In *15th International Teletraffic Congress Proceedings*, June 1997

Measurement-based connection admission control

R. J. Gibbens[a*] and F. P. Kelly[a]

[a]Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB

The performance of measurement-based admission control depends upon statistical interactions between several time-scales, ranging from the very short time scales associated with cell or packet queueing, through burst time-scales, to the time-scales associated with admission decisions and the holding times of connections. In this paper we continue the development of a modelling approach which attempts to integrate these several time-scales, and illustrate its application to the analysis of a family of simple and robust measurement-based admission controls. A subsidiary aim of the paper is to shed light on the relationship between the admission control proposed for ATM networks by Gibbens *et al* [9] and that proposed for controlled-load Internet services by Floyd [7]. We shall see that their common origin in Chernoff bounds allows the definition of a simple and general family of admission controls, capable of tailoring for several implementation scenarios.

1. Introduction

There is by now a fairly good understanding of the behaviour of a queue whose arrival process is the superposition of many independent stationary streams. Results, generally based on large deviations theory, have shed considerable light on how a resource in a multiservice broadband network can statistically multiplex a given collection of well-characterized sources. There arise, however, several difficulties in the extension of this work to understand the behaviour of measurement-based admission control for multiservice networks where sources may not be well characterized. One major difficulty is that buffer overflow in such a network is generally a consequence of the combined effects of both extreme measurement errors that allow too many sources admission *and* the subsequent extreme behaviour of admitted sources. The first effect is naturally analysed on the time scales associated with admission decisions and the holding times of connections, while any analysis of the second effect requires an understanding of the statistical characteristics of sources over time scales comparable to the typical busy period preceding a buffer overflow.

In [9] an attempt was made to study this interaction between time scales with a simple Markov chain model. In this paper we continue the development of this modelling approach, and illustrate its application to the analysis of a simple and general family of admission controls.

Some of the theoretical background to the admission control of this paper is described in the Appendices; this background has motivated our choice of the acceptance regions for our admission controls, but stops well short of providing a full understanding of the interaction

*This author wishes to acknowledge the support of the Royal Society through a University Research Fellowship and both authors the EPSRC (Grant GR/J31896) for computing equipment.

between time scales, and there remains much work to be done. Jamin and Shenker [11] have conducted a simulation-based comparison of several measurement-based admission control algorithms, including [7] and [9]. We view the approach of this paper to be complementary to the simulation approach: each has its strengths and weaknesses, and a fuller understanding of measurement-based admission control will require contributions from theory, modelling, simulation and experiment. For several other approaches to admission control, see [5,15,18].

2. The basic model

The description of our basic model is self-contained, but for further background see [9]. Let

$$S(t) = \sum_{j=1}^{J} S_j(t), \quad S_j(t) = \sum_{i=1}^{n_j(t)} X_{ji}(t) \tag{1}$$

where $X_{ji}(t)$, for distinct values of i, j and t, are independent random variables with

$$\mathbb{P}\{X_{ji}(t) = h_j\} = \frac{m_j}{h_j}, \quad \mathbb{P}\{X_{ji}(t) = 0\} = 1 - \frac{m_j}{h_j}. \tag{2}$$

We interpret $X_{ji}(t)$ as the load produced by a connection of class j at time t. There are n_j connections of class j each with peak rate h_j and mean rate m_j. The rate of load lost at a resource of capacity C is then $M(\mathbf{n}) = \mathbb{E}(S - C)^+$ where $\mathbf{n} = (n_1, n_2, \ldots, n_J)$. Let connections of class j arrive in a Poisson stream of rate ν_j, let the holding times of accepted connections be independent and exponentially distributed with parameter μ_j. Let $\mathbf{S}(t) = (S_1(t), S_2(t), \ldots, S_J(t))$ and let $A(\mathbf{n})$ be a subset of $\mathbb{N}^J$. Suppose that a connection arriving at time t is accepted if $\mathbf{S}(t) \in A(\mathbf{n})$ and is rejected otherwise; suppose also that if a connection is rejected no other arriving connection is considered for acceptance until *after* a connection currently in progress has ended. Call the period between the rejection of a connection and the time when the first connection then in progress ends the *backoff period* [1,2,9]. Let $d(t) = 1$ or 0 according as at time t the system is in a backoff period or not. Then $(\mathbf{n}(t), d(t))$ is a Markov chain, with the following off-diagonal transition rates

$$(\mathbf{n}, 0) \longrightarrow \begin{cases} (\mathbf{n} + \mathbf{e}_j, 0) & \text{at rate } \nu_j a(\mathbf{n}) \\ (\mathbf{n} - \mathbf{e}_j, 0) & \text{at rate } \mu_j n_j \\ (\mathbf{n}, 1) & \text{at rate } \sum_j \nu_j (1 - a(\mathbf{n})) \end{cases} \tag{3}$$

$$(\mathbf{n}, 1) \longrightarrow (\mathbf{n} - \mathbf{e}_j, 0) \qquad \text{at rate } \mu_j n_j\,. \tag{4}$$

Here $\mathbf{e}_j$ is a vector with a 1 in the jth component and zeros otherwise and $a(\mathbf{n})$ is the acceptance probability $a(\mathbf{n}) = \mathbb{P}\{\mathbf{S}(t) \in A(\mathbf{n})\}$.

The proportion of load lost is

$$L = \frac{\mathbb{E}\, M(\mathbf{n})}{\mathbb{E} \sum_{j=1}^{J} n_j m_j}\,, \tag{5}$$

where the expectation is taken over the state $\mathbf{n}$ of the Markov chain.

The time parameter t appearing in the above model describes the time-scale associated with admission decisions and the holding times of connections. However the measurements of load, that provide the random variables $S_j(t)$, are taken over a time period τ that is typically very much shorter and comparable with the length of the typical busy period preceding an overflow from a cell or packet buffer. The step between the queueing model, describing important correlation effects on the short time scale, and the above bufferless model, describing admission decisions on the longer time scale, is outlined in Appendix B.

Note that we do *not* allow the admission decision to depend upon the peak rate of the current connection request, since this will produce an implicit bias towards connections with low peak rates: for the above scheme blocking probabilities are constant across connection classes. In Section 5 we extend the admission mechanism to give explicit priority to certain connections when the resource is near capacity.

3. Admission control schemes

How should the region $A(\mathbf{n})$ be chosen? In this paper we investigate some choices, each motivated by a particular choice of Chernoff bound. The choices correspond to different tangents to the effective bandwidth function, and are described in more detail in Appendix A.

Tangent at peak: let

$$A_I(\mathbf{n}) = \left\{ \mathbf{S} \,:\, \sum_j \left(h_j \left(1 - e^{-sh_j}\right) n_j + e^{-sh_j} S_j \right) \leq C \right\} . \tag{6}$$

Tangent at arbitrary location: let

$$A_{II}(\mathbf{n}) = \left\{ \mathbf{S} \,:\, \sum_j \left\{ \frac{h_j e^{sh_j}}{\left(h_j + \overline{m}_j \left(e^{sh_j} - 1\right)\right)^2} \right\} \left(\overline{m}_j^2 \left(e^{sh_j} - 1\right) n_j + h_j S_j \right) \leq C \right\} \tag{7}$$

where $\overline{m}_j \in (0, h_j)$ are constants, which might be interpreted as predicted mean rates.

Tangent of slope one: Our next choice of acceptance region $A(\mathbf{n})$ is that suggested by the Hoeffding bound, a bound whose use for admission control has been carefully discussed by Floyd [7]. Let

$$A_{III}(\mathbf{n}) = \left\{ \mathbf{S} \,:\, \sum_j S_j + \frac{s}{4} \sum_j h_j^2 n_j \leq C \right\} . \tag{8}$$

Tangent at origin: let

$$A_{IV}(\mathbf{n}) = \left\{ \mathbf{S} \,:\, \sum_j e^{sh_j} S_j \leq C \right\} . \tag{9}$$

Note that the various acceptance regions may have different implications for implementations. For example, the acceptance region $A_{III}(\mathbf{n})$ does not depend upon the entire vector $\mathbf{S}$, but only upon the aggregate measurement $\sum_j S_j$. In contrast, the acceptance region $A_{IV}(\mathbf{n})$ depends upon the vector $\mathbf{S}$, but not upon the vector $\mathbf{n}$: this is referred to as the *load only* case in [9].

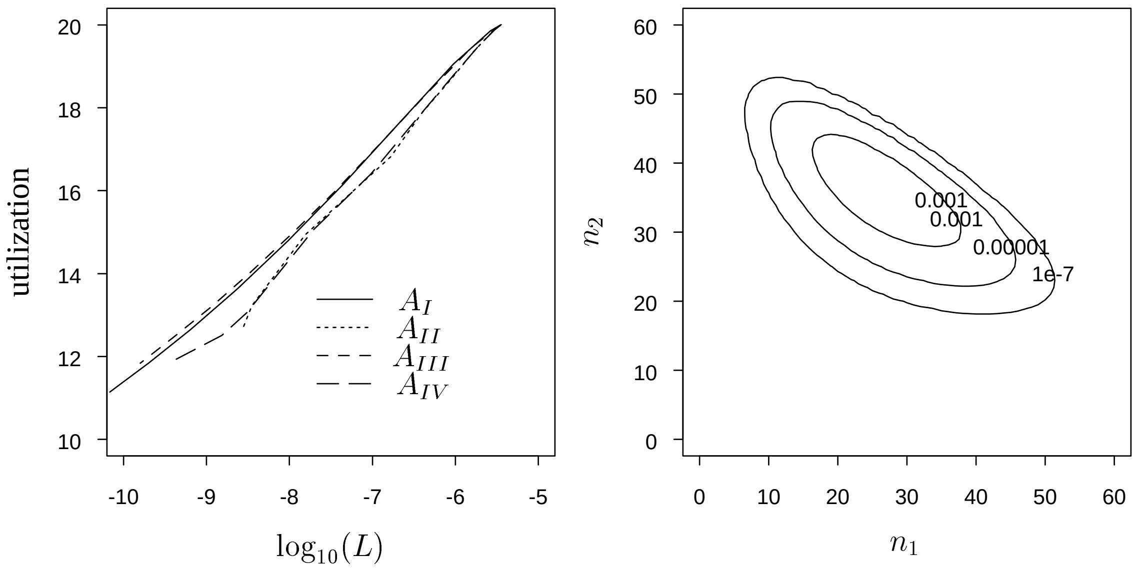

Figure 1. The left panel shows how the utilization decreases and the cell loss ratio improves, as the control parameter s increases. The right hand panel concerns the third scheme for a fixed value $s = 0.2$; note that the posterior distribution tails off rapidly as both n_1 and n_2 increase.

4. Numerical investigations

4.1. Trading off cell loss ratio against utilization

Any connection admission control must address the trade-off between cell loss and utilization (or, equivalently, connection blocking). In our schemes the parameter s controls this trade-off, which is illustrated in the left panel of Figure 1, under the following traffic conditions $(\lambda_1, h_1, m_1, \mu_1) = (15, 1, 0.5, 1)$, $(\lambda_2, h_2, m_2, \mu_2) = (5, 4, 0.1, 1)$, where $\lambda_j = \nu_j m_j$ is the *offered load* for a connection of type j and $C = 50$. Note that the first and third schemes behave similarly, as do the second and fourth schemes; no substantial difference is observed between the schemes.

The right panel of Figure 1 shows contours of the posterior distribution for the vector $\mathbf{n}$, given that a cell loss has just occurred: $\pi(\mathbf{n}|\text{cell loss}) = \pi(\mathbf{n})M(\mathbf{n})/\sum_{\mathbf{n}'} \pi(\mathbf{n}')M(\mathbf{n}')$, where $\pi(\mathbf{n})$ is the stationary distribution of the Markov chain $\mathbf{n}$. These results used the third scheme with the value $s = 0.2$.

4.2. Sensitivity to traffic mix

Next we consider the sensitivity of the third scheme to traffic mix. (In [9, Table 1, p 1112] a similar investigation was considered for the fourth scheme.) Suppose that the capacity $C = 50$ and there are three ($J = 3$) traffic types with parameters given by $(h_1, m_1, \mu_1) = (1, 0.5, 1)$, $(h_2, m_2, \mu_2) = (2, 0.5, 1)$ and $(h_3, m_3, \mu_3) = (4, 0.5, 1)$.

Table 1 show the results for the cell loss ratio, utilization and connection blocking probability for the third scheme under a range of traffic patterns. In this investigation the control parameter s is fixed throughout at $s = 0.75$, a value which ensures a satisfactory cell loss ratio over the entire range of traffic mixes considered.

Table 1
Cell loss ratio, utilization and connection blocking for the third scheme. Observe that the cell loss ratio remains well controlled over a wide range of traffic mix.

λ_1	λ_2	λ_3	Cell loss ratio, $\log_{10}(L)$	Utilization	Blocking
25.000	0.000	0.000	-11.01	23.66	0.054
12.500	12.500	0.000	-8.14	21.41	0.144
0.000	25.000	0.000	-12.38	13.47	0.461
0.000	12.500	12.500	-9.93	10.13	0.595
0.000	0.000	25.000	-10.23	7.18	0.713
12.500	0.000	12.500	-9.78	11.37	0.545
6.250	9.375	9.375	-9.57	12.08	0.517
6.250	6.250	12.500	-9.83	10.72	0.571
9.375	6.250	9.375	-9.53	12.48	0.501
12.500	6.250	6.250	-9.14	14.91	0.404
9.375	9.375	6.250	-9.20	14.35	0.426
6.250	12.500	6.250	-9.25	13.82	0.447
8.333	8.333	8.333	-9.44	12.90	0.484

5. Priorities

In this section we consider a model of two traffic classes with time-varying offered loads. The backoff mechanism discussed earlier is amended so that explicit priorities can be assigned to connections offered to the system. This is achieved by extending the state space with an indicator taking values of 0, 1 or 2. The precise description of the off-diagonal transition rates for the Markov chain are as follows

$$(\mathbf{n},0) \longrightarrow \begin{cases} (\mathbf{n}+\mathbf{e}_j,0) & \text{at rate } \nu_j a(\mathbf{n}) \\ (\mathbf{n}-\mathbf{e}_j,0) & \text{at rate } \mu_j n_j \\ (\mathbf{n},2) & \text{at rate } \sum_j \nu_j(1-a(\mathbf{n})) \end{cases} \tag{10}$$

$$(\mathbf{n},1) \longrightarrow \begin{cases} (\mathbf{n}-\mathbf{e}_j,1) & \text{at rate } \mu_j n_j(1-\alpha) \\ (\mathbf{n}-\mathbf{e}_j,0) & \text{at rate } \mu_j n_j \alpha \\ (\mathbf{n}+\mathbf{e}_j,1) & \text{at rate } \nu_j a(\mathbf{n}) p \\ (\mathbf{n},2) & \text{at rate } \sum_j \nu_j(1-a(\mathbf{n})) \end{cases} \tag{11}$$

$$(\mathbf{n},2) \longrightarrow (\mathbf{n}-\mathbf{e}_j,1) \qquad \text{at rate } \mu_j n_j\,. \tag{12}$$

In this model a connection of any traffic class is a high priority connection with probability p or a low priority connection with probability $1-p$, independent of the state of the Markov chain. If the indicator is in state 0 then both high and low priority connections are eligible for admission. If the indicator is in state 1 then only high priority connections are eligible for admission. If a connection is rejected by the scheme while the indicator is in states 0 or 1 (irrespective of the connection's priority) the indicator switches to state 2. In state 2, no connections are eligible for admission and the indicator remains in that state until the first connection clears down when it switches to state 1. If a connection clears down while the indicator is in state 1 then the indicator switches to state 0 with probability α and both high and low priority connections become eligible for admission. The parameter α is a simple device to manipulate the degree of priority given to the high priority traffic.

Figure 2 illustrates the operation of this modified form of backoff with the third scheme with the parameter $\alpha = 0.1$. The capacity is $C = 50$, the proportion of high priority connections is $p = 0.5$ and the choice $s = 0.3$ was selected. The parameters considered were $(h_1, m_1, \mu_1) = (1, 0.5, 1)$, and $(h_2, m_2, \mu_2) = (4, 0.1, 1)$. Thus there are essentially four types of connection, since priority can be high or low for each possibility of (h_j, m_j, μ_j).

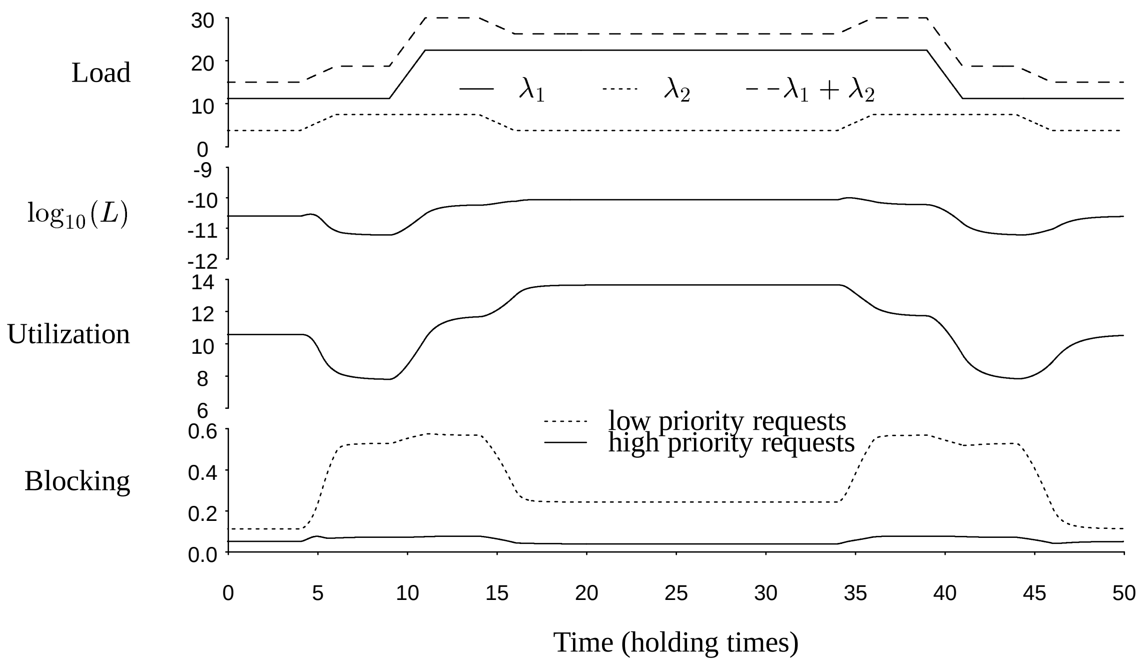

Figure 2. The figure shows the variation over time of cell loss ratio, utilization and connection blocking. Note that although the cell loss ratio is relatively constant, the utilization may vary substantially as the traffic mix, and hence the potential multiplexing gain, varies over time.

A. Chernoff bounds

Let $S = \sum_{k=1}^{K} X_k$ where $X_1, X_2, \ldots, X_K$ are independent non-negative random variables. Then $\mathbb{P}\{S \geq C\} \leq \mathbb{E}[e^{s(S-C)}]$, where here and throughout $s \geq 0$. Hence

$$\log \mathbb{P}\{S \geq C\} \leq s \left(\sum_k \alpha_k(s) - C \right) \tag{13}$$

where $\alpha_k(s) = s^{-1} \log \mathbb{E}\left[e^{sX_k}\right]$. Specialize now to the case where $\mathbb{P}\{X_k = h_k\} = m_k/h_k$ and $\mathbb{P}\{X_k = 0\} = 1 - m_k/h_k$. Thus m_k, h_k are respectively the mean and peak of an on-off source, and

$$\alpha_k(s) = \frac{1}{s} \log\left[1 + \frac{m_k}{h_k}(e^{sh_k} - 1)\right] . \tag{14}$$

A.1. Tangent at the peak ($A_I(\mathbf{n})$)

Regard $\alpha_k(s)$ as a function of m_k; it is a concave function of m_k, and hence bounded above by its tangent at the point $m_k = h_k$. Thus

$$\alpha_k(s) \le h_k - (h_k - m_k)\frac{e^{sh_k} - 1}{sh_k e^{sh_k}} . \tag{15}$$

Hence $\log \mathbb{P}\{S \ge C\} \le -\gamma$ is assured if there exists an s such that

$$\sum_k \left[h_k - (h_k - m_k)\frac{e^{sh_k} - 1}{sh_k e^{sh_k}}\right] + \frac{\gamma}{s} \le C . \tag{16}$$

The value of s minimizing the left-hand side of inequality (16) satisfies

$$\sum_k (h_k - m_k)\frac{e^{sh_k} - 1 - sh_k}{h_k e^{sh_k}} = \gamma , \tag{17}$$

and, with this value of s, inequality (16) becomes

$$\sum_k h_k - \sum_k e^{-sh_k}\,(h_k - m_k) \le C . \tag{18}$$

A.2. Tangent at arbitrary location ($A_{II}(\mathbf{n})$)

Regard $\alpha_k(s)$ as a function of m_k: it is a concave function of m_k, and hence bounded above by its tangent $a_k(s, \overline{m}_k) + b_k(s, \overline{m}_k)m_k$ at an arbitrary point $\overline{m}_k \in (0, h_k)$. Thus

$$\alpha_k(s) \le a_k(s, \overline{m}_k) + b_k(s, \overline{m}_k)m_k , \tag{19}$$

and so $\log \mathbb{P}\{S \ge C\} \le -\gamma$ is assured if there exists an s such that

$$\sum_k \left[a_k(s, \overline{m}_k) + b_k(s, \overline{m}_k)m_k\right] + \frac{\gamma}{s} \le C . \tag{20}$$

Using the minimizing choice of s this inequality becomes

$$\sum_k \left\{\frac{h_k e^{sh_k}}{\left(h_k + \overline{m}_k\left(e^{sh_k} - 1\right)\right)^2}\right\}\left(\overline{m}_k^2\left(e^{sh_k} - 1\right) + h_k m_k\right) \le C . \tag{21}$$

Information on good choices for $\overline{m}_k$, $k = 1, \ldots, K$ may come from a variety of sources, for example users may provide information through their tariff choices [12,17], or the network may have available long-term averages for traffic of different types.

A.3. Tangent of slope one ($A_{III}(\mathbf{n})$)

Suppose that $\sum_k m_k$ is known, but not individual values of m_k. Can we bound the right-hand side of inequality (13)? One method leads to the Hoeffding bound [10], which we obtain as follows.

Again regard $\alpha_k(s)$ as a function of m_k; it is a concave function of m_k, with a tangent of unit slope at the point $\overline{m}_k = s^{-1} - h_k(e^{sh_k} - 1)^{-1}$. Thus

$$\alpha_k(s) \leq m_k - \frac{1}{s} + \frac{h_k}{e^{sh_k} - 1} + \frac{1}{s} \log \frac{e^{sh_k} - 1}{sh_k} \leq m_k + \frac{sh_k^2}{8}\,; \tag{22}$$

see [10, pp. 106, 110] for the second inequality. Hence from inequality (13), $\log \mathbb{P}\{S \geq C\} \leq -\gamma$ is assured if there exists an s such that

$$\sum_k \left(m_k + \frac{sh_k^2}{8} \right) + \frac{\gamma}{s} \leq C\,. \tag{23}$$

The value of s minimizing the left-hand side of inequality (23) is $s = 2\left(\frac{2\gamma}{\sum_k h_k^2}\right)^{1/2}$, and with this choice of s inequality (23) can be written as either

$$\sum_k m_k + \frac{s}{4} \sum_k h_k^2 \leq C \quad \text{or} \quad \sum_k m_k + \left(\frac{\gamma}{2} \sum_k h_k^2 \right)^{1/2} \leq C\,; \tag{24}$$

the second inequality is familiar as the Hoeffding bound. The schemes considered in Table 1 and Figure 2 keep s, rather than γ, fixed as the traffic mix alters.

A.4. Tangent at the origin ($A_{IV}(\mathbf{n})$)

Regard $\alpha_k(s)$ as a function of m_k: it is a concave function, and bounded above by its tangent at the origin, $m_k = 0$. Thus

$$\alpha_k(s) \leq \frac{e^{sh_k} - 1}{sh_k}\, m_k \tag{25}$$

and so $\log \mathbb{P}\{S \geq C\} \leq -\gamma$ is assured if there exists an s such that

$$\sum_k \frac{e^{sh_k} - 1}{sh_k}\, m_k + \frac{\gamma}{s} \leq C\,. \tag{26}$$

Take s to be the value minimizing the left-hand side of inequality (26). Then a simple differentiation establishes that

$$\sum_k \frac{m_k}{h_k} \left[e^{sh_k}(sh_k - 1) + 1 \right] = \gamma\,. \tag{27}$$

With this value of s inequality (26) becomes

$$\sum_k e^{sh_k} m_k \leq C\,. \tag{28}$$

B. Large deviation limits

Suppose that the amount of work arriving at a queue over the period $[0, \tau]$ is

$$X[0,\tau] = \sum_{j=1}^{J} \sum_{i=1}^{n_j} X_{ji}[0,\tau] \tag{29}$$

where $(X_{ji}[0,\tau])_{ji}$ are independent processes with stationary increments whose distributions may depend upon j but not upon i. Let $L(C,b,n)$ be the proportion of workload lost, through overflow of a buffer of size $b > 0$, when the server has rate C and $n = (n_1, n_2, \ldots, n_J)$. Then [3,6,16]

$$\lim_{N\to\infty} \frac{1}{N} \log L(cN, bN, nN) = \sup_{\tau} \inf_{s} \left[s\tau \sum_j n_j \alpha_j(s,\tau) - s(b + c\tau) \right] \tag{30}$$

where

$$\alpha_j(s,\tau) = \frac{1}{s\tau} \log \mathbb{E}\left[e^{sX_{ji}[0,\tau]} \right] \tag{31}$$

is the *effective bandwidth*. A recent review of the effective bandwidth concept is given in [13]; in [8] the effective bandwidth function is obtained for traces of ethernet traffic and MPEG video sources across a broad range of time and space scales. This motivates the large deviations approximation for $\gamma = -\log L(c,b,n)$ of

$$\gamma \approx -\sup_{\tau} \inf_{s} \left[s\tau \sum_j n_j \alpha_j(s,\tau) - s(b + c\tau) \right] . \tag{32}$$

Henceforth let s, τ be the extremizing pair in relation (32). Then the approximation (32) aligns with the Chernoff bound (13) under the correspondence $X_j = X_{ji}[0,\tau]$, $C = b + c\tau$ and $\tau\alpha_j(s,\tau) = \alpha_j(s)$. The critical time scale τ has a straightforward interpretation under the large deviations limit as the time for which the server has been busy preceding a buffer overflow [6].

Next suppose that a source of type j is policed by leaky buckets labelled $k = 1, 2, \ldots, K$, so that $X_{ji}[0,T] \le \beta_{kj} + \rho_{kj} T$, for all $T > 0$, and for $k = 1, 2, \ldots, K$. Then with probability one

$$0 \le X_{ji}[0,\tau] \le h_j = \min_k \{\beta_{kj} + \rho_{kj}\tau\} , \tag{33}$$

and so, if $m_j = \mathbb{E}X_{ji}[0,\tau]$, then an upper bound for $\tau\alpha_j(s,\tau)$ is

$$\tau\alpha_j(s,\tau) \le \frac{1}{s} \log \left[1 + \frac{m_j}{h_j} \left(e^{sh_j} - 1 \right) \right] , \tag{34}$$

corresponding with equation (14) under the identification $\tau\alpha_j(s,\tau) = \alpha_j(s)$.

Acknowledgement

We are grateful to Sally Floyd, Sugih Jamin and Scott Shenker for sharing early drafts of their work and for their helpful insights.

REFERENCES

1. N. G. Bean. *Statistical Multiplexing in Broadband Communication Networks*. PhD thesis, University of Cambridge, 1993.
2. N. G. Bean. Robust connection acceptance control for ATM networks with incomplete source information. *Annals of Operations Research*, 1994.
3. D. D. Botvich and N. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 20:293–320, 1995.
4. C. Courcoubetis, F. P. Kelly, and R. R. Weber. Pricing resource usage in communication networks. In preparation, 1996.
5. C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand, and R. R. Weber. Admission control and routing in ATM networks using inferences from measured buffer occupancy. *IEEE Transactions on Communications*, 43:1778–1784, 1995.
6. C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a switch handling many traffic sources. *J. Appl. Probab.*, 33, 1996.
7. S. Floyd. Comments on measurement-based admission control for controlled-load services. Lawrence Berkeley National Laboratory, July 1996.
8. R. J. Gibbens. Traffic characterisation and effective bandwidths for broadband network traces. In F. P. Kelly, S. Zachary, and I. B. Ziedins, editors, *Stochastic Networks: Theory and Applications*, Royal Statistical Society Lecture Note Series 4, pages 169–179. Oxford University Press, 1996.
9. R. J. Gibbens, F. P. Kelly, and P. B. Key. A decision-theoretic approach to call admission control in ATM networks. *IEEE Journal on Selected Areas in Communications, special issue on Advances in the Fundamentals of Networking*, 13(6):1101–1114, 1995.
10. M. Hofri. *Probabilistic Analysis of Algorithms*. Springer-Verlag, New York, 1987.
11. S. Jamin, S. J. Shenker, and P. B. Danzig. Comparison of measurement-based admission control algorithms for controlled-load service. In *IEEE INFOCOM '97 Proceedings*, April 1997.
12. F. P. Kelly. Charging and accounting for bursty connections. In L. W. McKnight and J. P. Bailey, editors, *Internet Economics*. MIT Press, 1996.
13. F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. B. Ziedins, editors, *Stochastic Networks: Theory and Applications*, Royal Statistical Society Lecture Note Series 4, pages 141–168. Oxford University Press, 1996.
14. B. K. Ryu and A. Elwalid. The importance of long-range dependence of VBR traffic in ATM traffic engineering: myths and realities. Hughes Reseach Labs/Bell Labs, 1996.
15. H. Saito and K. Shiomoto. Dynamic call admission control in ATM networks. *IEEE Journal on Selected Areas in Communications*, 9:982–989, 1991.
16. A. Simonian and J. Guibert. Large deviations approximation for fluid sources fed by a large number of on/off sources. *IEEE J. Sel. Areas Commun.*, 13:1017–1027, 1995.
17. D. Songhurst and F. Kelly. Charging schemes for multiservice networks. In *15th Int. Teletraffic Congress Proceedings*. North Holland-Elsevier Science Publishers, 1997.
18. R. Warfield, S. Chan, A. Konheim, and A. Guillaume. Real-time traffic estimation in ATM networks. In *14th Int. Teletraffic Congress Proceedings*. North Holland-Elsevier Science Publishers, 1994.