Overflow probability upper bound for heterogeneous fluid queues handling general on-off sources

J. Guibert

CNET/PAA/ATR, France Telecom,
38-40, rue du Général Leclerc,
92130 Issy Les Moulineaux, France

Fluid queues arising from the superposition of independent and heterogeneous on-off sources with generally distributed burst volumes and silence durations are considered. In this paper, the so-called Benes upper bound for the equilibrium queue length distribution is first expressed as a contour integral. We then derive a closed formula which readily provides an asymptotic estimate for the queue length distribution tail. We evaluate the influence of some parameters in the case of a superposition of homogeneous sources and finally propose an application to the derivation of effective bandwidth.

INTRODUCTION

In high-speed packet switched networks, a critical problem is to evaluate the probability of buffer overflow when the superposition of a finite number of variable bit rate sources representing voice, data or video traffic, is offered to a statistical multiplexer. This is the basis of studies on traffic control procedures which aim at enabling the network to meet quality of service constraints appropriately defined for various classes of services. Our objective in the present paper is to provide analytic tools for expressing the distribution of the multiplexer queue length under fairly general assumptions about the incoming traffic.

It is well known [1-3] that fluid models, where discrete packet arrivals are assimilated to fluid flows, provide a useful description for studying such queues. In [4], where a Markovian input process is assumed, the complementary equilibrium queue length $Q(x)$ has been explicitly obtained. The use of the method of stages enabled the generalisation of the previous technique to Erlang and Hyper-Exponential distributions for input sources [1] but only asymptotic behaviour has been numerically considered there. An upper bound of the overflow probability has been introduced in [5] for reversible Markov Modulated Rate Process (MMRP). However, reversibility is a serious restriction since even Erlang-2 distribution does not lead to a reversible rate process.

More recently [6], the application of a method initiated by Benes for $G/G/1$ queues [7] and extended to fluid queues [8] enabled the derivation of the so-called Benes upper bound, which has been estimated [3] through the use of the saddle-point method. However, errors in the use of that method were not theoretically controled.

This paper is organized as follows. In Section 1, we give an alternative expression
of the Beneš bound as a contour integral in the complex plane. We then evaluate in Section 2 the formula corresponding to heterogeneous input sources whose burst volume and silence duration distributions have rational Laplace transform. In Section 3, we simplify the formulas in the homogeneous case enabling us to show the influence of some parameters on the behaviour of \(Q(x)\). Finally, in Section 4, we propose a definition for the effective bandwidth used by a source.

1. GENERAL BENĚŠ FORMULA

1.1. Model Description

We consider a fluid reservoir with constant output rate \(c\) and a stationary input process consisting of the superposition of \(N\) independent on/off sources, source \(i\) being defined as follows:

- alternating bursts \(B_i\) and silences \(A_i\) are independent;
- each silence duration has density \(a_i(t)\);
- each burst volume has density \(b_i(w)\);
- expected silence duration is \(1/\alpha_i\) and expected burst volume is \(1/\beta_i\);
- during a burst, work is generated at rate \(h_i\);
- mean cycle duration is \(D_i = 1/\alpha_i + 1/(h_i\beta_i)\) and the activity rate is \(\nu_i = 1/(h_i\beta_iD_i)\).

![Figure 1](image.png)

Denote by \(V_t\) the stationary process representing the work in system at epoch \(-t\), where \(0\) is an arbitrary instant, and let \(\Lambda_t\) be the input rate at \(-t\). Let \(W_t\) be the incoming workload over time interval \([-t,0)\), and let \(Q\) be the complementary distribution function of \(V_0\): \(W_t = \int_0^\infty \Lambda_u \, du\) and \(Q(x) = P(V_0 > x), x \geq 0\). The Beneš result for on/off fluid sources reads [3]:

\[
Q(x) = \int_0^\infty \Phi(t, x + ct) \, dt ,
\]

where \(\Phi(t, w) = \sum_{\varepsilon : \varepsilon \cdot h < w} (c - \varepsilon \cdot h) \, P(\Lambda_t = \varepsilon \cdot h; w < W_t \leq w + dw; V_t = 0) / dw\), and \(\varepsilon\) is an \(N\)-vector with 0-1 coordinates.
The following upper bound derived from (1) on ignoring the event \((V_i = 0)\) is a good approximation in many fluid systems of practical interest:

\[
Q(x) \leq q(x) = \int_0^\infty \varphi(t, x + ct) \, dt ,
\]

where \(\varphi(t, w) = \sum_{\epsilon : \epsilon \cdot h \leq \epsilon} (c - \epsilon \cdot h) \, \nu(\Lambda_1 = \epsilon \cdot h; w < W_i \leq w + dw) / dw .\)

This bound is accurate when \(c\) is much larger than the \(h_i\) or when the system load \(\rho = \sum_{i \in I} \frac{h_i}{h_i} \) is not too close to 1, i.e., when \(P(V_i = 0) \approx 1\). Nevertheless, we can still give a good approximation in all cases by considering the “heuristic” corrective multiplying factor given by [3]:

\[
\theta = c(1 - \rho) / \sum_{\epsilon : \epsilon \cdot h \leq \epsilon} (c - \epsilon \cdot h) \, \nu(1 - \nu)^{1 - \epsilon} ,
\]

where \(\nu = (\nu_i)_{i \in I}\) is the vector of the activity rates, and by convention \(u^y = \prod_{t \in I} u^y_t\).

1.2. Expressing the Beneš Result

In formula (2), the density \(\varphi(t, w)\) cannot be numerically computed directly in general. Actual computation requires its Laplace transform with respect to variable \(w\):

\[
\varphi^*(t, z) = \sum_{\epsilon : \epsilon \cdot h \leq \epsilon} (c - \epsilon \cdot h) \, [p^e_A(t, z)]^\epsilon \, [p^e_A(t, z)]^{(1 - \epsilon)} ,
\]

where \(p_A(t, \cdot)\) (resp. \(p_B(t, \cdot)\)) is the product of the densities \(p_A(t, \cdot)\) (resp. \(p_B(t, \cdot)\)) of the incoming workload due to source \(i\) which was “off” (resp. “on”) at epoch \(-t\). These densities are Dirac at 0 or \(h_i t\), respectively, so that \(\varphi(t, + ct)\) does not have any Dirac component.

We need now to invert the Laplace transform \(\varphi^*(t, z)\) and then integrate the result from 0 to \(\infty\). Formally, we use the inversion formula [9] which is valid for the smooth function \(\varphi(t, + ct)\):

\[
\varphi(t, x + ct) = \frac{1}{2i\pi} \int_\mathcal{L} \varphi^*(t, z) \, e^{(x + ct)z} \, dz ,
\]

where \(\mathcal{L}\) is a line of integration we may take strictly to the left of the imaginary axis. We thus write:

\[
q(x) = \int_0^\infty \left( \frac{1}{2i\pi} \int_\mathcal{L} \varphi^*(t, z) \, e^{(x + ct)z} \, dz \right) \, dt .
\]

Keeping a common line \(\mathcal{L}\) of integration, we can invoke Fubini’s theorem to invert the order of integration of the convergent integrand, which yields finally:

\[
q(x) = \frac{1}{2i\pi} \int_\mathcal{L} \varphi^*(-cz, z) \, e^{cz} \, dz .
\]

This reasoning can be made precise by first obtaining the Laplace transform of \(q(x)\) which is given by:

\[
q^*(z) = \varphi^*(-cz, z) - \int_0^\infty \int_0^t \varphi(t, w) \, e^{z(t-w)} \, dw \, dt ,
\]
for $\Re(z) < 0$, then showing that the last component in formula (7) is analytic and has null residue in the half plane defined by $\Re(z) < 0$. The basic assumption which is needed is the analyticity of the Laplace transforms $a_i^*(s)$ and $b_i^*(z)$ at 0 [10].

The general formula (6) is the basis of this paper and the key for obtaining the closed formulas of the next sections.

2. EVALUATING THE BENEŠ FORMULA

We can apply formula (6) when silence duration and burst volume densities have Cox-type distributions:

$$
\begin{align*}
    a_i^*[s] &= \sum_{j=1}^{k_{i,0}} \frac{\sigma_i \cdots \sigma_{i,j-1}(1 - \sigma_{i,j})}{(1 + \frac{s}{\sigma_{i,j}}) \cdots (1 + \frac{s}{\sigma_{i,j-1}})} , \quad 0 < \sigma_{i,j} \leq 1 , \quad \sigma_{i,0} = 1 , \quad \sigma_{i,k} = 1 , \quad \sum_{j=1}^{k_{i,0}} \frac{\sigma_{i,j-1}}{\sigma_{i,j}} = \frac{1}{\rho_i} , \\
    b_i^*[s] &= \sum_{j=1}^{k_{i,0}} \frac{\theta_i \cdots \theta_{i,j-1}(1 - \theta_{i,j})}{(1 + \frac{s}{\theta_{i,j}}) \cdots (1 + \frac{s}{\theta_{i,j-1}})} , \quad 0 < \theta_{i,j} \leq 1 , \quad \theta_{i,0} = 1 , \quad \theta_{i,k} = 1 , \quad \sum_{j=1}^{k_{i,0}} \frac{\theta_{i,j-1}}{\theta_{i,j}} = \frac{1}{\beta_i} .
\end{align*}
$$

This is not a serious restriction since these distributions, which have a rational Laplace transform, form a dense subset of all possible distributions.

2.1. Obtaining a Closed Formula for the Beneš Bound

We need to evaluate the Laplace transform of the workload density. Observing that the distribution of the burst duration is $\frac{1}{h_i} b_i(h_i)$, we can employ the formulas given in [6] which now read:

$$
\begin{align*}
    p_{Ai}^*(s,z) &= \frac{1 - \nu_i}{s} + h_i \nu_i (1 - \nu_i) D_i s \frac{z}{s} \frac{\tilde{a}_i^*[s] \tilde{b}_i^*[s/h_i + z]}{\tilde{a}_i^*[s] \tilde{b}_i^*[s/h_i + z]} , \\
    p_{Bi}^*(s,z) &= \frac{\nu_i}{s + h_i z} + h_i \nu_i (1 - \nu_i) D_i s \frac{z}{s + h_i z} \frac{\tilde{a}_i^*[s] \tilde{b}_i^*[s/h_i + z]}{\tilde{a}_i^*[s] \tilde{b}_i^*[s/h_i + z]} ,
\end{align*}
$$

where $\tilde{a}_i[.]$ (resp. $\tilde{b}_i[.]$) is the residual density of $A_i$ (resp. $B_i$), whose Laplace transform is given by

$$
\tilde{a}_i^*[s] = \alpha_i \frac{1 - a_i^*[s]}{s} , \quad \tilde{b}_i^*[s] = \beta_i \frac{1 - b_i^*[s]}{s} ,
$$

respectively.

Since all the functions involved in (8) are rational, it is straightforward to invert these formulas with respect to $s$:

$$
\begin{align*}
    p_{Ai}^*(t, z) &= \bar{r}_{Ai}(z) \cdot e^{\bar{s}_i(t)} , \\
    p_{Bi}^*(t, z) &= \bar{r}_{Bi}(z) \cdot e^{\bar{s}_i(t)} ,
\end{align*}
$$

where $\bar{s}_i(z)$ is the vector solution of the equation:

$$
a_i^*[s] \tilde{b}_i^*[s/h_i + z] = 1 ,
$$

and $\bar{r}_{Ai}(z)$ (resp. $\bar{r}_{Bi}(z)$) is the residue of $p_{Ai}^*(s, z)$ (resp. $p_{Bi}^*(s, z)$) at $\bar{s}_i(z)$.
We proceed by evaluating the double Laplace transform $\varphi^{**}(-cz, z)$. First,

$$
\varphi^{**}(t, z) = \sum_{\varepsilon : \varepsilon \cdot h < 0} \sum_{\bar{\varepsilon} \leq \bar{t}} (c - \varepsilon \cdot h) \bar{F}_H(z)^{\bar{\varepsilon}} \bar{F}_A(z)^{\bar{t} - \bar{\varepsilon}} \cdot \varepsilon \cdot \bar{s}(z). 
$$

(11)

where $\bar{F}_H = (\bar{F}_{H_i})_{i \leq N}, \bar{F}_A = (\bar{F}_{A_i})_{i \leq N}, \bar{t} = (\bar{t}_i)_{i \leq N}$ and $\bar{t}_i$ is a $t_i$-vector ($t_i = k_{ai} + k_{bi}$) with all coordinates 0 except one which is 1, and $\bar{t} \cdot \bar{s} = \sum_{i \leq N} \bar{t}_i \cdot \bar{s}_i$.

Taking the Laplace transform with respect to variable $t$ in (11), we obtain:

$$
\varphi^{**}(-cz, z) = \sum_{\bar{t}} \frac{-1}{(\bar{t} \cdot \bar{s}(z)) + c} \sum_{\varepsilon : \varepsilon \cdot h < 0} (c - \varepsilon \cdot h) \bar{F}_H(z)^{\bar{\varepsilon}} \bar{F}_A(z)^{\bar{t} - \bar{\varepsilon}}.
$$

(12)

Finally,

$$
q(x) = \sum_{\bar{t} \cdot \bar{s}(z) < 0, x} \kappa(\bar{t}, z) e^{\bar{z}x},
$$

(13)

with $\kappa(\bar{t}, z) = \frac{-1}{(\bar{t} \cdot \bar{s}(z)) + c} \sum_{\varepsilon : \varepsilon \cdot h < 0, \bar{\varepsilon} \leq \bar{t}} (c - \varepsilon \cdot h) \bar{F}_H(z)^{\bar{\varepsilon}} \bar{F}_A(z)^{\bar{t} - \bar{\varepsilon}}$.

This formula shows that $q(x)$ is a finite sum of exponential terms whose exponents are complex numbers having negative real part. The existence of these complex exponents derives from the “non-linearity” occurring when burst volumes and silence durations have general densities.

2.2. Asymptotic Results for the Beneš Bound

Formula (13) is somewhat intricate, so that we naturally look for some asymptotic estimates. Moreover, the formula will be used when the queue length distribution tail $q(x)$ is small which implies in general that $x$ is large. It is clear that the exponent $z$ with larger negative real part in (13) gives the asymptotic slope.

We order the roots $\bar{s}_i(z)$ by decreasing real part values. The maximal root of the equation:

$$
\bar{t} \cdot \bar{s}(z) + cz = 0
$$

(14)

will then be attained for $\bar{t} = ((1, 0, \cdots, 0), \cdots, (1, 0, \cdots, 0)), so that we have to solve the equation:

$$
\sum_{i \leq N} s_i(z) + cz = 0,
$$

(15)

where $s(z)$ is the vector of maximal roots in the $\bar{s}_i(z)$.

We thus obtain the following expression of the asymptotic for $q(x)$:

$$
q(x) \sim K(\eta) e^{-\eta x} \text{ as } x \to +\infty,
$$

(16)

where

- $\eta > 0$ is the smallest positive root of the equation.
\[ \sum_{i \leq N} s_i(-u) = cu \]  \hspace{1cm} (17)

- \( K(\eta) = \frac{-1}{[\sum_{i \leq N} x_i(-\eta + \varepsilon)]} \cdot \sum_{\varepsilon < h < \varepsilon} (c - \varepsilon \cdot h) r_H(-\eta)^\varepsilon \cdot r_A(-\eta)^{t-\varepsilon} \).

We can further approximate the above formula (16) when system load is high. If \( \eta \) is small, then the \( s_i(-\eta) \) are small, too, since the other roots of (10) necessarily have negative real part. Expanding \( a_i(s) \) and \( b_i(z) \) about 0, we finally obtain:

\[ \eta \sim 2c(1 - \rho)/(\sum_{i \leq N}(1 - \mu_i)^2) \rho_i (cv_{ai}^2 + cw_{ai}^2)/\beta_i \]  \hspace{1cm} (18)

where \( cv_{ai}^2 \) (resp. \( cw_{ai}^2 \)) is the squared coefficient of variation of silence \( i \) duration (resp. burst \( i \) volume) distribution, and \( \rho_i = h_i \mu_i / c \) the system load due to source \( i \).

Pursuing the approximation for the coefficients also yields:

\[ q(x) \sim \frac{1}{\theta} e^{-\left[2c(1 - \rho)/(\sum_{i \leq N}(1 - \mu_i)^2) \mu_i (cv_{ai}^2 + cw_{ai}^2)/\beta_i\right] x} \]  \hspace{1cm} (19)

where \( \theta \) is the corrective factor introduced in Section 1.1. The asymptotic equivalent for the exact queue length distribution tail is then:

\[ Q(x) \sim e^{-\left[2c(1 - \rho)/(\sum_{i \leq N}(1 - \mu_i)^2) \mu_i (cv_{ai}^2 + cw_{ai}^2)/\beta_i\right] x} \]  \hspace{1cm} (20)

We see here the effect of the determinism of the distributions of silence duration and burst volume. The more they are deterministic, the less buffer overflow occurs (i.e. the steeper is the asymptotic slope).

3. HOMOGENEOUS SOURCES

We can obtain simpler formulas when all the sources are identically distributed.

3.1. Closed Formula

The formula (4) in the homogeneous case reduces to:

\[ \psi(t, z) = \sum_{n, k \leq c} \binom{N}{n} (c - nh) [p_H^a(t, z)]^n [p_A^a(t, z)]^{(N-n)} \]  \hspace{1cm} (21)

We then derive from (13):

\[ q(x) = \sum_{\substack{N, \Re(z) < 0, z: \\
N \cdot \bar{s}(z) + cz = 0}} \kappa(\bar{N}, z) e^{\bar{z} x} \]  \hspace{1cm} (22)

with

\[ \kappa(\bar{N}, z) = \frac{-1}{\bar{N} \cdot \bar{s}(z) + c} \cdot \sum_{n, k \leq c, \bar{n} \leq \bar{N}} \left[ \bar{n}, \bar{N} - \bar{n} \right] (c - nh) (1 - \mu)^{N-n} \bar{r}_A(z)^{N-n} \bar{r}_H(z)^{-n} \]  \hspace{1cm} (23)

where \( \left[ \bar{n}, \bar{N} - \bar{n} \right] \) is the multinomial coefficient.
When silence durations and burst volumes have exponentially distributed densities, the exact probability of buffer overflow has been obtained in [4]. The above formulation gives the same set of real negative exponents in this special case. The formulas for the asymptotic case are more explicit than in the heterogeneous case. The term with largest negative value \( z \) in formula (22) is solution to the following equation:

\[
\alpha^*[\frac{z}{\lambda_0}] \cdot b^*[\left(1 - \frac{\alpha}{\lambda_0}\right)z] = 1 \tag{24}
\]

which gives the following expression of the asymptotic for \( q(x) \):

\[
q(x) \sim \left[\sum_{n=0}^{\infty} \binom{\lambda_0 x}{n} (1 - \lambda_0)(\frac{\lambda_0 x}{n})^{\gamma} \cdot \eta^n \right] \cdot \text{C}(\eta)^N K(\eta) e^{-\eta x} \text{ as } x \to +\infty, \tag{25}
\]

where

- \( \gamma = \frac{\lambda_0}{\lambda_0} < 1 \);
- \( \eta > 0 \) is the smallest positive root of the equation

\[
\alpha^*[\gamma h u] b^*[-(1 - \gamma)u] = 1 \tag{26}
\]

- \( \text{C}(\eta) = (1 - \nu) \rho D \frac{\alpha^*[\gamma h u] b^*[\gamma h u]}{\frac{\alpha^*}{\lambda_0} \frac{\alpha^*}{\lambda_0} - \frac{\alpha^*}{\lambda_0} \frac{\alpha^*}{\lambda_0} \frac{-(1 - \gamma)}{\lambda_0} \eta} \);
- \( K(\eta) = \frac{\eta}{\gamma} \frac{\alpha^*}{\lambda_0} \frac{\alpha^*}{\lambda_0} \frac{\alpha^*}{\lambda_0} \frac{-(1 - \gamma)}{\lambda_0} \eta} \frac{\alpha^*}{\lambda_0} \frac{\alpha^*}{\lambda_0} \frac{-(1 - \gamma)}{\lambda_0} \eta} \frac{\alpha^*}{\lambda_0} \frac{\alpha^*}{\lambda_0} \frac{-(1 - \gamma)}{\lambda_0} \eta} \)

As in the heterogeneous case, we can give a simpler formula at high load:

\[
\eta \approx \frac{2\beta}{(1 - \nu)^2} \frac{(1 - \rho)}{\rho (ce\alpha^2 + ce\alpha^2)} \tag{27}
\]

and (using the heuristic corrective factor)

\[
Q(x) \sim e^{-\frac{2\beta}{(1 - \nu)^2} \frac{(1 - \rho)}{\rho (ce\alpha^2 + ce\alpha^2)} x} \tag{28}
\]

### 3.2. Influence of the coefficient of variation

We let \( ce\alpha^2 = ce\alpha^2 = ce\alpha^2 \) vary while all other parameters remain constant (sources are deterministic for \( ce\alpha^2 = 0 \)).

We observe that the asymptotic slope is steeper when the determinism of burst volumes and silence durations is higher, i.e. when the \( ce\alpha^2 \) is smaller. This fact means that deterministic sources behave better than more variable sources.
3.3. Influence of peak rate

We show the influence of the parameter $h$ by examining the asymptotic behaviour of $q(x)$. We first note that $\nu = \frac{1}{\beta D}$, which implies $\rho := \frac{Nh\nu}{c} = \frac{N}{\beta D}$ so that workload is constant if $D$ remains constant while $\alpha$ varies.

By the implicit derivation theorem, we have:

$$\eta'(h) = \frac{c\eta}{h^2} e \left[ \frac{\nu}{h} \left( \frac{\nu}{h} \right) \frac{1}{\frac{N}{h}} \right] \frac{-1}{1 - \frac{\nu}{N/h}} \eta,$$

where the denominator is positive (the asymptotic $q(x)$ is positive) and the numerator is negative. We therefore conclude that $\eta(.)$ is decreasing, namely $h > h_0 \Rightarrow \eta(h) < \eta(h_0)$. The queue behaves better for smaller individual source rates. This fact means that the superposition of high peak rates is more damaging than the superposition of medium peak rates.

3.4. Influence of the number of multiplexed sources

We now let $N$ vary while $\frac{N}{D}$ remains constant: the cycle duration enlarges just like the number of sources to be multiplexed.

If $N$ gets large with $\frac{N}{D} \rightarrow c\beta\eta$, we obtain the Poisson limiting case where the slope is now independent of the individual rate $h$, since:

$$q(x) \approx \sum_{\nu < c} \left( c - nh \right)^{\left( c/h \right)^n} \frac{c^{1 - \rho + \rho b^\nu[-\eta]/h}}{n!} \frac{1}{-c \left[ 1 + \rho \beta b^\nu[-\eta] \right]} e^{-\nu \eta},$$

where $\eta > 0$ is the smallest positive root of equation $b^\nu[-\eta] = 1 + \frac{\nu}{\rho \beta}$.
4. EFFECTIVE BANDWIDTH

Our analytical approach makes it possible to define an effective bandwidth for source $i$ as derived for Markovian input in [11], by using relation (17). We define $g_i(u) = s_i(-u)/u$. When $u = \eta$, we have $\sum_{i \leq N} g_i(\eta) = c$, which leads us to consider $g_i(\eta)$ as the effective bandwidth used by source $i$.

Effective bandwidth is valid for $B \to \infty$ for any load, and satisfactory when $Q(x) \approx e^{-a_B}$ (i.e., when $\rho \to 1$ or $c/h \to 0$).

Under these conditions, we can approximate the probability of buffer overflow by: $Q(B) \sim e^{-a_B}$, where the buffer size $B$ is assumed to be large enough so that asymptotic conditions prevail. In view of offering the quality of service $\rho$, we then define the effective bandwidth for source $i$ to be $c_i = g_i(-\ln(\rho)/B)$. Recall that $\sum_{i \leq N} g_i(\eta) = c$ by definition of $\eta$. Expanding the function $g_i(u)$ about 0 yields $g_i(u) \approx -h_i \nu_i + [h_i \nu_i (1 - \nu_i)]^2 D_i \{a \epsilon_i^2 + c \epsilon_i^2 \} u/2$, showing it is strictly increasing.

Now, if $\sum_{i \leq N} e_i < c$, then the monotonicity of the function $g_i(.)$ implies $-\ln(\rho)/B < \eta$, hence that $Q(B) < p$: the required quality of service is met. But, if on the other hand, $\sum_{i \leq N} e_i > c$, then the probability of buffer overflow is larger than the required quality of service.

CONCLUSION

We have obtained a general analytic formula for expressing the Beneš bound as a contour integral in the complex plane. We easily derived closed formulas for the superposition of heterogeneous sources. These formulas are even more simple and explicit when superposing homogeneous sources. An immediate consequence of the analytic formula for the asymptotic queue length distribution tail was the easy definition of an effective bandwidth for each source.
REFERENCES


11. Elwalid A.I., Mitra D. “Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks”, IEEE Infocom ’93, March 1993