Effective Bandwidth Vectors for Multiclass Traffic Multiplexed in a Partitioned Buffer*

V.G. Kulkarni†

Department of Operations Research
University of North Carolina
Chapel Hill, N.C. 27599-3180

L. Gün
Motorola Codex
20 Cabot Boulevard
Mansfield Mass. 02048-1193

P. F. Chimento‡

Tele-Informatics and Open Systems
University Of Twente
Enschede, The Netherlands

August 31, 2000

Abstract

We consider a traffic model where a single source generates traffic having \( J \) (\( J \geq 2 \)) Quality of Service (QoS) classes. QoS in this case is described by a cell loss probability objective \( \epsilon_j \) for QoS class \( j \). We assume that \( \epsilon_1 \geq \epsilon_2 \geq \ldots \geq \epsilon_J \), in other words, class \( J \) has the most stringent QoS requirements and class 1 the least. The traffic from \( K \) such independent heterogeneous Markov-modulated fluid sources is multiplexed into a single buffer of size \( B \). There are \( J - 1 \) thresholds \( \{B_j, 1 \leq j \leq J - 1\} \) such that \( 0 < B_1 < B_2 < \ldots < B_{J-1} < B \). Let \( B_0 = 0 \) and \( B_J = B \). If the buffer content is in \( (B_{j-1}, B_j) \), \( 1 \leq j \leq J \) only traffic of class index \( j \) or above is accepted and all other traffic is rejected. For this system of \( K \) sources we define an effective bandwidth vector of size \( J \) such that QoS requirements for all classes are satisfied if each component of the vector is less than the channel capacity. We propose several bandwidth vectors that can be computed for each source separately. Numerical studies are reported on the efficacy of these bandwidth vectors.

*This work was performed while the last two authors were at IBM High Bandwidth Networking, Network Analysis Center, Research Triangle Park, NC. 27709, and the first author was visiting there. The work was presented at the INFORMS Telecom Conference in Boca Raton, Florida, March, 1995.
†work of this author is partially supported by NSF Grant No. NCR-9406823
‡Currently on personal leave from IBM
1 Introduction

The emerging high speed networks achieve efficiency and higher resource utilization by using statistical multiplexing. Under statistical multiplexing the traffic from several sources gets superimposed onto a single buffer and is transmitted to the network in a first-come-first-served fashion. In the high-speed networks using Asynchronous Transfer Mode (ATM), each source is assured a given quality of service (QoS), and the admission control scheme is designed to assure that a source is admitted into the network only if there is sufficient capacity in the network to guarantee the QoS.

Stochastic fluid models and the theory of large deviations have been used in the literature to define what is called "effective bandwidth" or "equivalent capacity" for each source. The effective bandwidth is a number that depends on the source characteristic, the QoS requirement of the source and the buffer size. The most useful property of the effective bandwidth is the following: If the sum of the effective bandwidths of all the sources multiplexed onto a buffer is less than the speed of the channel removing data from the buffer then (in certain asymptotic sense) the QoS requirements of all the sources are satisfied. This is what makes the concept of effective bandwidths very useful in admission control. Several authors have studied this concept from different angles: see [6, 9, 7, 11, 8, 3].

The above method of admission control indeed works very satisfactorily as long as the QoS requirements of all the sources are the same or at least close. In practice all the sources get the QoS of the most stringent source, since they all get the same QoS. This manifests itself in less than optimal resource allocation. This is a major drawback of this method.

When the source QoS requirements are varied, or when every source generates traffic belonging to multiple classes with varying QoS requirements (such as MPEG-2 multi-layered video, or sources policed by leaky bucket regulators), the above call admission procedure needs to be changed. In particular, when a source, such as a multi-layered video codec, transmits, the information from different layers that emanate from that single source may have very different QoS requirements. So, for example, some signals may have only enhancement effects on the final image, or enhancements of sound quality, and so can be transmitted at a much lower QoS than signals having to do with the more fundamental quality of the way the information is presented. (See [13].)

One method is to classify the traffic into multiple classes according to their QoS requirements and to employ a separate buffer for each class of traffic. Then the effective bandwidth methodology can be applied to each buffer separately. This, however, has several drawbacks: First, we need a
scheduler to schedule the transmission from the various buffers to ensure first in first out discipline for each connection (source). Second, the buffer utilization is inferior. Third, if the cells of different priorities emanate from a single source (that is, they are part of a single stream), then having separate buffers for differing QoS requirements forces resequencing at the destination, which is highly undesirable. Fourth, having separate logical buffers for each QoS requirement complicates the implementation.

To overcome the above disadvantages one can consider a shared buffer scheme. Under this scheme the traffic is classified as before. Suppose there are \( J \) classes, indexed 1, 2, ..., \( J \). We assume that class \( J \) has the most stringent QoS and class 1 has the least stringent one. The buffer is of size \( B \). There are \( J - 1 \) thresholds \( \{B_j, 1 \leq j \leq J - 1\} \) such that \( 0 < B_1 < B_2 < \ldots < B_{J-1} < B \). For convenience, define \( B_0 = 0 \) and \( B_J = B \). If the buffer content is in the interval \( (B_{j-1}, B_j), (1 \leq j \leq J) \) only traffic of class index \( j \) or above is accepted and all other traffic is rejected. The admission policy is more complicated when the buffer content is exactly at a threshold and will be stated precisely in the next section.

The above scheme of differentiating traffic from different traffic classes is called a buffer sharing scheme. It is intuitively clear that the above mechanism would provide highest QoS to class \( J \) and lowest QoS to class 1 traffic. Furthermore, the scheme maintains the first-in-first-out ordering among all accepted traffic and eliminates the need for schedulers. It is also easy to analyse (see [3, 13]) and is shown to be optimal within a large class of service policies (see [14, 12, 2]). For these reasons this scheme is very attractive.

In this paper we analyse this scheme and show how the concept of effective bandwidths can be extended to this configuration. The aim is construct a vector of size \( J \) (called the effective bandwidth vector) for each source such that the QoS requirements are satisfied for all classes if the each component of the sum of the effective bandwidth vectors is less than the channel speed.

The purpose of constructing such a vector is to try to reduce the bandwidth requirements of the source stream as a whole. So, for example, if no priority differentiation is made, then all the cells from that particular source must be treated as though they required the most stringent QoS. Intuitively, one would expect the bandwidth requirements to be less when there is priority differentiation. For a given number of sources and for a given trunk capacity, bandwidth savings can
be defined as the difference between the total effective bandwidth required by the sources with no differentiation and that required by the sources with two or more levels of priority.

The paper is organized as follows: The model is described in detail, and the bulk of the notation is introduced in Section 2. Section 3 contains the relevant results from the descriptive analysis of this model. This section essentially restates the results of [3] using our notation. In Section 4 we present the asymptotic analysis of the loss probability (QoS). The asymptotic region is precisely defined in this section. It also states a sufficient condition for the priority differentiation to produce savings. Using the results of this section, we construct an effective bandwidth vector for the system of \( K \) sources in Section 5. Unfortunately, this bandwidth vector is not additive, i.e., it cannot be written as a sum of \( K \) effective bandwidth vectors, one for each source. Hence we study several possible candidates for effective bandwidth vectors for a single source. The most logical candidate, unfortunately, provides only an approximation, and not a bound. Hence we propose several candidates that can be proved to provide bounds, although loose. In Section 6 we discuss numerical algorithms to compute the effective bandwidth vectors. We illustrate the concepts in Section 7 by numerical examples using a voice-multiplexing example.

2 The Model

In this section we describe the precise model of multiplexing \( K \) Markov modulated fluid sources, each producing fluid belonging to \( J \) different classes onto a single buffer of size \( B \) that is serviced by a channel of capacity \( c \). Let \( Z_k(t) \) be the state of the \( k^{th} \) source (i.e., the state of the external environment that controls the traffic stream generated by the \( k^{th} \) source) at time \( t \). \( \{Z_k(t), t \geq 0\} \) is assumed to be an irreducible Continuous Time Markov Chain (CTMC) with state space \( S_k = \{1, 2, ..., N_k\} \) and generator matrix \( G_k, (1 \leq k \leq K) \). When the \( k^{th} \) source is in state \( i \), it generates class \( j \) fluid at rate \( \lambda_{k,i}^j \). As explained in the introduction, the admission policy is based upon a space reservation scheme, using thresholds \( \{B_j, 1 \leq j \leq J-1\} \).

Let \( X(t) \) be the amount of fluid (of all classes) in the buffer at time \( t \). When \( B_{j-1} < X(t) < B_j \), only fluid belonging to classes \( \{j, j+1, ..., J\} \) is admitted into the buffer. Thus when \( 0 \leq X(t) < B_1 \), fluid of all classes is admitted, while when \( B_{J-1} < X(t) < B_J \), only fluid of class \( J \) is admitted. At threshold \( B_j \) fluid of class \( j \) is admitted at such rate that the buffer content does not rise above \( B_j \). To make this precise, we give an expression for \( I_{j}(t) \), the rate at which the fluid of class \( j \) is
admitted into the buffer at time \( t \). First we need the following notation

\[
\lambda_{k,i}^j = \sum_{r=j}^{J} \lambda_{k,i}^r.
\]  

(1)

With this we can write, for \( 1 \leq j \leq J \),

\[
I_j(t) = \begin{cases} 
\sum_{k=1}^{K} \lambda_{k,Z_k(t)}^j, & \text{if } X(t) < B_j, \\
\min\{\sum_{k=1}^{K} \lambda_{k,Z_k(t)}^j, (c - \sum_{k=1}^{K} \lambda_{k,Z_k(t)}^{j+1})\} & \text{if } X(t) = B_j, \\
0 & \text{if } X(t) > B_j.
\end{cases}
\]  

(2)

We use \((x)^+ = \max(0, x)\) and \((x)^- = \min(0, x)\). Now let \( \Phi(B_j) \) be the long run probability that the buffer content is above the threshold \( B_j \). We take this probability to be the surrogate for the loss fraction for fluid of class \( j \). The QoS guarantee is that this loss fraction is kept below a prespecified quantity \( \varepsilon_j \). Obviously, \( \varepsilon_1 > \varepsilon_2 > ... > \varepsilon_J > 0 \). Thus, using the surrogate, the QoS is satisfied for all the classes of traffic if

\[
\Phi(B_j) < \varepsilon_j, \quad \text{for all } 1 \leq j \leq J.
\]  

(3)

3 The Analysis

Let \( Z(t) = (Z_1(t), Z_2(t), ..., Z_K(t)) \) be the state vector of the \( K \) sources. It is clear that \( \{(X(t), Z(t)), t \geq 0\} \) is a Markov process with piecewise deterministic paths. The dynamics of the process is given by

\[
\frac{dX(t)}{dt} = \begin{cases} 
(\sum_{j=1}^{J} I_j(t) - c)^+ & \text{if } X(t) = 0, \\
\sum_{j=1}^{J} I_j(t) - c & \text{if } 0 < X(t) < B, \\
(\sum_{j=1}^{J} I_j(t) - c)^- & \text{if } X(t) = B.
\end{cases}
\]  

(4)

Now \( \{Z(t), t \geq 0\} \) is a CTMC with state-space \( \mathcal{N} = \mathcal{N}_1 \times \mathcal{N}_2 \times ... \times \mathcal{N}_K \) where \( \mathcal{N}_k = \{1, 2, ..., N_k\} \) is the state-space of \( \{Z_k(t), t \geq 0\} \). The generator of \( \{Z(t), t \geq 0\} \) is given by

\[
G = G_1 \oplus G_2 \oplus ... \oplus G_K,
\]  

(5)

where \( \oplus \) represent the Kronecker sum. Let

\[
p_k(i) = \lim_{t \to \infty} P\{Z_k(t) = i\}, \quad (i \in \mathcal{N}_k, k = 1, 2, ..., K),
\]  

(6)

\[
p(n) = \lim_{t \to \infty} P\{Z(t) = n\}, \quad (n \in \mathcal{N})
\]  

(7)

be the limiting distribution of \( Z \). Now let

\[
\pi(x, n) = \lim_{t \to \infty} P\{X(t) \leq x, Z(t) = n\}, \quad (x \geq 0, n \in \mathcal{N})
\]  

(8)

and

\[
\pi(x) = [\pi(x, n)]_{n \in \mathcal{N}}.
\]  

(9)
For $1 \leq j \leq J$, we use the notation
\[ \pi^j(x) = \pi(x), \quad B_{j-1} < x < B_j. \] (10)

Next we state the differential equations satisfied by $\pi^j(x)$. First we need the following notation:
\[ \Lambda^j_k = \text{diag}(\Lambda^j_{k,1}, \Lambda^j_{k,2}, ..., \Lambda^j_{k,N_k}), \quad 1 \leq j \leq J, \] (11)
\[ \Lambda^j = \Lambda^j_1 \oplus \Lambda^j_2 \oplus ... \oplus \Lambda^j_K, \quad 1 \leq j \leq J, \] (12)
\[ D^j_j = \Lambda^j - cI, \quad 1 \leq j \leq J. \] (13)

The $\{\pi^j(x), 1 \leq j \leq J\}$ satisfy the following differential equations:
\[ \frac{d}{dx} \pi^j(x)D^j_j = \pi^j(x)G, \quad B_{j-1} < x < B_j. \] (14)

Next we state the boundary conditions. For $1 \leq j \leq J$, let
\[ \mathcal{N}^j_+ = \{n \in \mathcal{N} : D^j(n,n) > 0\}, \] (15)
\[ \mathcal{N}^j_- = \{n \in \mathcal{N} : D^j(n,n) < 0\}. \] (16)

For simplicity we assume that $\{n \in \mathcal{N} : D^j(n,n) = 0\} = \emptyset$. (The final results hold true even when this assumption is not satisfied.) Note that
\[ \mathcal{N}^j_+ \cup \mathcal{N}^j_- = \mathcal{N}, \quad 1 \leq j \leq J, \] (17)
\[ \mathcal{N}^1_- \subset \mathcal{N}^2_- \subset ... \subset \mathcal{N}^J_-, \] (18)
\[ \mathcal{N}^1_+ \supset \mathcal{N}^2_+ \supset ... \supset \mathcal{N}^J_. \] (19)

With this notation we can write the boundary conditions as follows (see Elwalid and Mitra [3] for their derivation):
\[ \pi^1(0,n) = p(n), \quad n \in \mathcal{N}_+ \] (20)
\[ \pi^j(B_{j-}, n) = \pi^{j+1}(B_{j+}, n), \quad n \in \mathcal{N}^j_- \cup \mathcal{N}^{j+1}_+, \] (21)
\[ \pi^j(B_{j-}, n) = 0, \quad n \in \mathcal{N}^j_. \] (22)

The above boundary conditions are intuitive. Equation (21), for example, says that there is no mass at $(B_j, n)$ in steady state if the drift on either side of $B_j$ is away from $B_j$. The spectral solution to (14) is given by:
\[ \pi^j(x) = \sum_r \alpha_r^j \phi_r^j e^{\eta_r^j x}, \quad 1 \leq j \leq J, \] (23)
where each \((\eta^j_r, \phi^j_r)\) pair is an (eigenvalue,eigenvector) pair satisfying
\[
\eta \phi^j r = \phi^j G. \tag{24}
\]

Elwalid and Mitra in [4] show how the (eigenvalue, eigenvector) problem of the above equation involving the large \(G\) matrix can be reduced to a coupled (eigenvalue,eigenvector) problem involving the smaller matrices \(G_k\). We briefly state the main result.

For \(j, 1 \leq j \leq J\), and \(k, 1 \leq k \leq K\), define
\[
A^j_k(\eta) = \Lambda^j_k - \frac{1}{\eta} G_k. \tag{25}
\]

**Theorem 3.1 Elwalid-Mitra**

Let \(1 \leq j \leq J\) be fixed.

(i) A pair \((\eta, \phi)\) satisfies (24) if and only if the following equations hold:
\[
\begin{align*}
g^j_k(\eta) \phi^j_k &= \phi^j_k A^j_k(\eta), & 1 \leq k \leq K, \\
\sum_{k=1}^{K} g^j_k(\eta) &= c, \\
\text{and} \quad \phi &= \phi^j_1 \otimes \phi^j_2 \otimes \ldots \otimes \phi^j_K.
\end{align*}
\]

(ii) For \(\eta < 0\) the solution \(g^j_k(\eta)\) to (26) with the maximum real part is a simple real solution, called the maximal real eigenvalue, denoted by \(g^j_*(\eta)\) and it decreases monotonically from \(A^j_{k,\text{max}}(= \max_{i=1,..,N_k} \Lambda^j_{k,i})\) to \(A^j_{k,\text{mean}}(= \sum_{i=1}^{N_k} p_k(i) \Lambda^j_{k,i})\) as \(\eta\) increases from \(-\infty\) to 0.

(iii) For \(A^j_{k,\text{mean}}(= \sum_{k=1}^{K} A^j_{k,\text{mean}}) < c < A^j_{k,\text{max}}(= \sum_{k=1}^{K} A^j_{k,\text{max}})\), the dominant eigenvalue \(\eta^j_\ast\), is given by the unique solution in \((-\infty, 0)\) to
\[
\sum_{k=1}^{K} g^j_k(\eta^j_\ast) = c. \tag{29}
\]

Furthermore, \(\eta^j_\ast\) is a monotonic, strictly decreasing function of \(c \in (A^j_{k,\text{mean}}, A^j_{k,\text{max}})\)

We consider the case of distinct eigenvalues, in which case we have \(|\mathcal{N}|\) eigenvalues.

The unknowns \(\{a^j_r, 1 \leq j \leq J, r = 1, 2, \ldots, |\mathcal{N}|\}\) are obtained by using (20)-(22) which provide the correct number of equations for them.

Now, for \(1 \leq j \leq J\),
\[
\Phi(B_j) = P\{\text{Buffer Content} \geq B_j\} \tag{30}
\]
\[
= \sum_{n \in \mathcal{N}} \pi^j(B_j, n) \tag{31}
\]
Thus the QoS requirements can be roughly written as

\[ \Phi(B_j) < \epsilon_j, \quad 1 \leq j \leq J. \]  

(32)

Now that we have a way of computing the QoS requirements, we next study the asymptotic behaviour of \( \Phi(B_j) \) as buffer size gets large and the QoS requirements get more stringent.

4 Asymptotics

In this section we study the following asymptotic case:

\[ B_j \to \infty, \quad 1 \leq j \leq J, \]  

(33)

and

\[ \epsilon_j \to 0, \quad 1 \leq j \leq J, \]  

(34)

so that

\[ \frac{B_j - B_{j-1}}{B_j} \to b_j, \quad 1 \leq j \leq J, \]  

(35)

and

\[ \frac{\log \epsilon_j}{B_j} \to \gamma_j, \quad 1 \leq j \leq J, \]  

(36)

where \( b_j > 0, \sum_{j=1}^{J} b_j = 1 \) and \( -\infty < \gamma_j < 0 \). Note that the condition \( b_j > 0 \) implies that all threshold separations \( B_j - B_{j-1} \) tend to infinity in the asymptotic region.

The next theorem gives the asymptotic behaviour of \( \Phi(B_j) \).

**Theorem 4.1** In the asymptotic case described by (33) and (35) we have

\[ \Phi(B_j) = \exp\{\sum_{r=1}^{j} \eta^{r}\ast(B_r - B_{r-1})\}\{\text{constant} + o(1)\}, \quad 1 \leq j \leq J. \]  

(37)

**Proof** The proof is a tedious exercise in matrix algebra. We start by introducing the required notation:

\[ \mathcal{R}_{+}^{j} = \{ r : \eta_r^{j} \geq 0 \}, \]  

(38)

\[ \mathcal{R}_{-}^{j} = \{ r : \eta_r^{j} < 0 \}, \]  

(39)

\[ a_{i}^{j} = [a_{i}^{j}]_{i=1, \ldots, N}, \]  

(40)

\[ a_{+}^{j} = [a_{+}^{j}] : r \in \mathcal{R}_{+}^{j}, \]  

(41)

\[ a_{-}^{j} = [a_{-}^{j}] : r \in \mathcal{R}_{-}^{j}, \]  

(42)

\[ \phi_{i}^{j}(A) = [\phi_{i}^{j}(n)]_{n \in A}, \text{ for } A \subseteq N \]  

(43)

\[ \phi_{+}^{j}(A) = [\phi_{+}^{j}(n)]_{n \in A, r \in \mathcal{R}_{+}^{j}}, \]  

(44)

\[ \phi_{-}^{j}(A) = [\phi_{-}^{j}(n)]_{n \in A, r \in \mathcal{R}_{-}^{j}}, \]  

(45)
\[ p(A) = [p(n)]_{n \in A}, \]  
(46)

\[ E_j^i(x) = \text{Diag}[e^{\eta_j^i x}]_{r=1,...,N}, \]  
(47)

\[ E_+^i(x) = \text{Diag}[e^{\eta_+^i x}]_{r \in R_+^i}, \]  
(48)

\[ E_-^i(x) = \text{Diag}[e^{\eta_-^i x}]_{r \in R_-^i}. \]  
(49)

With the above notation the QoS requirements of (32) can be written as:

\[ \Phi(B_j) = a^j E_j^i(B_j) \phi^j(N) 1 \leq e_j, \text{for } 1 \leq j \leq J, \]  
(50)

where 1 is a column vector of dimension \( |N| \). Using the above notation we shall prove the theorem for the case \( J = 2 \). The general case follows in a similar fashion. The boundary conditions in (20)-(22) can be written as:

\[ a^1 \phi^1(N_+^1) = p(N_+^1), \]  
(51)

\[ a^1 E^1(B_1) \phi^1(N_+^1) = a^2 E^2(B_1) \phi^2(N_+^1), \]  
(52)

\[ a^1 E^1(B_1) \phi^1(N_+^2) = a^2 E^2(B_1) \phi^2(N_+^2), \]  
(53)

\[ a^2 E^2(B_2) \phi^2(N_+^2) = 0. \]  
(54)

Equation (54) yields:

\[ a^2_+ E^2_+ \phi^2_+ (N_+^2) + a^2_- E^2_- \phi^2_-(N_+^2) = 0, \]  
(55)

which yields

\[ a^2_+ E^2_+(B_2) = -a^2_- E^2_-(B_2) \phi^2_-(N_+^2) \phi(N_+^2)^{-1}. \]  
(56)

Next, equation (53) yields

\[ a^1 E^1(B_1) \phi(N_+^1) = a^2_+ E^2_+ (B_1) \phi^2_+(N_+^2) + a^2_- E^2_- (B_1) \phi^2_-(N_+^2) \]  
(57)

\[ = a^2_+ E^2_+ (B_2) E^2_+(B_1 - B_2) \phi^2_+(N_+^2) + a^2_- E^2_- (B_1) \phi^2_-(N_+^2) \]  
(58)

\[ = a^2_+ E^2_+(B_2) [-\phi^2_-(N_+^2) \phi^2_+(N_+^2)^{-1} E^2_+(B_1 - B_2) \phi^2_+(N_+^2) \]  
\[ + E^2_-(B_1 - B_2) \phi^2_-(N_+^2)] \]  
(59)

The last equation follows by substituting (56) in (58). Now, in the asymptotic region under consideration, \( E^2_+(B_1 - B_2) \) remains bounded above by 1, while \( E^2_-(B_1 - B_2) \) goes to \( \infty \). Hence the first term in the square brackets in (59) can be ignored in the asymptotic region. Hence we get

\[ a^2_- E^2_-(B_2) = a^1 E^1(B_1)[E^2_-(B_2 - B_1)] \]  
(60)
where $[.]$ represents a matrix that does not dependent on $B_1$ or $B_2$. Using this convention we get

\begin{equation}
\begin{aligned}
a^2 E^2(B_2) &= a^2 E^2_+ (B_2) + a^2 E^2_- (B_2) \\
&= a^2 E^2_+ (B_2)[.] \\
&= a^1 E^1_+ (B_1)[.]E^2_-(B_2 - B_1)[.]
\end{aligned}
\end{equation}

(61)

(62)

(63)

A similar analysis shows that

\begin{equation}
a^1 E^1(B_1) = p(N_1)[.]E^1(B_1).
\end{equation}

(64)

Combining (63)-(64) we get

\begin{equation}
a^2 E^2(B_2) = [.]E^1_+(B_1)[.]E^2_-(B_2 - B_1)[.].
\end{equation}

(65)

Now, (64) shows that, in the asymptotic region, $\Phi(B_1)$ is a linear combination of terms of the type $e^{\eta^1_1 B_1}, r \in \mathcal{R}_1^1$, and (65) shows that $\Phi(B_2)$ is a linear combination of terms of the type $e^{\eta^1_1 B_1 + \eta^2_2 (B_2 - B_1)}, r \in \mathcal{R}_1^1, s \in \mathcal{R}_2^2$. Using the dominant terms corresponding to $\eta^{1*}$ and $\eta^{2*}$ we get the theorem. **End Proof.**

Since there can be a non-zero probability mass at $B_j$, it is important to realize that $\Phi(B_j)$ in the above theorem is in fact $\Phi(B_j^+)$, that is, it includes that mass. The next theorem gives a sufficient condition to check if the QoS criterion is satisfied for the traffic of all classes.

**Theorem 4.2** The QoS criterion of (32) is satisfied in the asymptotic region of (33)-(36) if there exist negative numbers $\{\gamma_{jr}, 1 \leq j \leq J, 1 \leq r \leq j\}$ such that

\begin{equation}
\sum_{r=1}^{j} b_r \gamma_{jr} = \gamma_j \sum_{r=1}^{j} b_r
\end{equation}

(66)

and

\begin{equation}
\sum_{k=1}^{K} g_k^{r*}(\gamma_{jr}) < c, \quad 1 \leq r \leq j.
\end{equation}

(67)

**Proof** From Theorem 4.1, in the asymptotic region of (33)-(36) we can write

\begin{equation}
\Phi(B_j) = \frac{1}{e_j} \exp\{\delta_j B_j\} \{\text{constant} + o(1)\}, \quad 1 \leq j \leq J.
\end{equation}

(68)

where

\begin{equation}
\delta_j = \sum_{r=1}^{j} \eta^{r*} b_r - \gamma_j \sum_{r=1}^{j} b_r.
\end{equation}

(69)
Now, $\Phi(B_j)/e_j \to 0$ if $\delta_j < 0$ and $\Phi(B_j)/e_j \to \infty$ if $\delta_j > 0$ in the asymptotic region under consideration. It can be easily seen that $\delta_j < 0$ if and only if there exists a set of negative numbers $\{\gamma_{jr}, 1 \leq r \leq j\}$ satisfying (66) and

$$\eta^{rs} < \gamma_{jr} \quad 1 \leq r \leq j.$$  

(70)

Now, from EM it follows that $\eta^{rs} < \gamma_{jr}$ if $\sum_{k=1}^{K} g_k^{rs}(\gamma_{jr}) < c$. Hence the theorem follows. **End Proof.**

One basic question in priority traffic is: will the creation of several priority classes always result in savings? The next theorem shows that the answer is a conditional yes.

**Theorem 4.3** The priority differentiation results in savings if

$$\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_J.$$  

(71)

**Proof:** We shall give a very restricted proof, but it will provide an idea behind the general result. Consider two systems, each with $K$ sources, that are identical in all respects except that priority one traffic of system one is treated as priority two traffic in system two. Using an additional superscript $l$ to denote the system index $l = 1, 2$, we have

$$\lambda_{k,i}^{l,2} = \begin{cases} 
0 & \text{for } j = 1 \\
\lambda_{k,i}^{1,1} + \lambda_{k,i}^{2,1} & \text{for } j = 2 \\
\lambda_{k,i}^{1,1} & \text{for } j = 3, \ldots, J.
\end{cases}$$

(72)

This implies that

$$g_k^{j*,2}(\eta) = \begin{cases} 
g_k^{1*,1}(\eta) & \text{for } j = 1, 2 \\
g_k^{2*,1}(\eta) & \text{for } j = 3, \ldots, J.
\end{cases}$$

(73)

Now for system two, the sufficient conditions of Theorem 4.2 need to be satisfied only for $j = 2, \ldots, J$. Suppose this is the case. The condition in (67) for $j = 2$, when applied to system two, becomes

$$\sum_{k=1}^{K} g_k^{2*,1}(\gamma_{21}) < c,$$  

(74)

$$\sum_{k=1}^{K} g_k^{2*,2}(\gamma_{22}) < c.$$  

(75)

Using (73) and the fact that $g_k^{1*,1}(\eta)$ is a decreasing function of $\eta$ the above equations can be written as

$$\sum_{k=1}^{K} g_k^{1*,1}(\min\{\gamma_{21}, \gamma_{22}\}) < c.$$  

(76)

11
Since the maximum value of $\min\{\gamma_21, \gamma_22\}$ is $\gamma_2$ we see that the above equation implies

$$\sum_{k=1}^{K} g_k^{1s,1}(\gamma_2) < c. \quad (77)$$

Now, if $\gamma_2 \leq \gamma_1$, then (77) implies

$$\sum_{k=1}^{K} g_k^{1s,1}(\gamma_1) < c. \quad (78)$$

Furthermore, $g_k^{2s,1}(\eta) \leq g_k^{1s,1}(\eta)$ implies

$$\sum_{k=1}^{K} g_k^{2s,1}(\gamma_2) < c. \quad (79)$$

Thus condition (67) is satisfied for system one for $j = 1$ with $\gamma_{11} = \gamma_1$ and for $j = 2$ with $\gamma_{21} = \gamma_{22} = \gamma_2$. For $j = 3, ..., J$ the conditions in (67) are the same for both the system. Thus, if the sufficient conditions of Theorem 4.2 are satisfied for system two, then they are satisfied for system one, provided $\gamma_2 \leq \gamma_1$. This proves the result in this case. The general result follows in a similar fashion. **End Proof.**

Note that the condition in (71) implies that the ratio $B_j/B_{j-1}$ must be larger than $log(\epsilon_j)/log(\epsilon_{j-1})$ in order to achieve savings through priority differentiation. In fact, when $log(\epsilon_j)/B_j$ is a constant, (independent of $j$) one can see that the bandwidth requirements of a source with priority differentiation is the same as the one which treats all traffic as the highest priority traffic. This analytic insight is itself a benefit of this analysis.

We note here that in [1], Choudhury, Lucantoni and Whitt discuss the notion of effective bandwidth and point out some of its drawbacks. In general, an effective bandwidth approximation based only on large buffer asymptotics can significantly over-estimate or under-estimate the number of sources that can be multiplexed on a trunk. If the sources are more 'bursty' than Poisson, then effective bandwidth tends to under-estimate the number of sources that can be multiplexed, and if the sources are less 'bursty' then the number of sources are over-estimated. This argues that effective bandwidth results should be used carefully, perhaps combined with other procedures as in [5].

## 5 Effective Bandwidth Vectors

The sufficient condition of Theorem 4.2 can be written in the following form:

$$e^{j*}(K) < c, \quad 1 \leq j \leq J, \quad (80)$$
where
\[ c^j_s(K) = \min_{(\gamma_{j_1}, \ldots, \gamma_{j_J}) \in \Delta_J} \max_{r=1, \ldots, j} \left\{ \sum_{k=1}^{K} g^{r,s}_k(\gamma_{jr}) \right\}, \quad 1 \leq j \leq J, \] (81)

where
\[ \Delta_J = \{(x_1, \ldots, x_J) : x_r < 0, 1 \leq r \leq j, \sum_{r=1}^{j} x_r b_r = \gamma_{j} \sum_{r=1}^{j} b_r\}. \] (82)

(83)

This leads to the following natural definition:

**Definition.** The vector
\[ c^*(K) = (c^1_s(K), c^2_s(K), \ldots, c^J_s(K)) \] (84)
is called the **effective bandwidth vector** for the system of K sources multiplexed onto a single buffer. However, it is clear from the definition that the effective bandwidth vector of the K-source system cannot be represented as a sum of K effective bandwidth vectors, each associated with a single source. In the next section we study one candidate for effective bandwidth vectors of a single source, which will serve as an approximation.

### 5.1 An Approximate Bandwidth Vector

Here we consider the case of independent and identical sources. Thus the generator matrix of the external environment driving each source is \( G \), and the rate at which a source produces traffic of class \( j \) in state \( i \) is \( \lambda^j_i \). Thus the maximal real eigenvalues are the same for all the sources, i.e.,
\[ g^{j,s}_k(\eta) = g^{j,s}(\eta). \] Then we get
\[ c^j_s(K) = \min_{(\gamma_{j_1}, \ldots, \gamma_{j_J}) \in \Delta_J} \max_{r=1, \ldots, j} \left\{ \sum_{k=1}^{K} g^{r,s}_k(\gamma_{jr}) \right\} \] (85)
\[ = K c^j_s \] (86)

where
\[ c^j_s = \min_{(\gamma_{j_1}, \ldots, \gamma_{j_J}) \in \Delta_J} \max_{r=1, \ldots, j} \{ g^{r,s}(\gamma_{jr}) \}. \] (87)

Thus we can define
\[ c^* = (c^1_s, c^2_s, \ldots, c^J_s) \] (88)
as the effective bandwidth vector of a single source. (This is a direct generalization of the effective bandwidth vector defined in [10]). Thus as long as the sources are identical and independent, the effective bandwidth vector of the K-source system is the sum of the effective bandwidth vectors of the individual sources, i. e.,
\[ c^*(K) = K c^*. \] (89)
This motivates us to study the effective bandwidth vector defined by (88) as a candidate even if the sources are distinct.

Now consider the $K$ distinct sources as described in Section 2. For the $k^{th}$ source, define the effective bandwidth vector as $c^*_k = (c^1_k, c^2_k, \ldots, c^j_k)$ where

$$c^j_k = \min_{(\gamma_{j1}, \ldots, \gamma_{jj}) \in \Delta_j} \max_{r=1, \ldots, j} \{ g^*_r(\gamma_{jr}) \}$$

(90)

The question is, how does $\sum_{k=1}^K c^*_k$ compare with the system bandwidth vector $c^*(K)$ defined by (81)? Unfortunately, the sum provides neither an upper bound nor a lower bound. It only provides an approximation. This is because moving the sum in (81) outside the max operator increases the right hand side, while further moving it outside the min operator decreases the right hand side.

The next theorem gives an important result. Let $\gamma_{jr}^i, 1 \leq r \leq j$ be the point where the minimum on the right-hand side of (90) is achieved. Also let $\gamma_{jr}^i, 1 \leq r \leq j$ be the point where the minimum on the right-hand side of (81) is achieved. We introduce the following notation:

$$q_{jr}^i = \max_{k=1,2,\ldots,K} \gamma_{kr}^i, \quad 1 \leq r \leq j \leq J,$$  (91)

$$q^i_j = (q^i_1, q^i_2, \ldots, q^i_j),$$  (92)

$$w_i = \frac{1}{b_i} \sum_{r=1}^j (\gamma^i_j - q^i_r)b_r, \quad 1 \leq i \leq j \leq J,$$  (93)

$$\Gamma^i_j = (\Gamma^i_1, \Gamma^i_2, \ldots, \Gamma^i_j)$$

$$= q^i_j + w_i \epsilon^j_i, \quad 1 \leq i \leq j \leq J,$$  (94)

where $\epsilon^j_i$ is a vector with $j$ components, all of which are zero, except the $i^{th}$ one, which is equal to one. Note that the definition of (95) implies that $\Gamma^i_j \in \Delta_j$.

**Theorem 5.1** The vector $(\gamma_{j1}^i, \gamma_{j2}^i, \ldots, \gamma_{jj}^i)$ is a convex combination of the $j$ vectors $\{\Gamma^i_j, i = 1, 2, \ldots, j\}, 1 \leq j \leq J$.

**Proof:** The theorem is obvious for $j = 1$ since in that case $\Delta_1 = \{(\gamma_1)\}$ is a singleton. Now consider the case where $2 \leq j \leq J$. We further assume that the vectors $\{\gamma^i_k = (\gamma_{j1}^i, \ldots, \gamma_{jj}^i), k = 1, \ldots, K\}$ lie in the interior of $\Delta_j$. (The theorem holds even when they are on the boundary, but the proof is more tedious.) Then, from (90), we have

$$g^*_r(\gamma_{jr}^i) = c^i_r, \quad \text{for all } 1 \leq r \leq j.$$  (96)
Now consider a fixed \( r, 1 \leq r \leq j \). Since \( g_k^i(\eta) \) is a decreasing function of \( \eta \), it follows that for all \( x \in \Delta_j \) with \( x_r \geq \gamma_k^{jr*} \) we have \( g_k^i(x_r) \leq c_k^i \) and there exists at least one \( i \neq r \) such that \( g_k^i(x_i) \geq c_k^i \). Hence

\[
\max_{i=1,\ldots,j} g_k^i(x_i) = \max_{i=1,\ldots,j; i \neq r} g_k^i(x_i). \tag{97}
\]

Now let \( x \in \Delta_j \) be such that \( x_r < \gamma_k^{jr*} \). Then there exists a \( y \in \Delta_j \) such that \( x_r < y_r \leq \gamma_k^{jr*} \) and \( y_i < x_i \), for \( i \neq r \). Hence we have

\[
\max_{i=1,\ldots,j} g_k^i(x_i) = \max_{i=1,\ldots,j; i \neq r} g_k^i(x_i) \geq \max_{i=1,\ldots,j; i \neq r} g_k^i(y_i) = \max_{i=1,\ldots,j} g_k^i(y_i). \tag{98}
\]

Hence, for each \( r, 1 \leq r \leq j \), and \( x \in \Delta_j \) such that \( x_r > q_k^{i_r} \) there is a \( y \in \Delta_j \) such that \( x_r > y_r \geq q_k^{i_r} \) and \( y_i > x_i \), for \( i \neq r \), and

\[
\max_{i=1,\ldots,j} g_k^i(x_i) \geq \max_{i=1,\ldots,j} g_k^i(y_i) \tag{100}
\]

for all \( k = 1,\ldots,K \). This implies that \( \gamma^{jr*} \leq q_k^{i_r} \). The set \( \{ x \in \Delta_j : x_r \leq q_k^{i_r} \} \) is the convex hull of the vectors \( \{ \Gamma_i^{jr*}, i = 1,\ldots,j \} \). Hence the theorem follows. \textbf{End Proof.}

The above theorem can be used in two ways: First, it reduces the space over which one has to search for the system-optimal vector \( (\gamma_1^{*}, \ldots, \gamma_j^{*}) \). Secondly, it provides another candidate for an effective bandwidth vector with desirable properties, as described in the next section.

### 5.2 Additive Bandwidth Vectors

We begin by defining a class of effective bandwidth vectors. For \( k = 1,2,\ldots,K \) define

\[
e_k^{i*} = \max_{r=1,\ldots,j} \{ g_k^{i*}(\xi_r) \}, \quad \text{for } 1 \leq j \leq J \tag{102}
\]

where \( \{ \xi_r, 1 \leq r \leq j \leq J \} \) is a given set of numbers so that \( \xi_j = (\xi_{j1}, \ldots, \xi_{jj}) \in \Delta_j \). Define

\[
e_k^* = (e_k^{1*}, e_k^{2*}, \ldots, e_k^{J*}). \tag{103}
\]

We study \( e_k^* \) as a possible candidate for an effective bandwidth vector for the \( k^{th} \) source.

\textbf{Theorem 5.2}

\[
(i) e_k^* \geq c_k^*, \quad \text{where } c_k^* \text{ is as defined in (90).} \tag{104}
\]
\[(ii) \sum_{k=1}^{K} e_k^* \geq c^*(K) \text{ where } c^*(K) \text{ is defined by (81).} \quad (105)\]

**Proof:**

(i) Using (102) and (90) we get

\[e_k^* = \max_{r=1, \ldots, j} \{g_k^*(\xi_{jr})\}, \quad (106)\]

\[\geq \min_{r=1, \ldots, j} \max_{k=1} \{g_k^*(\gamma_{jr})\} \quad (107)\]

\[= c_k^*. \quad (108)\]

The inequality follows because \(\xi_{jr}^*\) belongs to the set of \(\gamma_{jr}\) over which the minimum is taken.

(ii) Using (102) and (90) we get

\[\sum_{k=1}^{K} e_k^* = \sum_{k=1}^{K} \max_{r=1, \ldots, j} \{g_k^*(\xi_{jr})\} \quad (109)\]

\[\geq \max_{r=1, \ldots, j} \sum_{k=1}^{K} \{g_k^*(\xi_{jr})\} \quad (110)\]

\[\geq \min_{r=1, \ldots, j} \sum_{k=1}^{K} \{g_k^*(\gamma_{jr})\}. \quad (111)\]

**End Proof.**

The property in (105) provides the justification for calling \(e_k^*\) an additive bandwidth vector.

Note that \(e_k^*\) is not an effective bandwidth vector of the usual type, i.e., it is not defined in terms of the characteristics of the \(k^{th}\) source alone. In fact it depends upon the source characteristics of all the sources in the system. However, it still provides an implementable method of call admission.

Next we consider two special sets of vectors \(\{\xi_j, 1 \leq j \leq J\}\).

1. **Average of Extreme Points.** As a first choice, consider

\[\xi_j = \frac{1}{j} \sum_{i=1}^{j} \Gamma_{i}^{j*}, 1 \leq j \leq J \quad (112)\]

where \(\Gamma_{i}^{j*}\) is an extreme point of \(\Delta_j\) as defined in (94). Clearly, \(\xi_j \in \Delta_j\) for all \(1 \leq j \leq J\).

2. **Average over Sources.** As a second choice, consider

\[\xi_{jr} = \frac{1}{K} \sum_{k=1}^{K} \gamma_{kr}^{j*}, 1 \leq r \leq j \leq J \quad (113)\]

where \(\gamma_{kr}^{j*} = (\gamma_{k1}^{j*}, \ldots, \gamma_{kj}^{j*})\) is the point where the minimum over the right hand side of (90) is achieved. Thus the vector \(\xi_j\) is an average of the vectors \(\gamma_{kr}^{j*}\) and hence is in \(\Delta_j\).
5.3 A Special Case

Here we consider a special case where each source produces a single priority fluid. However, the total number of priorities is still $J$. The motivation is to study the multiplexing of sources of differing QoS requirements into a single buffer.

We say that a source is of type $j$ if it produces traffic with QoS requirement $\epsilon_j$ (i.e. priority $j$ traffic). Let $K_j$ be the set of sources of type $j$. Now consider a source $k \in K_t$. We have

$$\lambda_{k,i}^j = \begin{cases} 0 & \text{if } j \neq t \\ \lambda_{k,i}^j & \text{if } j = t \end{cases}$$

(114)

and

$$A_{k,i}^j = \begin{cases} 0 & \text{for } t < j \leq J \\ \lambda_{k,i}^j & \text{for } 1 \leq j \leq t \end{cases}$$

(115)

This, along with (25) implies that

$$A_{k}(\eta) = \begin{cases} A_k^j(\eta) & \text{for } 1 \leq j \leq t \\ \frac{1}{\eta} G_k & \text{for } t < j \leq J \end{cases}$$

(116)

which further implies that

$$g_{k}^j(\eta) = \begin{cases} g_k^j(\eta) & \text{for } 1 \leq j \leq t \\ 0 & \text{for } t < j \leq J \end{cases}$$

(117)

Now consider the computation of the approximate bandwidth vector $c_k^*$ as defined in (90). We have

$$c_{k}^{j*} = \min_{(\gamma_{j1},...,\gamma_{j2}) \in \Delta_j} \max_{r=1,...,J} \{ g_{k}^{j*}(\gamma_{jr}) \}, \quad 1 \leq j \leq J$$

(118)

$$= \min_{(\gamma_{j1},...,\gamma_{j2}) \in \Delta_j} \max_{r=1,...,J} \{ g_{k}^{t}(\gamma_{jr}) \}, \quad 1 \leq j \leq t,$$

(119)

$$= \min_{(\gamma_{j1},...,\gamma_{j2}) \in \Delta_j} \max_{r=1,...,t} \{ g_{k}^{t}(\gamma_{jr}) \}, \quad t < j \leq J.$$  

(120)

The above minimum is achieved at $\gamma_{jr} = \gamma_k^{j*}$ where

$$\gamma_k^{j*} = \begin{cases} \gamma_j & \text{for } 1 \leq r \leq j \leq t \\ 0 & \text{for } 1 \leq r \leq j, t < j \leq J \end{cases}$$

(121)

Thus computing the $\xi_j$ vectors (using the average over sources method) for the additive bandwidth vectors using (25) is easy. We get

$$\xi_{jr} = \frac{\sum_{t=1}^{r} |K_t|}{K} \gamma_{j}, \quad 1 \leq r \leq j \leq J.$$  

(122)

Note that this method is particularly easy to use, since it involves using the fixed parameters $\gamma_j$ and keeping track of $|K_t|$ for each $t = 1, 2, ..., J$. 

17
6 Computational Algorithm

We begin by describing a numerical algorithm to compute the minimum in (81) and (90). The algorithm uses ideas of steepest descent as well as of binary search, and is based upon the following simple theorem, which we state without proof.

**Theorem 6.1** Let $h_i : (-\infty, 0] \to [0, \infty), 1 \leq i \leq j$ be bounded monotone decreasing functions. Let $(\eta_1, \eta_2, ..., \eta_j)$ be the value of $x = (x_1, x_2, ..., x_j)$ where the function $h(x) = \max_{i=1,2,...,j} h_i(x_i)$ achieves its minimum over $x \in \Delta_j$. If, for a given $x \in \Delta_j$, $h_r(x_i) = \max_{i=1,...,j} \{h_i(x_i)\}$ then, $\eta_r \geq x_r$.

This is a direct result of the structure of the minimization problem and the monotone nature of the $h_i$ functions. It yields the following algorithm to find $(\eta_1, \eta_2, ..., \eta_j)$.

**ALGORITHM A:**

**Given:** Monotone decreasing bounded functions $h_i : (-\infty, 0] \to [0, \infty)$. Two small positive numbers $\text{tolerance1}$ and $\text{tolerance2}$ to dictate the stopping criterion.

**Aim:** To find $(\eta_1, \eta_2, ..., \eta_j)$ as defined in Theorem 6.1.

**Step 0:** Set $L = \gamma_j \sum_{i=1}^{j} b_i (\frac{1}{b_1}, \frac{1}{b_2}, ..., \frac{1}{b_j})$.

**Step 1:** Compute

$$\delta = \frac{\gamma_j \sum_{i=1}^{j} b_i}{\sum_{i=1}^{j} L_i b_i}.$$ 

Set $x = \delta L$.

**Step 2:** Let

$$h_{\max} = \max_{i=1,2,...,j} h_i(x_i)$$

and

$$h_{\min} = \min_{i=1,2,...,j} h_i(x_i).$$

If $h_{\max} - h_{\min} \leq \text{tolerance1}$, stop. $x$ is the desired $\eta$.

If $-\max_{i=1,2,...,j} \{x_i\} \leq \text{tolerance2}$, stop. $x$ is the desired $\eta$.

**Step 3** Set $L_r = x_r$ and go to step 1.

**Remark:** If the desired $\eta$ vector is in the interior of $\Delta_j$, the algorithm stops due to the first
stopping condition in Step 2. If \( \eta \) is on the boundary of \( \Delta_j \), the algorithm stops due to the second stopping condition in Step 2.

Note that this algorithm can be used to compute \( c^*_k \) in (90) as well as \( c^*(K) \) in (85) because both \( g^*_k(x_i) \) and \( \sum_{k=1}^{K} g^*_k(x_i) \) satisfy the properties of Theorem 6.1. Once the algorithm produces the \( (\gamma^*_k) \) vectors, they can be used to compute the additive bandwidth vector defined by (103). This algorithm is used in all the numerical work reported below.

7 Numerical Results

Though most of this paper has concentrated on theoretical results, and the construction of approximate bandwidth vectors, we will show some numerical results. One can reasonably ask whether the procedures outlined above actually produce enough benefit to be worth while. In this short section, we give an example to show that priority differentiation may indeed be worth while. We use the same example as in [10]. Here we extend the voice source example of that paper to explore the effects of multiple priorities and multiple homogeneous sources.

Once again, we assume that the peak bit rate, including ATM overhead, is 72,170 bits per second. The parameters of the two-state Markov chain controlling the source are: average talkspurt = 350 ms. and average silence duration = 650 ms. We used four different configurations:

1. No priority differentiation.

   In this case, there is only one buffer threshold, 100 cells, which is the size of the buffer. The target loss probability used was the most stringent requirement of the traffic stream, \( 10^{-10} \).

2. Two priorities.

   In this case, we used two buffer thresholds: 55 cells, for low priority traffic and 100 cells, for high priority traffic. The target loss probability for low priority traffic is \( 10^{-1} \) and for high priority traffic it is \( 10^{-10} \). The peak bit rate is split evenly between the priorities.

3. Three priorities.

   For three priorities, we set the buffer thresholds at 55 cells (low priority), 91 cells (medium priority) and 100 cells (high priority). The target loss probabilities we set to \( 10^{-1} \) for low priority, \( 10^{-7} \) for medium priority, and \( 10^{-10} \) for high priority. We split the peak bit rate among the priorities as follows: 25% each for high and low priority traffic and 50% for medium priority.
<table>
<thead>
<tr>
<th>Sources</th>
<th>Number of Priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>1,515,568</td>
</tr>
<tr>
<td>22</td>
<td>1,587,738</td>
</tr>
<tr>
<td>23</td>
<td>1,659,908</td>
</tr>
<tr>
<td>24</td>
<td>1,732,078</td>
</tr>
<tr>
<td>25</td>
<td>1,804,248</td>
</tr>
</tbody>
</table>

Table 1: Homogeneous sources: Total bandwidth used

4. Four priorities.

Finally, for four priorities we set the buffer thresholds to 55 cells, 60 cells, 91 cells, and 100 cells (lowest priority to highest). We set the target loss probabilities to $10^{-1}$, $10^{-4}$, $10^{-7}$, and $10^{-10}$ (again, lowest priority to highest). The peak bit rate was again split evenly among all the priorities.

In this experiment, we computed the additive effective bandwidth vector $e_k^*$ defined in (103). Recall that the $e_k^*$ vectors provide an upper bound for $c_k^*(K)$ (105) but that since all the sources in this case are identical, the system effective bandwidth vector and the additive effective bandwidth vectors are the same.

Table 1 shows the result of the computation of the effective bandwidth for the four different configurations described above. The entries in Table 1 show the total effective bandwidth (in bits per second) required to support a given number of sources with a given number of priorities differentiated. We can look at the savings achieved from two different points of view: First is to look at the total effective bandwidth required by each of the priority configurations. Having two priorities saves about 8% of the bandwidth over no differentiation. Three priorities saves about 11% of the bandwidth and having four priorities saves about 15% of the bandwidth in this case.

Second, we can look at the savings from the point of view of the number of connections that would fit on a 1.544 Mb/s link. Note that we have ignored the usual framing and that also, ATM overhead prevents the usual 24 64Kb/s connections from fitting in the DS1.

In this example, using 4 priorities, we can fit about 19% more connections into the given bandwidth than with no priority differentiation. With 2 and 3 priorities, we can fit 9.5% more connections.

In this case, we did not find optimal buffer thresholds, as we did in [10] and so the bandwidth savings are less than shown in that paper.
References


