, we can then use the classic Erlang B loss equation to find the blocking probability for class k as follows [3].

$$B_k = 1 - \frac{\sum_{\mathbf{n} \in \Omega_k} \prod_{k \in K} \frac{(\lambda_k / \mu_k)^{n_k}}{n_k!}}{\sum_{\mathbf{n} \in \Omega} \prod_{k \in K} \frac{(\lambda_k / \mu_k)^{n_k}}{n_k!}} \tag{2}$$

The key element to this analysis is the definition of the state space, Ω . The traditional approach has been to assume

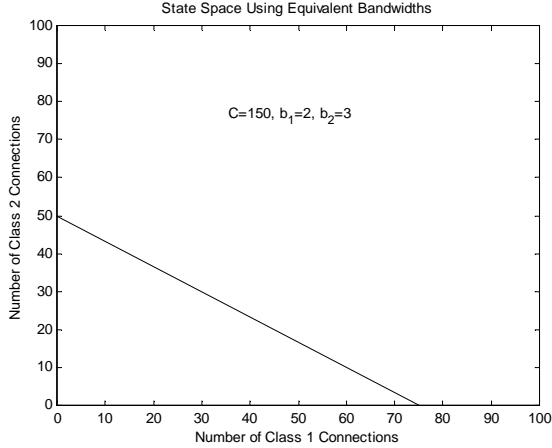


FIGURE 1 - EQUIVALENT BANDWIDTH STATE SPACE

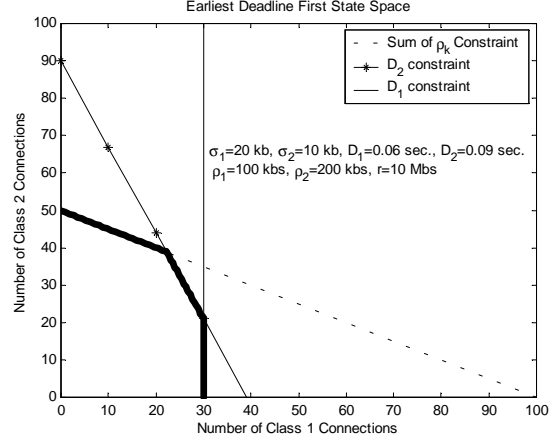


FIGURE 2 - EDF STATE SPACE

that each connection in class k uses an equivalent bandwidth, b_k , that provides the connection with a statistical expectation of performance related to rate, loss, and delay. The state space, Ω , is then defined, for a given capacity, C , of a network link, as

$$\Omega = \{ \mathbf{n} \in Z^+ : \sum_{k=1}^K b_k n_k \leq C \}. \quad (3)$$

Based on this definition of Ω , a large body of work has been conducted to find simplifications and approximations for (2) [3, 4]. Figure 1 illustrates a sample state space boundary for two classes.

In our analysis, we use loss network analysis methods to find blocking probabilities for loss network state spaces that provide deterministic delay bounds through ingress routers. We assume that sources are shaped by traffic shapers to a (σ, ρ) specification [5], where σ is the burstiness of the source and ρ is the average flow rate. We then find the state space of possible vectors, \mathbf{n} . This state space has also been called the *schedulable region*, but has not previously been defined in terms of \mathbf{n} as we do here. Figure 2 illustrates the state space boundary for an Earliest Deadline First scheduler as derived in the next section. In a later section, we propose mechanisms to balance the blocking probabilities fairly between classes.

Two basic categories of schedulers that provide delay bounds are deadline-ordered and rate-based schedulers [6]. Many variants of these schedulers have been developed, but follow the same basic approach. Other simpler types of schedulers, such as fixed priority schedulers will not be considered because they do not inherently provide a fixed upper bound on delay. The Earliest Deadline First (EDF) and Generalized Processor Sharing (GPS) schedulers are investigated here as the fundamental bases of deadline-ordered and rate-based schedulers.

A. Earliest Deadline First (EDF)

Deadline-ordered schedulers assume that packets carry with them a specification for the maximum delay, D_i , that can be incurred at a node. Packets are then scheduled according to those delay specifications. To provide bounds on delay, an EDF scheduler must admit only as many connections as can all have their delay constraints satisfied.

To find the state space for EDF policies, we expand on the derivations in [7]. To simplify the discussion, we do not consider packetization issues here, but instead assume that packets arrive as continuous flows. Let the non-negative vector, $\mathbf{D}=(D_1, D_2, \dots, D_N)$ be an ordered list of upper bounds on delay for connections 1 through N . Given a link rate, r , an EDF scheduler can then support this set of N connections if the set of following constraints is met.

$$\sum_{i=1}^N (\sigma_i + \rho_i (D_n - D_i)) U(D_n - D_i) \leq r D_n, \quad 1 \leq n \leq N \quad (4)$$

These equations assume that all sources are greedy sources starting at time $t=0$, (i.e., they transmit as much as possible as allowed within the (σ_i, ρ_i) specification). Then, at each specific time, D_n , the total amount of traffic that has entered the node that should depart by time, D_n , is less than or equal to the amount that could be transmitted at rate r . We also assume that the sum of the input rates is less than r (which is necessary to ensure bounded delays),

For our purposes, we do not assume that all connections are different, but can rather be grouped into connections of common (σ_i, ρ_i) specifications and D_i requirements. The delay specification vector then becomes $\mathbf{D}=(n_1 D_1, n_2 D_2, \dots, n_K D_K)$ where $n_1 + n_2 + \dots + n_K = N$ and $D_i \leq D_j$ if $i < j$. The resulting set of constraints becomes

$$\begin{aligned} \sum_{k=1}^K (\sigma_k + \rho_k (D_n - D_k)) U(D_n - D_k) &\leq r D_n, \quad 1 \leq n \leq K \\ \sum_{k=1}^K n_k \rho_k &\leq r, \end{aligned} \quad (5)$$

or in expanded form for two classes,

$$\begin{aligned} n_1 \sigma_1 &\leq r D_1 \\ n_1 (\sigma_1 + \rho_1 (D_2 - D_1)) + n_2 \sigma_2 &\leq r D_2 \\ n_1 \rho_1 + n_2 \rho_2 &\leq r. \end{aligned} \quad (6)$$

Figure 2 shows an example of this state space. Notice that each of the three constraints forms part of the heavy boundary line. The next section describes how the state space boundary changes depending on traffic parameters.

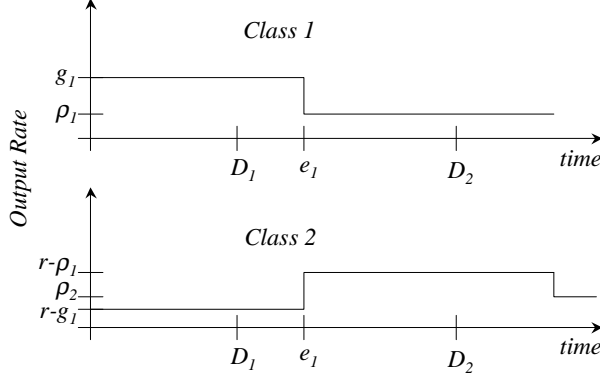


FIGURE 3 - GPS OUTPUT FLOWS FOR $e_1 < D_2$

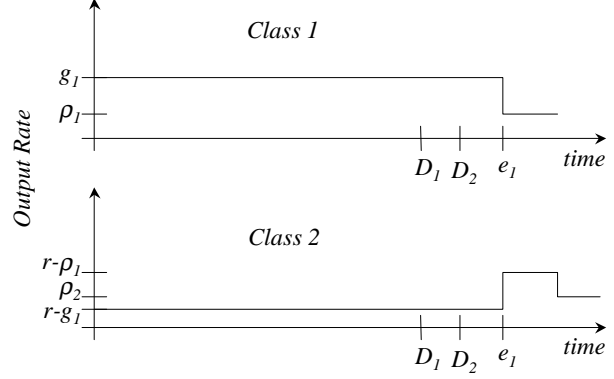


FIGURE 4 - GPS OUTPUT FLOWS FOR $e_1 > D_2$

B. Generalized Processor Sharing (GPS)

Generalized Processor Sharing (GPS) is the fundamental scheduling philosophy behind rate-based schedulers. The basic concept is to view each connection as a flow that is segregated from other flows and allocated a certain portion of the capacity. Each flow is guaranteed a minimum rate, and is able to use more of the capacity if other sources are idle. Alternative forms of GPS scheduling have been developed under names such as Packet-by-Packet GPS (PGPS), Weighted Fair Queueing (WFQ), Worst-case Fair Weighted Fair Queueing (WF²Q), etc.

By shaping a flow before it arrives at the scheduler, a minimum rate can be determined to provide bounded delay to the connection. Figure 3 illustrates the process for determining these minimum rates, and the corresponding state spaces, for two classes using GPS. It is assumed that class 1 is allocated a minimum rate, g_1 , and class 2 uses as much of the remaining capacity as it needs. Following the feasible ordering approach of [8, Section VI.C], e_1 is defined as the time when class 1 empties the backlog in its queue. To obtain bounded delay, a rate g_1 must be allocated to class 1 so that a burst σ_1 that arrives at $t=0$ for class 1 can be transmitted out of the node by D_1 . The rate g_1 must be less than or equal to r and also greater than the source's average rate, ρ_1 (so queues will empty at a faster average rate than they can be filled).

$$\begin{aligned} \rho_1 &\leq g_1 \leq r \\ D_1 &\leq \frac{\sigma_1}{r} \end{aligned} \tag{7}$$

Class 2 will transmit at rate $r-g_1$ for as long as class 1 needs to transmit (i.e., as long as there is a backlog for class 1). The backlog in class 1 clears at time e_1 , when the burst that arrived at $t=0$ is cleared, as well as any new traffic that arrived at rate ρ_1 . After time e_1 , class 1 traffic arrives at rate ρ_1 and only needs output rate ρ_1 . This, therefore, leaves a

rate $r - \rho_1$ now available to class 2. To meet the D_2 deadline for class 2, there must be sufficient rate for class 2 to send out its burst σ_2 that arrived at $t=0$. The following two constraints must be met.

$$\begin{aligned} r - \rho_1 &\geq \rho_2 \quad \rightarrow \quad \rho_1 + \rho_2 \leq r \\ \sigma_1 + \rho_1 D_2 + \sigma_2 &\leq r D_2. \end{aligned} \tag{8}$$

The sum of arrival rates must be less than the capacity, and the total amount of data that GPS will attempt to transmit by D_2 must be less than the total amount possible to transmit. By combining (7) and (8) together, as well as incorporating the same concepts of traffic classes used for EDF, we find the following constraint space for GPS.

$$\begin{aligned} n_1 \sigma_1 &\leq r D_1 \\ n_1 (\sigma_1 + \rho_1 D_2) + n_2 \sigma_2 &\leq r D_2 \\ n_1 \rho_1 + n_2 \rho_2 &\leq r. \end{aligned} \tag{9}$$

Note that this form follows very closely the form of the constraints for EDF in (6), except that the second constraint has a $\rho_1 D_2$ term for class 1 instead of a $\rho_1(D_2 - D_1)$. This difference is explored in Section IV. Because of the similarity of the constraint equations, plots for the state space for GPS are similar to that given for EDF in Figure 2.

For the sake of completion, it is important to notice that it is possible for $e_1 > D_2$, as shown in Figure 4. This occurs when D_1 is close to D_2 . In such cases, it is necessary for the $r - \rho_1$ rate available to class 2 to be sufficient for meeting the D_2 delay bound, even if class 1 has not yet emptied its backlog. It can be shown that in this case,

$$\begin{aligned} e_1 &\geq D_2 \\ D_1 &\geq \frac{1}{\frac{1}{D_2} + \frac{\sigma_1}{\rho_1}}. \end{aligned} \tag{10}$$

and the constraint space becomes

$$\begin{aligned} n_1 \sigma_1 &\leq r D_1 \\ n_1 \left(\frac{\sigma_1 D_2}{D_1} \right) + n_2 \sigma_2 &\leq r D_2 \\ n_1 \rho_1 + n_2 \rho_2 &\leq r. \end{aligned} \tag{11}$$

For the sake of brevity, we will defer derivations of the GPS state space for more than 2 classes to the sequel to this paper. We will also consider the effect of packetized versions of EDF and GPS, like Non-preemptive EDF and Packet-by-Packet GPS in the sequel.

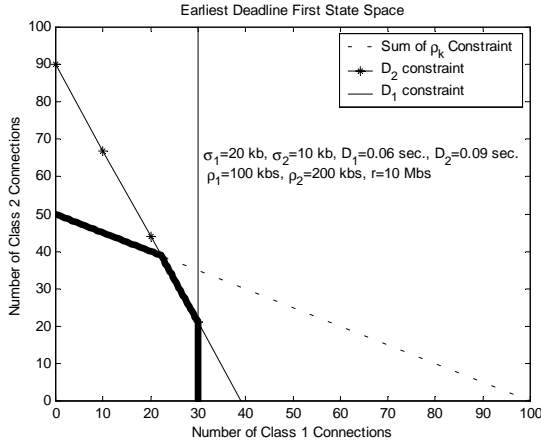


FIGURE 5 - BASE EDF STATE SPACE

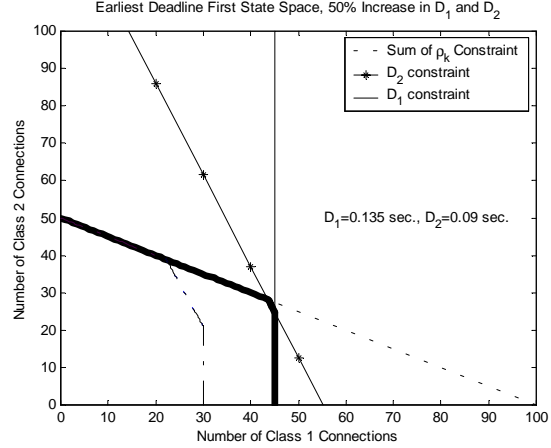


FIGURE 6 - 50% INCREASE IN ALLOWABLE DELAY

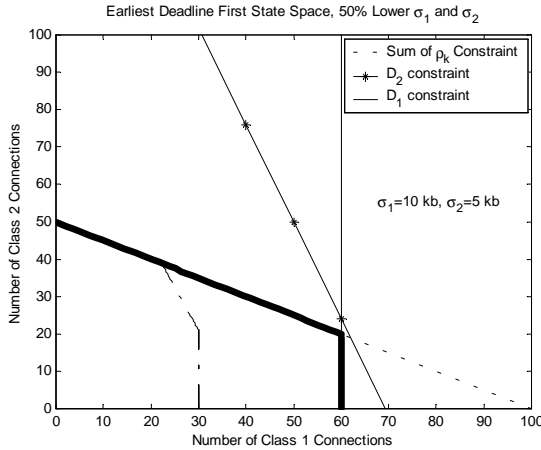


FIGURE 7 - 50% DECREASE IN BURST SIZE

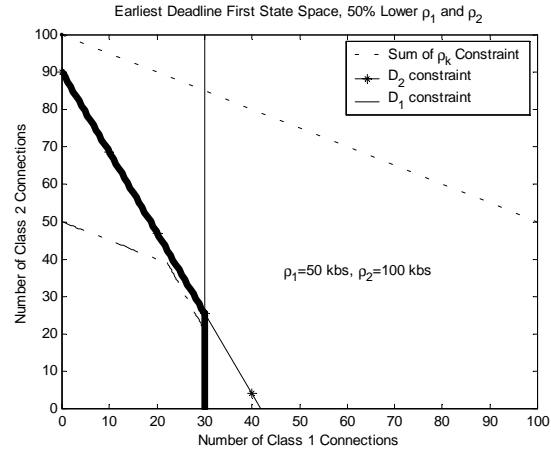


FIGURE 8 - 50% DECREASE IN AVERAGE RATE

III. IMPACTS OF TRAFFIC PARAMETERS ON STATE SPACE BOUNDARIES

This section provides a discussion of how state spaces change as traffic parameters and delay requirements are modified. More specifically, the goal is to enlarge the state space to make better use of an existing contract into a Diffserv domain. Figure 5 shows the same plot as Figure 2 of the boundary of the schedulable region (shown with a heavy line), and is used as the basis of comparison. Figures 6 through 8 show the new state spaces, and Table 1 shows the new blocking probabilities as computed using (2). For blocking probabilities we assume that $\lambda_1=\lambda_2=24.5$ connections per second for all cases and $\mu_1=\mu_2=1$ connection per second. The final column in the table shows how much higher arrival rates can be supported to provide the same sum of blocking probabilities.

Figure 6 shows the increased size of the state space from making both classes tolerate 50% higher delay and Figure 7 shows a similar change in size for 50% changes in burst size. Both cases provide more connections for class 1. Table 1

shows that the changes in blocking probabilities are very significant for these two cases and that by making these modifications, 26% more load can be supported. A 50% decrease in the average rate, ρ , as shown in Figure 8 does not show the same marked decrease in blocking. Overall, it appears that managing the burstiness and delay requirements of connections will significantly affect the use of capacity.

TABLE 1 - BLOCKING PROBABILITIES FOR CHANGES IN STATE SPACES

Case	B_1	B_2	B_1+B_2	ARRIVAL RATES FOR EQUIVALENT $B_1+B_2=0.1$
Base	0.076	0.030	0.106	24.30 connections per sec.
50% Larger Delay Requirement	0.0025	0.0053	0.0077	30.57 connections per sec. (25.8% increase)
50% Decrease in Burst Size	0.0024	0.0053	0.0077	30.62 connections per sec. (26.0% increase)
50% Decrease in Average Rate	0.059	0.016	0.075	25.50 connections per sec. (4.9% increase)

IV. COMPARISON OF EDF AND GPS POLICIES

In Section II, we mentioned that the derivation of state spaces for GPS and EDF yield slightly different results. The potential differences between the policies are illustrated in Figure 9. The EDF region is shown to be bounded by lines with bold "+" symbols; the GPS region boundary has no symbol. The main difference in GPS and EDF policies is the fact that GPS policies provide unnecessarily short delays in the time period after time e_1 (from Figure 3). In contrast, an EDF policy does not waste capacity this way. Figure 10 shows for a certain set of parameters the amount of extra load that could be supported for an EDF as compared to a GPS policy and still provide the same sum of blocking probabilities. By empirical investigation, we were able to formulate cases where up to 25% more load could be supported. These benefits, however, could only be obtained for a fairly narrow range in D_1 values.

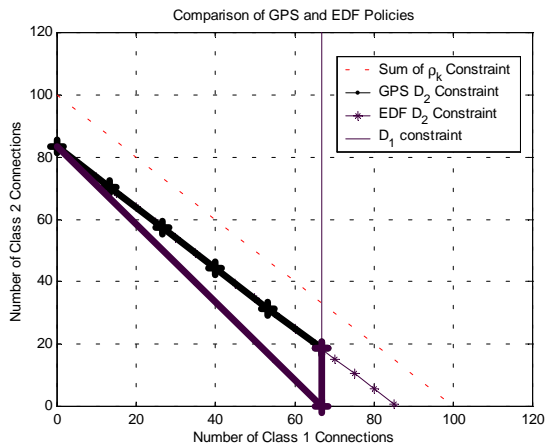


FIGURE 9 - GPS VERSUS EDF STATE SPACE

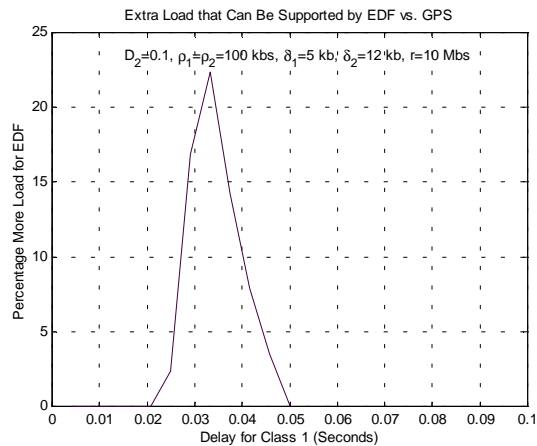


FIGURE 10 - EXTRA LOAD SUPPORTED FOR EDF POLICIES

V. PRIORITIZED RESOURCE MANAGEMENT

In this paper, we have established the state spaces for EDF and GPS policies and the ability to compute blocking probabilities for those policies. We have also discussed how parameters could be modified to expand the state spaces. Another consideration in effective management of resources at Diffserv ingress routers, however, is the ability to *balance* blocking between classes. In some cases a class will even need markedly lower blocking probabilities because of user priorities (e.g., priority users for emergency activities). In general, this can be viewed as an optimization problem, where one objective might be to optimize a weighted sum of blocking. In [9], a mechanism was proposed for optimizing weighted blocking by implementing *upper limit* policies where a limit, L_k , is set on the maximum numbers of connections allowed for certain classes. An optimization method then optimized weighted blocking for the state space in (3) based on an asymptotic blocking approximation. By using the same methodology for the state space definitions for EDF and GPS classes, the following simple linear programming problems for EDF and GPS weighted blocking optimization can be formulated. There are K classes of traffic, w_k are the weights for each class, expected service rates are assumed to be unity, and maximum and minimum blocking probabilities can be specified.

<u>EDF</u>	<u>GPS</u>
$\min \left(- \sum_{k=1}^K \frac{w_k L_k}{\lambda_k} \right)$	$\min \left(- \sum_{k=1}^K \frac{w_k L_k}{\lambda_k} \right)$
subject to	subject to
$L_k + s_{k,1} = L_{k,\max} = \lambda_k (1 - P_{B_{k,\min}})$	$L_k + s_{k,1} = L_{k,\max} = \lambda_k (1 - P_{B_{k,\min}})$
$L_k - s_{k,2} = L_{k,\min} = \lambda_k (1 - P_{B_{k,\max}})$	$L_k - s_{k,2} = L_{k,\min} = \lambda_k (1 - P_{B_{k,\max}})$
$L_1 \sigma_1 + s_3 = rD_1$	$L_1 \sigma_1 + s_3 = rD_1$
$L_1 (\sigma_1 + \rho_1 (D_2 - D_1)) + L_2 \sigma_2 + s_4 = rD_2$	$L_1 (\sigma_1 + \rho_1 D_2) + L_2 \sigma_2 + s_4 = rD_2$
$L_1 \rho_1 + L_2 \rho_2 + s_5 \leq r$	$L_1 \rho_1 + L_2 \rho_2 + s_5 \leq r$
$L_k, s_{k,1}, s_{k,2}, s_i \geq 0$	$L_k, s_{k,1}, s_{k,2}, s_i \geq 0$

Given the system in Figure 5, the optimum upper limit policy for $w_1=0.1$, $w_2=0.9$, and $P_{k,\max}=0.9$ is $L_1=20$ and $L_2=40$ for $\lambda_1=\lambda_2=40$. Without the upper limit policy, the weighted sum of blocking would be 0.236; with the upper limit policy, the weighted sum equals 0.157.

VI. CONCLUSION

In this paper we have addressed the concern that Differentiated Services ingress routers will become a choke point

for packet delay without the implementation of advanced packet scheduling policies. We have derived the state spaces for EDF and GPS schedulers so that ingress routers and resource managers can have simple equations to assess the numbers of users that can be supported and the blocking probabilities that can be expected. We then propose the use of upper limit policies so that contention for resources between users can be fairly distributed based on the needs of the traffic and the importance of the activities.

REFERENCES

- [1] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services", Internet Engineering Task Force, RFC 2475, December 1998.
- [2] V.Jacobson, K.Nichols, K.Poduri, "An Expedited Forwarding PHB", Internet Engineering Task Force, RFC 2598, June 1999.
- [3] K. W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer-Verlag, London, 1995.
- [4] F. P. Kelly, "Loss Networks," *Annals of Applied Probability*, Vol. 1, No. 3, pp. 319-378, 1991.
- [5] R. L. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation," *IEEE Transactions in Information Theory*, 37:132-141, January 1991.
- [6] R. Guerin and V. Peris, "Quality of service in packet networks: basic mechanisms and directions," *Computer Networks*, 31:169-189, 11 February 1999.
- [7] L. Georgiadis, R. Guerin, A. Parekh, "Optimal multiplexing on a single link: Delay and buffer requirements", *IEEE Transactions on Information Theory*, 43 (5), September 1997, 1518–1535.
- [8] A.K. Parekh, R.G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case", *IEEE/ACM Transactions on Networking*, 1(3), June 1993, 344–357.
- [9] Cory C. Beard and Victor S. Frost, "Connection Admission Control for Differentiating Priority Traffic on Public Networks," *1999 IEEE Military Communications International Symposium*, Atlantic City, New Jersey.