A Hybrid Admission Control Scheme for Broadband ATM Traffic

Ernst Nordström

Department of Computer Systems, Uppsala University
Box 325, S–751 05 Uppsala, SWEDEN
Ernst.Nordstrom@docs.uu.se

Abstract

Link admission control (LAC) in broadband ATM networks is based on evaluation of expected traffic performance. The traditional LAC approach relies on approximate analytical performance models, and can lead to an over controlled network. This paper presents a hybrid LAC scheme which uses a multi layer perceptron (MLP) to refine the performance estimate of a traditional analytical approximation. An accurate but complex traffic performance model is used to derive MLP target data. The numerical results show that the MLP can improve the bandwidth efficiency.

1 Introduction

The Broadband Integrated Services Digital Network (B–ISDN) will be based on a transport mode called Asynchronous Transfer Mode (ATM). ATM is a packet– and connection–oriented switching and multiplexing technique, designed to support virtually all communication services expected in the future. The bandwidth flexibility, the capability to handle all services in an uniform way, and the possible use of statistical multiplexing are advantageous properties of ATM. However, advanced traffic control procedures are needed to maintain the performance of ATM connections. The underlying idea in broadband traffic control is to prevent congestion rather than to react to it. The backbone in ATM traffic control will therefore consist of connection admission control (CAC), traffic enforcement and traffic shaping.

This paper deals with a subroutine of CAC called link admission control (LAC), which is performed at each node along the network path connecting the source and destination node. The purpose of LAC is to establish whether the link can accept a new connection request or if it should be rejected. In particular, we focuse on performance evaluation of a single ATM multiplexer (output buffer in an ATM switch) since this forms the basis for LAC. A trade off between performance evaluation accuracy and computational complexity is necessary to achieve acceptable connection set up delays. The traditional approach is based on analytical approximations, which may result in an over controlled network thereby under utilizing the network resources, see e.g. [3,5,9,11]. This contribution presents a hybrid approach, which use neural networks to improve the link bandwidth efficiency. The suggested scheme comprises a supervised multi layer perceptron (MLP) [4], combined with a tractable analytical approximation. To this end, we employ a variant of the approximate fluid flow performance formula proposed in [8]. This simplified analytical approach will always underestimate link performance. Hence, the MLP's role is to refine the evaluation result, by applying complementary knowledge of the actual performance. Here, the accurate heterogeneous fluid flow model [6,7] and approximations of it, are used to derive additional performance data.

2 Traffic Burst Model

The traffic carried by an individual link are assumed to originate from independent and periodic on/off sources. An periodic on/off source alternates deterministically between burst (on) periods and silent (off) periods with constant characteristic durations. During the burst period, the source transmits packets (called "cells" in ATM terminology), at a constant characteristic cell rate. According to [8], it is widely believed, although not formally proven, that the periodic on/off source represents the "worst case" output of the traffic enforcement function (e.g. implemented by two running windows or two Leaky buckets). Hence, if the connections are represented by periodic on/off sources, the performance analysis will be conservative. The periodic on/off source is characterized by three parameters [8,9]:

- p: the *peak cell rate*, defined as the reciprocal of the constant inter–cell time within a burst,
- m: the *mean cell rate*, defined as the as fraction of time the source is active, multiplied with the peak cell rate, and
- t_{on}: the *burst period duration*

It is very difficult to evaluate the link performance in the deterministic case. However, it is known that more variable burst– and silent–period durations lead to larger queueing build up [8]. In particular, if exponentially distributed durations are used, as in the fluid flow cased below, the performance analysis will again be conservative.

3 Link Performance Evaluation

The total or superposed traffic stream offered to an ATM multiplexer is usually analyzed in two time scales, called the cell scale and the burst scale. The cell scale considers queue build–up (and buffer overflow) at the output buffer due to simultaneous cell arrivals. The burst scale considers queue build–up due to fluctuations in the total arrival *rate*, which may exceed the link capacity during some periods. Throughout this paper we only consider burst scale performance in terms of the cell loss probability, and assume that the buffer is dimensioned to resolve the cell scale conflicts.

3.1 Fluid Flow Cell Loss Probability

The heterogeneous fluid flow queueing model [6,7] is a very accurate tool for evaluating burst scale performance. This model analyses an ATM multiplexer loaded with connections of c different classes, each comprising N_j independent and homogeneous on/off sources (i.e. with equal traffic parameters p_j, m_j and $t_{on(j)}$). The multiplexer is characterized by a buffer of B cells, and an output capacity of C cells/second. The fluid approach models the traffic stream as a continuos (fluid) rate stream. A c–dimensional Markov chain emulates the arrival rate process. The Markov properties implies exponential active– and passive–period durations, which is of practical importance. The underlying Markov state space $\boldsymbol{S}$ is indexed by the possible active–source combinations, i.e. $\boldsymbol{S} = \{ \boldsymbol{k}=(k_1, ..., k_c) : 0 \leq k_j \leq N_j \}$. The state space comprises a total number of $N_s=(N_1+1)(N_2+1)...(N_c+1)$ states. Hence, the size of the state space increases rapidly (geometrically) with the class sizes N_j.

Two time–independent probability distribution functions are of prime importance in the fluid analysis:

$\pi = \{\pi_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \boldsymbol{S}}$	: the stationary arrival rate distribution,
$\boldsymbol{F}(x)= \{F_{\boldsymbol{k}}(x)\}_{\boldsymbol{k} \in \boldsymbol{S},\ 0 \leq x \leq B}$	: the bivariate buffer occupancy distribution.

$\pi_{\boldsymbol{k}}$ is the overall probability that the sources are in state $\boldsymbol{k}$, and is given by the multi–binomial distribution. $F_{\boldsymbol{k}}(x)$ is the probability that the sources are in state $\boldsymbol{k}$, and that the buffer content is less or equal to x. The $F_{\boldsymbol{k}}$'s are determined by solving a set of first order differential equations associated with the Markov states. The standard spectral expansion (eigenvalue–eigenvector) solution yields [6]:

$$\boldsymbol{F}(x) = \sum_{\boldsymbol{n} \in \boldsymbol{S}} a_{\boldsymbol{n}} \exp(z_{\mathbf{n}} x)\, \phi_{\mathbf{n}} , \qquad (1)$$

where $a_{\boldsymbol{n}}$ are coefficients found from boundary conditions, $z_{\boldsymbol{n}}$ are eigenvalues and $\phi_{\boldsymbol{n}}$ are the corresponding (N_s dimensional) eigenvectors. The coefficients $a_{\boldsymbol{n}}$ constitute the main computational burden and dictates the overall complexity of the fluid flow model. The coefficients are calculated by solving a system of N_s coupled linear equations which requires $O(N_s{}^3)$ numerical operations. However, a simplified computational approach for estimating the coefficients can be found in [1]. The approximation reduces the overall complexity to $O(N_s{}^2)$.

Once these probability distributions are found, the cell loss probability is easily calculated as the ratio between the loss cell rate and mean arrival cell rate:

$$P_{loss} = \frac{\sum_{\{ \boldsymbol{k} \mid \boldsymbol{k} \cdot \boldsymbol{p} > C \}} (\boldsymbol{k} \cdot \boldsymbol{p} - C)\left(\pi_{\boldsymbol{k}} - F_{\boldsymbol{k}}(B)\right)}{\sum_{\boldsymbol{k} \in \boldsymbol{S}} (\boldsymbol{k} \cdot \boldsymbol{p})\pi_{\boldsymbol{k}}} , \qquad (2)$$

where $\cdot$ denotes scalar vector multiplication, and $\boldsymbol{p} = (p_1, ..., p_c)$. The loss rate is obtained by adding the loss rates of all overload states. For the multi–binomial arrival rate distribution, the mean arrival rate simply equals $\boldsymbol{N} \cdot \boldsymbol{m}$, where $\boldsymbol{N} = (N_1, ..., N_c)$ and $\boldsymbol{m} = (m_1, ..., m_c)$.

3.2 Bounds of the Fluid Flow Cell Loss Probability

The fact that the computational complexity of fluid flow model rapidly increases with the size of the state space N_s, severely limits its practical use. In particular, heterogeneous traffic mixes are hardy tractable in real time. Hence, the upper bound of the fluid flow cell loss probability proposed in [8] is therefore of great importance. Although no formal proof is given that validates the approach, it is supported by our extensive numerical tests. The upper bound is approximated by the expression:

$$\overline{P}_{\mathrm{loss}} = \mathrm{P}_{\mathrm{loss}}(0) \exp(z_\infty \mathrm{B}) \tag{3}$$

The quantity z_∞ represents the asymptotic exponential decay rate as the buffer size approaches infinity, and is given by the largest negative eigenvalue in the spectral expansion (1). The quantity B denotes the actual buffer size. The magnitude $P_{loss}(0)$ is found by substituting the empty buffer boundary condition $(F_{\boldsymbol{k}}(0)=0$ for $\boldsymbol{k} \cdot \boldsymbol{p} > C)$ in expression (2). The resulting expression involves only the stationary arrival rate distribution:

$$P_{loss}(0) = \frac{\sum_{\{\boldsymbol{k} \mid \boldsymbol{k} \cdot \boldsymbol{p} > C\}} (\boldsymbol{k} \cdot \boldsymbol{p} - C)\, \pi_{\boldsymbol{k}}}{\sum_{\boldsymbol{k} \in \boldsymbol{S}} (\boldsymbol{k} \cdot \boldsymbol{p})\, \pi_{\boldsymbol{k}}} . \tag{4}$$

As for the computational burden, the decay rate z_∞ has only a complexity which increases linearly with the number of classes c [6]. However, the magnitude $P_{loss}(0)$ may require significant computational effort since the complexity behaves like $O[(c+N)N_s]$, where N_s equals the size of the state space and $N=N_1+....+N_c$ [9]. As already mentioned, N_s will increase rapidly when the number of classes, or the class sizes increase. Fortunately, there exist approximations of $P_{loss}(0)$ with tractable complexity, e.g. Chernoff bound [5,9] or large deviation [5,11]. Both these methods require $O(c)$ operations, and reduce the overall complexity of the final estimate $\hat{\overline{\mathrm{P}}}_{\mathrm{loss}}$ to the same order.

The hybrid LAC scheme presented in the next section will also make use of a lower bound of the fluid flow cell loss probability, based on the traditional asymptotic fluid flow formula in [7], which states that the spectral expansion (1) can be approximated by the component associated with the largest negative eigenvalue. The lower bound of the cell loss probability is approximated by the expression:

$$\underline{P}_{loss} = -a_\infty \exp(z_\infty B) \frac{\sum_{\{\boldsymbol{k} \mid \boldsymbol{k} \cdot \boldsymbol{p} > C\}} (\boldsymbol{k} \cdot \boldsymbol{p} - C)\, \phi_\infty(\boldsymbol{k})}{\sum_{\boldsymbol{k} \in \boldsymbol{S}} (\boldsymbol{k} \cdot \boldsymbol{p})\, \pi_{\boldsymbol{k}}} , \tag{5}$$

where a_∞ and ϕ_∞ are the coefficient respectively eigenvector associated with the dominant eigenvalue z_∞. The computational approach described in [1] is recommended for estimating the coefficient. In this case, the overall computational complexity of the lower (5) and upper bound (3) become equivalent, i.e. $O[(c+N)N_s]$ in the above notation. However, an efficient lower bound approximation $\underline{\hat{P}}_{\mathrm{loss}}$ of the Chernoff/Large deviation type is a subject for future work.

4 Hybrid Link Admission Control

The first objective of link admission control (LAC), is to maintain the link performance, and the second is to optimize the link bandwidth utilization. LAC based on the tractable upper bound yields link performance robustness, but not always link bandwidth efficiency. This section presents an hybrid LAC approach, which combines the tractable upper bound with nonlinear regression (function approximation), to improve this condition. The regression is performed by means of the standard multi layer perceptron (MLP) neural network [4], and concentrates on traffic load situations which are critical in terms of bandwidth utilization.

The hybrid approach is based on the observation that an *upper (lower)* bound of the true cell loss probability always yields correct LAC *accept (reject)* decisions. The inequalities $\left(\underline{\hat{P}}_{loss} <\right) \underline{P}_{loss} < \mathrm{P}_{\mathrm{loss}} < \overline{P}_{\mathrm{loss}} < \hat{\overline{\mathrm{P}}}_{\mathrm{loss}}$ is then simply applied in this context. Note that the tractable lower bound is parenthesized since no efficient approximation is known at present, and that the true cell loss probability is represented by the exact fluid flow cell loss probability. The hybrid cell loss probability estimate is defined by a single side inequality:

$$\hat{\mathrm{P}}_{\mathrm{loss}} = \begin{cases} \hat{\overline{\mathrm{P}}}_{\mathrm{loss}} & \text{when } \hat{\overline{\mathrm{P}}}_{\mathrm{loss}} < \epsilon \\ \hat{\mathrm{P}}_{\mathrm{loss}}(\boldsymbol{L},\boldsymbol{w}) & \text{when } \hat{\overline{\mathrm{P}}}_{\mathrm{loss}} > \epsilon, \end{cases} \tag{6}$$

where $\hat{\mathrm{P}}_{\mathrm{loss}}(\boldsymbol{L},\boldsymbol{w})$ denotes the MLP estimate for input link state vector $\boldsymbol{L}$ and MLP weight vector $\boldsymbol{w}$ (comprising all network weights), and the ε denotes the acceptable cell loss probability (e.g. $\varepsilon=10^{-9}$). According to definition (6), the MLP is only activated when the tractable upper bound states that the performance is unacceptable. The role of the MLP is simply to refine the upper bound estimate, by applying performance knowledge derived by a more accurate/complex analytical model.

The input link state vector $\boldsymbol{L}$ is composed of the tractable upper bound estimate, together with six statistics characterizing the superposed heterogeneous traffic stream [2]. The statistics represent the class parameters $\{N_j, p_j, m_j, t_{on(j)}\}_{j=1,\dots,c}$ in condensed, approximate form. The most computational demanding statistic is the dominant eigenvalue z_∞. See [2,10] for further details on the preprocessing.

The range of burst level traffic situations considered in the MLP training is derived from a model of offered traffic at the connection level [5]. In the present approach, the offered connection traffic is represented by a traffic intensity function ϱ defined for connection class tuples (N, p, m, t_{on}). In the context of connection level traffic simulation, the class parameters can be seen as stochastic variables with joint distribution function $\varrho(N, p, m, t_{on})$ if ϱ is properly normalized. A traffic situation is simply a set of class tuples drawn from this distribution. Note that the upper bound (3) and lower bound (5) should be consulted before the exact fluid flow model, since one of these bounds could have sufficient accuracy (w.r.t LAC).

5 Numerical Examples

In this section, numerical examples are used to evaluate the accuracy of the hybrid LAC scheme and to compare it to other methods tractable in real time. The Integrated Large Deviation (ILD) [11], Integrated Chernoff Bound (ICB) [9], Equivalent Capacity (EC) [3], and the tractable fluid flow upper bound (implemented by the ILD method) are evaluated separately.

The traffic situations are drawn from independent, uniform and discrete class parameters distributions, where $N \in \{1, 2, ..., 25\}$, $p \in \{2, 4, ..., 14\}$ Mb/s, $m \in \{1, 2, ..., 14\}$ Mb/s and $t_{on} \in \{1, 2, ..., 25\}$ msec. Each traffic situation is composed of three class tuple samples ($c=3$), and fulfill the criterion for statistical multiplexing ($\boldsymbol{N} \cdot \boldsymbol{m} < C < \boldsymbol{N} \cdot \boldsymbol{p}$). The multiplex system is characterized by B=100 cells and C=135 Mb/s, and the LAC by $\varepsilon=10^{-9}$. Given the above constraints, a set of 23512 traffic situations have been generated. According to the exact fluid flow model, 23.2% of these are accept– and 76.8% are reject–LAC situations. The following table shows the adequacy/applicability of the hybrid performance models.

Lower Bound	Exact Fluid Flow	Upper Bound	Tractable Upper Bound
70.8 %	10.4 %	0.001 %	18.7 %

As shown, the lower bound is sufficient (w.r.t. LAC) for many traffic situations. An MLP training set of 20000 samples and test set of 3512 samples are constructed from the generated data. The target performance values are logarithmically transformed to enable representation of tiny probabilities. A weighted MLP error function is used to enhance errors near the LAC decision boundary ε.

The LAC test set results are given as % bad accepts and % bad rejects, which can be interpreted as measures of performance robustness and bandwidth efficiency respectively. The hybrid result is a median over several backpropagation [4] training sessions. As shown in the following table, the hybrid approach yields fewest bad rejects, and therefore has potential for high bandwidth efficiency. Even though it makes some bad accepts, the true cell loss probability is always less than 10^{-8} in these cases.

	Hybrid	Tractable UB	ILD	ICB	EC
Bad Accepts	0.5%	0%	0%	0%	1.1%
Bad Rejects	1.0%	4.3%	13.0%	15.5%	9.1%

6 Summary

In this paper, we have presented and evaluated a hybrid scheme for link admission control (LAC) in broadband ATM networks. The suggested scheme comprises an approximate analytical performance formula, combined with a supervised multi–layer perceptron (MLP). The analytical formula is a computationally efficient variant of an upper bound of the fluid flow cell loss probability proposed in [8]. The MLP is trained using refined performance knowledge, derived (off–line) from different fluid flow performance approximations. An efficient lower bound of the cell loss probability is of interest in the presented scheme, but is part of future work.

The numerical results show that the lower bound of the fluid flow cell loss probability is adequate for many generated traffic situations. The results also show that the hybrid scheme has potential to improve the bandwidth efficiency.

Acknowledgements

The author would like to thank Mats Gustafsson, Olle Gällmo and Lars Asplund, Uppsala University, and Søren Blaabjerg, ELLEMTEL, for stimulating discussions. The financial support from ELLEMTEL, Sweden is also greatly acknowledged.

References

[1] A. Baiocchi, N. Bléfari–Melazzi, A. Roveri, F. Salvatore, ”Stochastic Fluid Analysis of an ATM Multiplexer Loaded with Heterogeneous ON–OFF Sources: an Effective Computational Approach”, *INFOCOM’92*, Session 3C.3, Florence, Italy, May 1992.

[2] S. Blaabjerg, ”On Two Approximations of the Heterogeneous On/ Off Fluid Queue”, *COST 224, Final Seminar*, Oct. 1991.

[3] R. Guérin, H. Ahmadi, M. Naghshineh, ”Equivalent Capacity and its Application to Bandwidth Allocation in High–Speed Networks”, *IEEE JSAC*, vol. 9, no. 7, pp. 968–981, Sep. 1991.

[4] J. Hertz, A. Krogh, R. Palmer, *Introduction to the theory of neural computation*, Addison–Wesley, 1991.

[5] J. Y. Hui, ”Resource Allocation for Broadband Networks”, *IEEE JSAC*, vol. 6, no. 9, pp. 1598–1608, Dec. 1988.

[6] S. Jacobsen, L. Dittman, ”A Fluid Flow Queueing Model for Heterogeneous On/Off Traffic”, RACE BLNT Workshop, Munich, July 1990.

[7] L. Kosten, Stochastic Theory of Multi Handling Systems with Groups of Multiple Sources, *Int. symp. on perf. of comp. comm. systems*, IFIP, Zürich, March 1984 (published by Elsevier).

[8] K. Kvols, S. Blaabjerg, ”Bounds and Approximations for the Periodic On/off Queue with Applications to ATM Traffic Control”, *INFOCOM’92*, Session A.3.1, Florence, Italy, May 1992.

[9] Y. Miyao, ”A Call Admission Control Scheme in ATM networks”, *ICC’91*, Session 14.5.1, Denver, Colorado, June 1991.

[10] E. Nordström, O. Gällmo, M. Gustafsson & L. Asplund, ”Statistical Preprocessing for Service Quality Estimation in a Broadband Network”, *WCNN’93*, Portland, Oregon, vol. 1, pp. 295–299, July 1993.

[11] J. Virtamo, ”A Note on Hui’s Admission Control Algorithm”, *COST 224*, TD(90) 038, 1990.