Required Capacity for ATM Virtual Paths for QoS on Cell and Connection Level

Markus Fiedler[1] and Åke Arvidsson[2]

[1]University of Karlskrona/Ronneby, Dept. of Telecommunications and Signal Processing (ITS), S-371 79 Karlskrona. E-mail: Markus.Fiedler@its.hk-r.se.

[2]Ericsson Utvecklings AB, Soft Center (Etapp III), S-372 25 Ronneby. E-mail: Ake.Arvidsson@uab.ericsson.se.

Abstract

We show approximation formulae for required capacities (bit rates) of ATM Virtual Paths that cover allocation on cell and connection level at the same time. Such formulae are valuable for dimensioning and routing purposes, as they allow for fast and precise calculations of required capacities and show the kind of dependency on the traffic demand for given traffic characteristics and Quality of Service requirements. We discuss how to determine the coefficients of the approximation and its precision for connections with on-/off characteristics (voice, data). Finally, we show the gain that might be obtained if statistical multiplexing of variable bit rate connections is taken into account on the cell level.

1 Introduction

In long-distance networks, we face the need for efficient bandwidth allocation due to the higher cost of bandwidth compared to local-area networks. Long-distance networks are increasingly based on the *Asynchronous Transfer Mode* (ATM) that offers an outstanding flexibility with regard to allocation of bandwidth to connections. Especially if connections with *variable bit rates* (VBR) are grouped, statistical multiplexing of their cell streams may lead to considerable savings in allocated bandwidth. We shall assume that such a group is created by assigning the respective connections to the same *Virtual Path* (VP); moreover, we shall confine ourselves to homogeneous connection characteristics within a VP.

For issues like how to dimension a network link that should support certain VPs, or how to route a VP through the network, it is of vital importance to know the amount of bandwidth (= bit rate) that has to be allocated to a specific VP so that the *Quality of Service* (QoS) demands of the connections within that VP are met. This quantity is called *required capacity* of the VP. Considerable attention has been paid to this required capacity, but mostly limited either to

- the *connection level* — specifying the number of connections that are required to keep the *blocking probability* for connections requests under a certain level, given a certain load [1], [2] *et al.*
- the *cell level* — specifying the bit rate that is required for a fixed number of connections to keep the *cell loss probability* or *probability of saturation* under a certain level [3], [4] *et al.* In this paper, the notion "cell level" shall also include the influence of the burst level, *i.e.* of the variability of the bit rate over time.

We will look at *capacity approximation formulae* that address *capacity allocation on both levels, i.e.* that take the (offered) load on connection level and QoS parameters, connection and system characteristics of both connection and cell level into account. As these formulae are tight in the sense that they avoid capacity underestimations (w.r.t. QoS) as well as huge overestimations (w.r.t. high network utilization), it gives insight into the underlying law that governs capacity allocation on different time scales.

A multi-level capacity allocation problem is studied by Hui [5], but no underlying law is discussed. Mitra *et al.* [6] present an Equivalent Bandwidth approach, but they take a kind of *mean cell loss probability* into account that leads to more optimistic (= lower) capacity allocation, but also to QoS degradation on cell level below the negotiated value if the number of existing connections keeps being close to its limit for a longer time.

The paper is organized as follows: Section 2 deals with the model under investigation and comments on QoS determination. Section 3 defines required capacities on different levels, and section 4 describes the basic approximation functions. Section 5 treats approximations of required capacities for CBR and VBR connections and shows the gain by taking the VBR property into account. Section 6 summarizes the paper and comments on open issues.

2 The Model

The model that we use to concentrate cell streams of different connections into a VP consists of a buffered multiplexer fed by homogeneous sources. The following subsections deal with aspects of the model and the calculation of QoS parameters that belong to the different time scales.

2.1 Connection Level Model

On connection level, we assume a *Poisson process* of connection requests with intensity λ, and i.i.d. holding times with mean $1/\mu$. Thus, the *offered load* becomes

$$A = \frac{\lambda}{\mu}, \tag{1}$$

and the *load* is given by

$$Y = A(1 - P_{\text{block}}). \tag{2}$$

If N' connections are supported on connection level, the *blocking probability* P_{block} is given by the well-known Erlang-B formula $B(N', A)$.

2.2 Cell Level Model

On cell level, we shall deal with a number of N'

- CBR connections as a model for voice traffic;
- VBR connections with *on-/off characteristics* and *exponentially distributed on-/off phase durations* as a model for silence-suppressed voice and packetized data traffic.

Both types are characterized by their *peak bit rate h*, the VBR connections additionally by their *mean bit rate m* and their *mean burst length* (= mean number of cells within an on phase) hb. The buffer size of the multiplexer is denoted by K.

The buffer shall be large enough so that loss only occurs if the system is temporarily overloaded; thus, the *fluid flow model* can be used. For this model, numerically stabilized algorithms to evaluate *loss probabilities* $P_{\text{loss}}(N')$ even for large systems with $N' = 400 \ldots 500$ connections are available [7]. The upper bound on N' depends mainly on calculation time and stability restrictions, the latter may vary with different values of h/m and K/hb.

3 Required Capacity

3.1 Connection Level

To meet a given blocking probability $P_{\text{block}} = 10^{-2}$, the *required number of connections* $N_{\text{r}}(A)$ has to be allocated (the index "r" stands for "reference"):

$$N_{\text{r}}(A) = \min\{N' \,|\, P_{\text{block}}(N', A) \leq 10^{-2}\}. \tag{3}$$

This function is a multi-step function with positive increments. As we are interested in a conservative approximation $N(A) \geq N_{\text{r}}(A)$, we might restrict our attention to the steps themselves: Let $A_{\min}(N')$ be the minimal offered load that leads to allocation of N' connections. Then, a function $N(A)$ that fulfills $N' \leq N(A_{\min}(N'))$ as tightly as possible is the approximation that we are looking for; this is illustrated by Figure 1. Finally, the number of connections to be allocated is determined by the integer $\lfloor N(A) \rfloor$.

For the numerical investigations, the search for $A_{\min}$ has been carried out up to a precision of 10^{-8}.

3.2 Cell Level

To meet a given loss probability on cell level of $P_{\text{loss}} \leq 10^{-9}$, if N' connections exist, the *required capacity* $C_{\text{r}}(N')$ has to be allocated:

$$C_{\text{r}}(N') = \min\{C \,|\, P_{\text{loss}}(C, N') \leq 10^{-9}\}. \tag{4}$$

Closed formulae for C_{r} are restricted to special assumptions, see [8] for details. In most cases, this capacity value has to be searched based on loss probability calculations. We stop the search as soon as $(1 - 10^{-4}) \cdot 10^{-9} \leq C_{\text{r}}(N') \leq 10^{-9}$ is fulfilled.

3.3 Both Levels

To ensure that the QoS on cell level is met, even if $N_{\text{r}}(A)$ connections exist, the *required capacity* for the joint capacity allocation problem is given by

$$\begin{aligned} C_{\text{r}}(A) &= C_{\text{r}}(N_{\text{r}}(A)) \\ &= \min\{C \,|\, P_{\text{block}}(N_{\text{r}}(A), A) \leq 10^{-2}, \\ &\qquad P_{\text{loss}}(C, N_{\text{r}}(A)) \leq 10^{-9}\}. \end{aligned} \tag{5}$$

For CBR connections, this simplifies to

$$C_{\text{r}}(A) = h N_{\text{r}}(A), \tag{6}$$

i.e. to the connection level problem. Observe that these reference values for the required capacity are valid only if not more than the required number of connections $N_{\text{r}}(A)$ are allocated on connection level.

4 Evaluation of Approximation Functions

The approximation function which is used for both required number of connections and required capacity has the structure

$$f(A) = k_0 + k_1 A + k_2 A^{k_3}. \tag{7}$$

The choice of the coefficients k_0 and k_1 will be explained in the following subsections, whereas k_2 and k_3 are determined by a Genetic Algorithm [9]; the number of significant decimals is limited to 2. As opposed to the Least

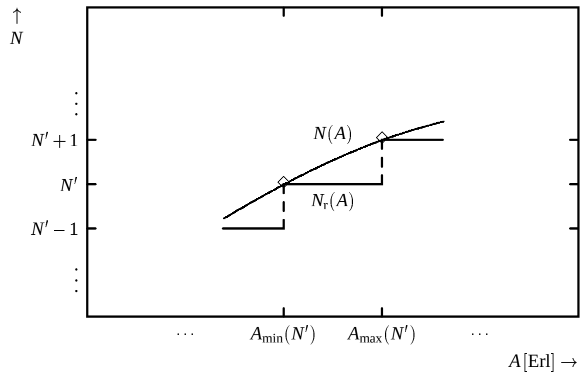

Figure 1: Required number of connections $N_r(A)$, critical points ($\diamond$) and an optimal approximation function $N(A)$.

Square Method, a Genetic Algorithm is able to avoid underestimations. Linear Programming cannot be used due to the power term in (7). More details might be found in [10].

4.1 Connection Level Approximation

Already for $A > 0.01$, 2 connections are required, and therefore, we fix $k_0 := 2$. In a former study, k_1 was approximated by a Genetic Algorithm with the result $k_1 \simeq 0.99 = Y/A$. For this reason, we will use the representation

$$N(Y) = 2 + Y + k_2 Y^{k_3}\,,\ Y = 0.99A\,. \tag{8}$$

The number of connections to be allocated to keep $P_{\text{block}} \leq 10^{-2}$ is given by $\lfloor N(Y) \rfloor$.

The relative error at the critical points is given by

$$e_{\text{conn}}(N') = \frac{N(Y_{\min}(N'))}{N'} - 1 \tag{9}$$

with $Y_{\min}(N') = 0.99 A_{\min}(N')$. The goal for the optimization of the coefficients of a capacity approximation function consists in a minimal mean relative error

$$\bar{e}_{\text{conn}} = \sum_{N'=2}^{N'_{max}} e_{\text{conn}}(N')\,. \tag{10}$$

A similar formula to (8) is used for dynamic determination of the required number of connections $N_d(\bar{Y})$, based on the *measured load* $\bar{Y}$ in a previous time interval [1]:

$$N_d(\bar{Y}) = \left\lceil \bar{Y} + 1.29\,\bar{Y}^{0.39} \right\rceil. \tag{11}$$

4.2 Approximation for Both Levels

For $A \to 0^+$, we already allocate 2 connections on connection level. The required capacity for each of those connections might be approximated by the equivalent bandwidth given in [3], [11]:

$$c = \frac{h}{2}\left(1 + \frac{\kappa h}{h-m}\right) - \frac{h\kappa}{2}\sqrt{\left(\frac{h}{h-m} + \frac{1}{\kappa}\right)^2 - 4\,\frac{m}{(h-m)\,\kappa}} \tag{12}$$

with

$$\kappa = \frac{1}{\ln(10^{-9})}\,\frac{K}{hb} \simeq -0.048255\,\frac{K}{hb} \tag{13}$$

for a desired $P_{\text{loss}} \leq 10^{-9}$. In parallel to the allocation problem on connection level, a suitable choice was found as

$$k_1 = mY\,, \tag{14}$$

so that also in the joint case, we formulate the approximation formula in dependence of Y rather than of A:

$$C(Y) = 2c + mY + k_2 Y^{k_3}\,,\ Y = 0.99A\,. \tag{15}$$

To construct such approximations, we again focus on the critical steps that occur at the same values $Y_{\min}(N')$ as described in subsection 4.1. Now, the *relative error* is given by

$$e_{\text{both}}(N') = \frac{C(Y_{\min}(N'))}{C_r(N')} - 1\,. \tag{16}$$

The definition of the mean relative error $\bar{e}_{\text{both}}$ is similar to (10).

5 Capacity Approximations

In this section, we will present some examples of capacity approximations that have been carried out up to $N'_{\max} = 500$, *i.e.* $Y \simeq 468$, if not stated otherwise.

5.1 CBR Connections

The approximation is given by

$$\frac{C(Y)}{h} = 2 + Y + 2.70\,Y^{0.40}\,,\ \bar{e}_{\text{conn}} \simeq 0.62\%\,. \tag{17}$$

The mean relative error is already quite small and might be lowered by taking more digits into account.

5.2 VBR On-/Off Connections

Table 1 shows results for a buffer size that corresponds to the mean burst size, and the results in Table 2 are based on a buffer that is ten times as big, which allows for additional statistical multiplexing gain.

For the small buffer, the curves of the approximation are shown in Figure 2. Comparing (17) with Table 1 shows that the quality of the approximation is mostly close to that for the CBR connections, with exception of the case $h/m = 8$ that is illustrated in Figure 3. For small load values, the approximation curve is quite close to the reference values, but the difference gets bigger as the load grows. In that case, a larger coefficient $k_0 > 2c$ would have led to a better approximation, but this is the price that has to be paid for a simple, universal formulation of the coefficients of the approximation.

$\frac{h}{m}$	$\frac{C(Y)}{h}$	$\bar{e}_{\text{both}}$
2	$1.9081 + 0.5Y + 3.48Y^{0.48}$	0.61 %
4	$1.9051 + 0.25Y + 3.21Y^{0.46}$	0.67 %
8	$1.9042 + 0.125Y + 2.93Y^{0.43}$	1.98 %

Table 1: Approximated required capacities for VBR connections, $P_{\text{block}} = 10^{-2}$, $P_{\text{loss}} = 10^{-9}$, buffer size $K = hb$.

$\frac{h}{m}$	$\frac{C(Y)}{h}$	$\bar{e}_{\text{both}}$
2	$1.4247 + 0.5Y + 1.73Y^{0.55}$	0.33 %
4	$1.2344 + 0.25Y + 1.56Y^{0.54}$	0.51 % [a]
8	$1.1391 + 0.125Y + 1.21Y^{0.53}$	0.49 % [b]

Table 2: Approximated required capacities for VBR connections, $P_{\text{block}} = 10^{-2}$, $P_{\text{loss}} = 10^{-9}$, buffer size $K = 10hb$ ($N'_{\max} = 436^{\text{a}}/452^{\text{b}}$).

Table 2 shows that a better approximation quality is obtained for the larger buffer. Figure 4 shows the corresponding approximations. The nonlinearity is less than that of the curves belonging to the smaller buffer; as the exponents k_3 do not differ substantially from each other, this may also be seen from $k_2(m, K = 10hb) < k_2(m, K = hb)$.

Finally, we shall look at the gain that comes along with taking the VBR property into account,

$$G(Y) = \frac{hN(Y)}{C(Y)} - 1. \quad (18)$$

Figure 5 shows that quite remarkable gain might be achieved. The gain, which is upper-bounded by $h/m - 1$, increases, if any of the following values

- load
- ratio of peak and mean bit rate
- ratio of buffer and mean burst size

becomes larger. If the latter ratio is quite high, a considerable gain is already obtained for small loads.

Most of the gain stems from the factor m in the proportional term of (15), compared to the factor h in (17). As we use the same, simple structure of the approximation formula for capacity allocation on connection level and both levels, the effort to allocate that gain is restricted to the determination of two coefficients.

6 Summary and Open Issues

We presented approximation formulae for the required number of connections to meet a given blocking probability as well as for the required capacity of a VP to meet given blocking and cell loss probabilities, both as functions of the traffic load. The formulae, which might be used for dimensioning and routing purposes, are always of the same structure: They consist of a constant, a linear and a power term. The first two coefficients may be fixed by general considerations, while the two coefficients belonging to the power term are found by using an optimization technique, here a Genetic Algorithm. In most of the cases, mean relative errors less than 1 % have been obtained. Furthermore, the structure of the formulae gives valuable insight into the fundamental relationship between load on connection level and capacity requirement on cell level.

Upto now, the two coefficient values that depend on the parameterization of connections, the size of the buffer and the QoS parameters are given numerically. One open issue consists in finding suitable interpolation formulae for them that reflect the parameterization and QoS demands on connection and cell level. Other open issues address connections that do not have exponentially distributed on-/off-phase durations or more than two states of activity, and connections with non-homogeneous parameters within a VP.

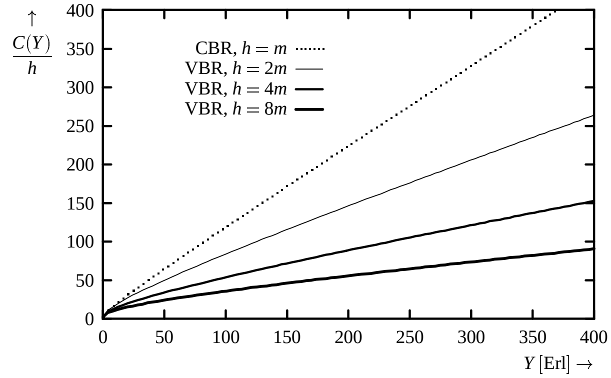

Figure 2: Approximated required capacity above load for CBR and VBR on-/off connections, $P_{\text{block}} = 10^{-2}$, $P_{\text{loss}} = 10^{-9}$, buffer size $K = hb$.

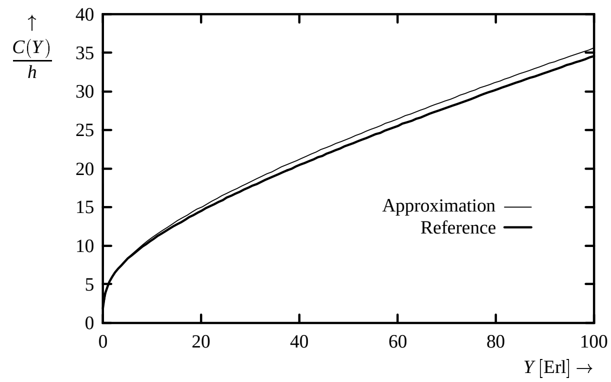

Figure 3: Approximated and reference values of required capacities above load for VBR on-/off connections, $h = 8m$, $P_{\text{block}} = 10^{-2}$, $P_{\text{loss}} = 10^{-9}$, buffer size $K = hb$.

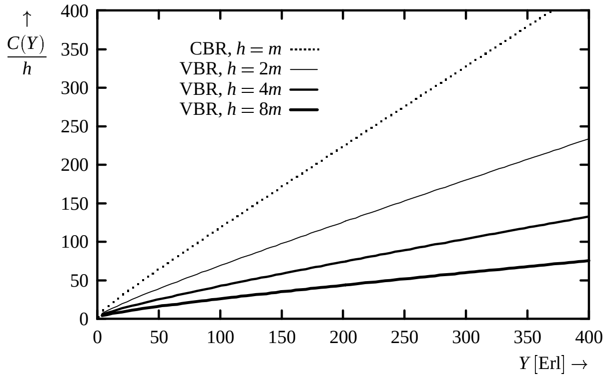

Figure 4: Approximated required capacity above load for CBR and VBR on-/off connections, $P_{\text{block}} = 10^{-2}$, $P_{\text{loss}} = 10^{-9}$, buffer size $K = 10hb$.

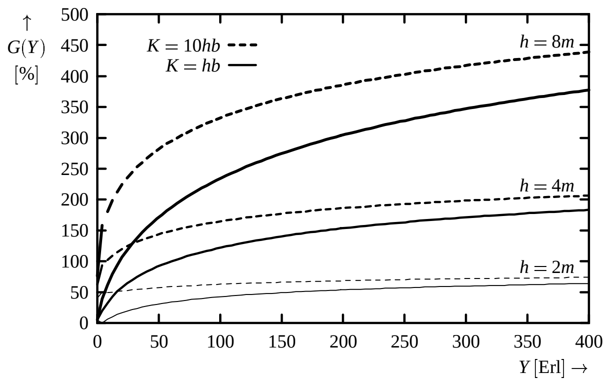

Figure 5: Gain above load for VBR on-/off connections, $P_{\text{block}} = 10^{-2}$, $P_{\text{loss}} = 10^{-9}$.

References

[1] J. Virtamo and S. Aalto. Remarks on the effectiveness of dynamic VP bandwidth management. *COST temporary document 257TD(97)15*, Jan. 1997, http://www-info3.informatik.uni-wuerzburg.de/cost/html/jan97.html.

[2] S.-O. Larsson and Å Arvidsson. Performance evaluation of a local approach for VPC capacity management. *IEICE Transactions on Communications*, E81-B(5):92–98, May 1998.

[3] R. Guérin, H. Ahmadi and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Journal on Selected Areas in Communications*, 9(7):968–981, Sept. 1991.

[4] F. Brichet. Comparison of measurement-based admission control methods using bufferless multiplexing. *COST temporary document 257TD(98)28*, May 1998, http://www-info3.informatik.uni-wuerzburg.de/cost/html/may98.html.

[5] J. Y. Hui. Layered required bandwidth for heterogeneous traffic. *Proc. INFOCOM'92*, 1:13–20.

[6] D. Mitra, M. I. Reiman and J. Wang. Robust admission control for heterogeneous ATM systems with both cell and call QoS requirements. *Proc. ITC-15: Teletraffic Contributions for the Information Age*, 2b:1421–1432, Elsevier Science B.V., 1997.

[7] M. Fiedler and H. Voos. *Fluid flow-Modellierung von ATM-Multiplexern. Mathematische Grundlagen und numerische Lösungsmethoden.* München: Utz Verlag, 1997.

[8] M. Fiedler. Formula for the required capacity of an ATM multiplexer. *Proc. NTS-14*, 367–379, Copenhagen, Aug. 1998.

[9] D. E. Goldberg. *Genetic Algorithms.* München: Addison-Wesley Publishing Company, 1989.

[10] M. Fiedler and Å. Arvidsson. A resource allocation law to satisfy QoS demands on ATM burst and connection level. *COST temporary document 257TD(99)06*, Jan. 1999, http://www-info3.informatik.uni-wuerzburg.de/cost/html/jan99.html.

[11] A. I. Elwalid and D. Mitra. Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. *IEEE/ACM Transactions on Networking*, 1(3):329–343, June 1993.