TOWARDS PRECISION TOOLS FOR ATM NETWORK DESIGN, DIMENSIONING AND MANAGEMENT

Sándor Molnár[1], András Faragó[1], Tamás Henk[1], Søren Blaabjerg[2]

Abstract

It is a critical issue in network dimensioning that the characterization of traffic at the call level should be accurate enough to provide the designer with reliable tools for dimensioning the transmission and switching capacities. Since in B-ISDN the nature of traffic is expected to be very different from traditional telephone traffic with much more complex features, therefore, new methods are needed to provide a satisfactory description. In this paper we present an approach that characterizes the traffic demand at the call level in a refined way, namely, by using a two-parameter description instead of the traditional one-parameter characterization. This approach contributes to the more accurate description of traffic demands at the call level, in order to provide the network designer and manager with precision tools to handle traffic demands and their consequences in dimensioning and related issues, while retaining simplicity, algorithmic feasibility and practical applicability.

[1] Department of Telecommunications and Telematics, Technical University of Budapest, Sztoczek u. 2., H-1111 Budapest, Hungary,
E-mail: molnar@bme-tel.ttt.bme.hu, farago@ttt-202.ttt.bme.hu, henk@bme-tel.ttt.bme.hu

[2] Institute of Telecommunications, Technical University of Denmark, DK-2800 Lyngby, Denmark,
E-mail: sbb@it.dtu.dk

1 Introduction

ATM networks, the standardized carriers of the future B-ISDN, provide a major challenge to many aspects of networking. Among them, a very important point is the methodology of network design and dimensioning that also influence the tools for efficient network management.

For traditional telephone networks many such methods have been developed, see e.g. [Gir90]. Quite a few of them can already be considered as classical. There are several reasons, however, that make the applicability of traditional methods very restricted in ATM networks. Let us list some of the arguments that call for a substantially upgraded methodology.

- The nature of B-ISDN traffic is expected to be very different from plain telephone traffic for a number of reasons:
  1. the arrival of connection requests can change from service to service;
  2. the holding time of calls are expected to deviate from the traditionally considered exponential distribution and the nature of this deviation depends on the service class;
  3. burstiness related properties play an essential role, for example, virtual LANs can be also configured on top of the physical ATM network, generating a traffic pattern that is bursty on every time scale;
  4. the network carries multiclass traffic, where even the definition of very basic concepts, such as blocking probability, becomes a nontrivial problem etc.

- Routing and related functions are expected to become more complex in a high speed ATM network, because
  1. multiclass traffic has to be routed, where the different classes have different nature, while sharing the same infrastructure network, using e.g. statistical multiplexing;
  2. network safety and reliability has increased importance, since the failure of a high capacity link or node effects many users;
  3. flexible reconfiguration capability, including re-routing of traffic, is essential in case of failures or changes in demand etc. Clearly, the increased complexity of routing and related functionalities makes the statistical estimation of traffic descriptor quantities, such as link load, route load, various blocking measures etc. substantially more complex.

- Network design and dimensioning functions are much less separated from management functions than it has been in traditional telephone networks. For example, flexible reconfiguration, controlled by the operator or in some cases the customers, requires the capability of running design/dimensioning type algorithms very fast.
- What data and when should be collected to support network management optimally? This is a difficult question in an integrated network. Too much information implies excessive overhead and longer reaction time, while too little information may degrade

safety and performance. This also raises the problem that what part of the data should be collected by measurement and what is the part that can already be computed or estimated from the measured data, thus providing a trade-off between transmission, measurement and processing overhead.

The above (extendable) list of problems clearly call for a substantially upgraded methodology to support efficient network design, dimensioning and management in ATM carried B-ISDN. Such a collection of new methods, of course, involves a longer and complex development process with continuous feedback from the practical applications.

In the present paper we would like to contribute to this development process by considering a specific issue on dimensioning, namely what can be gained if the usual one parameter traffic demand estimation is enlarged to a two parameter estimation. The experience with telephone traffic is that the Poisson process which is described by a single parameter constitutes a natural and accurate model for the arrival of call attempts, and its memoryless property ensures that the so-called insensitivity property i.e. most quantities of interest depends on the distribution of the holding time only through the mean.

It is highly unlikely that the Poisson property carries over to many other services in a B-ISDN context. The connection request process for some services may have a rather regular pattern while for other services it may come out very bursty. The first way to think of to solve this problem is to include in the traffic demand matrix a two parameter description, one parameter for the usual demand and a parameter characterizing the variability of the arrival of connection requests.

The objective of the present paper is three fold. In section 2 a variability parameter which it is feasible to measure is suggested. In section 3, two approximate models are presented which from the traffic demand and the variability measure compute the occupancy distribution of a link to which this traffic is offered. Finally, in section 4 the results obtained for the single link are used in a network model to find blocking probabilities on an end-to end basis. To illustrate the potential gain which can be obtained from a two parameter dimensioning model, we consider a link partitioning example and a network partitioning example.

2 OBTAINING A MEASURABLE VARIABILITY MEASURE

From traditional telephone traffic, the issue of variability in the arrival process of connection requests has been investigated mostly for overflow traffic in systems with alternative routing [DPK87].

Two measures has been intensively used. The most straightforward is the squared coefficient of variation of the interarrival time between two consecutive connection requests [Cox62]. In the case where the arrival process is well described by a renewal process this gives a complete second order characterization [Cox62]. However in the general case, it only gives one component in the characterization [CLe66]. Another important disadvantage comes from the fact that the blocking probability and the occupancy distribution also depends on the distribution and not only the mean of the holding time in the case without Poisson arrivals. A more accurate variability measure is therefore needed. The generalized peakedness measure as defined by [Eck82] has the advantage that it is a complete second order characterization of the arrival process and furthermore also takes the holding time process into account. The disadvantage is that it is not as straight forward to understand.

The definition is as follows: Assume that the arrival of connection requests are offered to a link with infinite capacity. Let *L(t)* be the amount of bandwidth occupied at time *t*. Then the generalized peakedness *Z* is defined as:

$$Z = \frac{Var\{L(t)\}}{E\{L(t)\}} \tag{1}$$

On a route in a real network it is possible only to measure the actual occupied bandwidth and here only accepted connections contributes. However, since also the amount of blocked connections needs to be monitored, it is possible by combining the occupancy distribution of carried call and the process of connection requests which are blocked to obtain an estimate of the occupancy distribution in the infinite capacity case and thereby get a measured estimate of the peakedness of the connection request on the route.

By characterizing both the arrival process and holding time the computation of peakedness is also possible. In [Eck82] the peakedness of the occupancy distribution has been studied assuming only that the arrival process is stationary. [Eck82] presents several different formulations for the peakedness, but here we shall restrict ourselves to two different formulations.

Let $U(x)$ denote the renewal function of the process S that is: $U(x) = E[N(a, a+x)]$ with $N(a,b)$ denoting the number of arrivals in the interval $]a,b]$, when an arrival occurred at time a.

Furthermore, define $H_2^c(x) = \int_{-\infty}^{\infty} (1-H(u))(1-H(u-x))\,du$ (2)

According to formula (3) in [Eck82] then the peakedness Z of the complementary holding time distribution 1-H is

$$Z(H) = 1 + 2\mu \int_{0-}^{\infty} H_2^c(x)\,dU(x) - \frac{m}{\mu} \tag{3}$$

If the holding time is exponentially distributed with mean $1/\mu$, the peakedness formula reduces to:

$$Z_{exp}(\mu) = 1 + U^*(\mu) - \frac{m}{\mu} \tag{4}$$

in which U^* denotes the Laplace-Stieltjes Transform of U.

Restricting the holding time distributions to be of the class which,

$$1 - H(x) = \int_{-\infty}^{\infty} e^{-xt} a(t)\,dt \quad \text{for } x > 0 \tag{5}$$

and a is a generalized function, [Eck82] obtains the following relation between the peakedness of H and the peakedness of an exponential holding time distribution with mean $1/\mu$.

$$Z[H] = 1 + 2\mu \int_{0-}^{\infty} \alpha(y)\,(Z_{exp}(y) - 1)\,dy \tag{6}$$

where

$$\alpha(y) = a(y) \int \frac{a(x)}{x+y}\,dx \tag{7}$$

In Appendix A it is shown how expression (5), (6) and (7) are applied when the holding time distribution is Coxian and a closed form expression for the generalized peakedness has been derived.

3 ONE LINK ANALYSIS

In this section we consider two approximations based on matching the mean and the variance of the occupancy distribution in the infinite capacity case. Furthermore, blocking probabilities derived from the truncated occupancy distribution are described and investigated.

3.1 The BPP Approximation

The BPP arrival process is a state dependent Poisson process characterized by two parameters α and β such that the Poissonian arrival intensity when k servers are occupied is $\alpha + k\beta$. In the case $\beta = 0$ it reduces to the plain Poisson process, for $\beta < 0$ it represents a process of less variability than Poisson (finite source model), and for $\beta > 0$ it represents a process of higher variability than Poisson see[Del81]. The mean and variance of the occupancy distribution turns out to be:

$$M = \frac{\alpha}{\mu - \beta} \qquad V = \frac{\alpha\mu}{(\mu - \beta)^2} \tag{8}$$

where $1/\mu$ is the mean holding time.

With the mean M and variance V given (from e.g the Eckberg approach) we suggest to approximate the occupancy distribution with the BPP distribution with the same mean and variance. The BPP parameters α and β then should be chosen as:

$$\beta = 1 - \frac{1}{Z} \qquad \alpha = M(1 - \beta) \tag{9}$$

assuming a mean holding time of 1.

As blocking probability we use the traffic congestion defined as:

$$TC = \frac{OT - CT}{OT} \tag{10}$$

where OT and CT denote the offered and carried traffic, respectively.

The offered traffic is the mean number of occupied servers in the infinite system (mean holding time = 1), and the carried traffic is the mean of the distribution obtained by truncating the occupancy distribution in the infinite server case at C (the link capacity) and renormalizing it:

$$OT = M,\ CT = \frac{\sum_{j=0}^{C} jp_j}{\sum_{i=0}^{C} p_i} \tag{11}$$

with $p_i = p_{i-1}\frac{\alpha + (i-1)\beta}{i}$ for $i > 0$, $p_i = 0$ for $\beta < 0$ and $i > N$ (in the case of $\beta < 0$ the β is modified to the closest value such that $N = -\alpha/\beta$ must be an integer), and

$p_0 = (1-\beta)^{\frac{\alpha}{\beta}}$ for $\beta \neq 0$, $p_0 = e^{-\alpha}$ for $\beta = 0$.

3.2 The Maximum Entropy Approximation

The concept of entropy appears in the mathematical theory of interconnecting networks and in the queueing theory [Ben65], [Kou86]. In queueing theory we only know of its existence in the 1-server case and usually it has been applied in a way where only average quantities like mean queue length and utilization has been matched. Here we shall apply this technique on a many server loss system where also the variance is matched.

The basic idea of the method is based on Bernoulli's principle of insufficient reason [Har92], which states that all events over a sample space should have the same probability unless there is evidence to the contrary. The entropy plays as a measure of the certainness of the outcome of an event. The more uncertain the value of a random variable is, the bigger the entropy is. In order to fulfil the Bernoulli's principle the entropy has to be maximized under the constrains of the mean and the variance which we would like to be matched.

Consider a stationary stochastic process $X(t)$ in a discrete state space and let p_i be the probability of being in state i. The entropy of $X(t)$ with stationary distribution $\{p_i\}$ is defined as:

$$H(\boldsymbol{p}) = -\sum_i p_i ln p_i \tag{12}$$

The idea here is, as an approximation for the occupancy distribution, to take the one which maximizes the entropy under the constrains that

- it should be a proper probability distribution i.e $\sum_{i=0}^{\infty} p_i = 1$

- the mean should be correct i.e $\sum_{i=0}^{\infty} ip_i = E(X) = \frac{m}{\mu}$
- and the second moment should be correct i.e.

$$\sum_{i=0}^{\infty} i^2 p_i = E(X^2) = \frac{m}{\mu}\left(Z(1-H) + \frac{m}{\mu}\right)$$

In Chapter 8.4.1 of [Har92] the following theorem is presented and proved

THEOREM: The probability mass function $\{p_i\}$ which maximizes $H(\boldsymbol{p}) = -\sum_i p_i ln p_i$

subject to $\sum_{i=0}^{\infty} p_i = 1$

and $\sum_i f_j(i) p_i = \bar{f}_j$ for $1 \le j \le k$

(where $\{\bar{f}_j | (1 \le j \le k)\}$ are prescribed mean values of the functions $\{f_j\}$)

is:

$$p_i = g \prod_{j=1}^{m} x_j^{f_j(i)} \tag{13}$$

where g is the normalization constant.

If this result is applied to a single server queue and the mean queue length is matched, the queue length distribution which maximizes entropy is geometrical thus yielding the exact distribution for the queue length in the M/M/1 case.

When matching the first and second moment in the infinite server case, a straightforward application of the theorem shows that the occupancy distribution which maximizes entropy is:

$$p_i = P\{X = i\} = g x_1^i x_2^{i^2} = g e^{i ln x_1} e^{i^2 ln x_2} \tag{14}$$

Thus the maximum entropy approach enforces a distribution which comes from sampling the normal density function at non-negative integer values. The 3 equations needed to match the mean, variance and obtain a proper distribution is:

$$\sum_{i=0}^{\infty} g x_1^i x_2^{i^2} = 1 \qquad \sum_{i=0}^{\infty} g i x_1^i x_2^{i^2} = E\{X\} \qquad \sum_{i=0}^{\infty} g i^2 x_1^i x_2^{i^2} = E\{X^2\} \tag{15}$$

and we have only been able to solve them by heuristic methods. However, applying the corresponding continuous approach fitting a normal density function restricted to the positive line yields a system of equation which can be solved in an exact way as demonstrated in Appendix B. Our numerical results indicates that the difference between the continuous and the discrete results are very small.

From the occupancy distribution (see Appendix B) $f(x) = \begin{cases} ge^{-\alpha x}e^{-\beta x^2} & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x < 0 \end{cases}$ a traffic congestion (10) can be derived with $OT = M$ and (16)

$$CT = \frac{\frac{-\alpha}{2\beta}\left(\Phi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) - \Phi\left(-\sqrt{2\beta}C - \frac{\alpha}{\sqrt{2\beta}}\right)\right) + \frac{1}{\sqrt{2\beta}}\left(\varphi\left(\frac{\alpha}{\sqrt{2\beta}}\right) - \varphi\left(\sqrt{2\beta}C + \frac{\alpha}{\sqrt{2\beta}}\right)\right)}{\Phi\left(\sqrt{2\beta}C + \frac{\alpha}{\sqrt{2\beta}}\right) - \Phi\left(\frac{\alpha}{\sqrt{2\beta}}\right)}$$

obtained by truncating the distribution at the link capacity (C) and renormalizing it. Φ is the standard normal distribution and φ is the corresponding density function

3.3 Numerical Example

Here we consider a simple numerical example which demonstrates the effect of using the variability measure.

Consider a 150 Mbit/s link which is loaded to 140 Mbit/s and carries two different traffic types: Traffic type 1 is 2 Mbit/s circuit emulation with peakedness of 0.25, and the Traffic type 2 is 2 Mbit/s frame relay with peakedness of 15. Consider a situation where 60 Mbit/s traffic offered to the link from both traffic types and we would like to share the capacity of the link to the two traffic streams such that the total carried traffic will be maximum. If we do not use any variability measure we are restricted to share the capacity only based on the offered traffic and using e.g. Erlang formula to compute the blocking probabilities. This way we get the equally partition solution: 70-70 Mbit/s. By using the peakedness as a variability measure and computing the blocking probabilities based on the above described methods by the BPP and the Maximum Entropy approximations and partition the capacity such that the total carried traffic is maximized we get the following results (Figure 1):

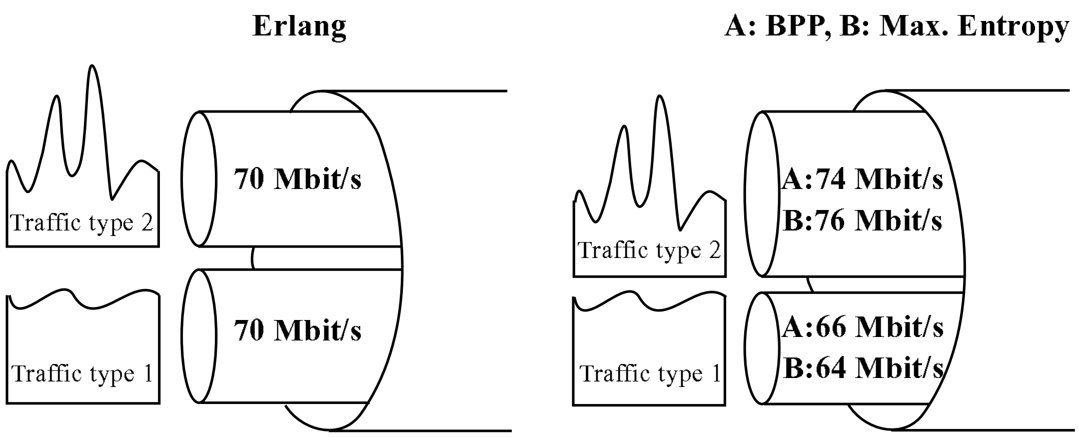

Figure 1 Link partitioning

From the result we can see that the bursty traffic (Traffic type 2) requires a bigger capacity and the smooth traffic (Traffic type 1) requires smaller capacity compared to the equally partitioning case. Also we can conclude that the BPP and Maximum Entropy methods give practically the same results.

The results clearly illustrates that in order to achieve the optimum capacity sharing related to the maximum total carried traffic we need to take into account the variability measure.

4 APPLICATIONS TO NETWORK PARTITIONING

In this section we demonstrate the effect of the variability measure in an ATM network configuration problem.

Consider an ATM network with some given logical subnetworks and traffic demand for each route, where we want to find the partition of the physical capacities related to the subnetworks that maximizes the total carried traffic. For this problem a solution can be found in [FBH94] which is based on the Erlang fixpoint method [Kel91].

Now we consider an extension of this network dimensioning algorithm with using the variability measure. The main purpose of this extension is to improve the network dimensioning algorithm to fulfil the expected nature of the future ATM, where the call arrival process will differ significantly from the Poisson process and the holding time distribution will be deviate from the exponential distribution.

For dimensioning purposes the traffic offered to each route is characterized by the mean and the peakedness. Based on these two parameters we are using the proposed blocking measures (based on the BPP and Maximum Entropy methods) to compute link blocking probabilities. To upgrade the dimensioning algorithm in [FBH94] only the replacement of the Erlang formula to the suggested blocking measures $B_j = B\left(\rho_j, \sigma^2_j, C_j\right)$ are needed [Mol94], where the blocking probability for link j computed from the proposed blocking measures with ρ_j aggregated offered traffic, σ^2_j variance and C_j capacity.

For the computation of the aggregated offered traffic on a link we use the same reduced load and link independence assumption as in [FBH94] and so: $\rho_j = (1-B_j)^{-1}\sum_r A_{jr}\nu_r \prod_i (1-B_i)^{A_{ir}}$ where ν_r is the average rate of the offered traffic to route r and a call on route r requires A_{jr} units of capacity on link j. For the calculation of the variance of the aggregated offered traffic we simple assume that the variance of the offered traffic is thinned by the same factor as the mean. It means that we keep the peakedness at the same value. However, it should be noted that the changing of the peakedness of the traffic stream going through the network is affected by the congestions on each link and very dependent on the burstiness of the offered traffic. This effect is rather complex and we use this simple approach as a first approximation for the changing of the peakedness. Therefore the variance of the carried traffic can be computed by $\sigma^2_j = (1-B_j)^{-1}\sum_r A^2_{jr}\omega^2_r \prod_i (1-B_i)^{A_{ir}}$ where ω^2_r is the variance of the number of the offered calls.

4.1 Numerical example

The algorithm is demonstrated in a small network example. In this 4 nodes network arranged in a ring (Figure 2.) carries two fully connected logical subnetworks with 3 nodes as shown in Figure 2.

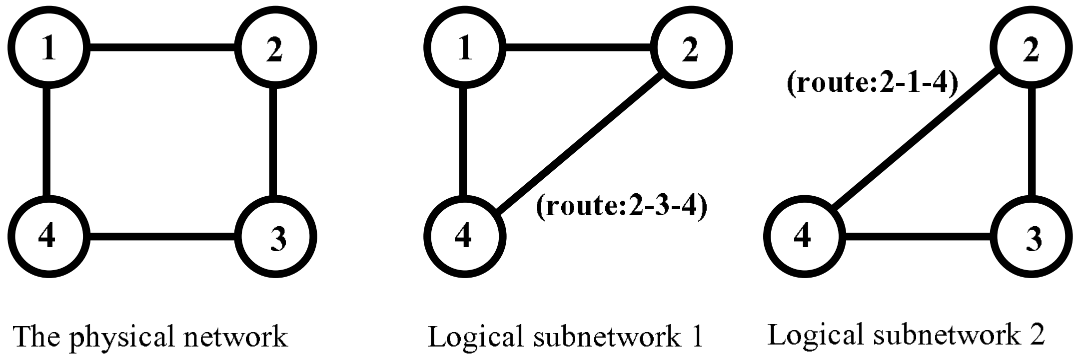

Figure 2 A network example with two logical subnetworks

Each physical link of the network has 45 Mbit/s capacity and each logical subnetwork carries two types of traffics:

- Frame relay: effective bandwidth: 0.75 Mbit/s, mean holding time: 60 s
- DS-1 circuit emulation: effective bandwidth: 1.5 Mbit/s, mean holding time: 480 s

The arriving rate of the calls from different traffic types are set such that the load be equally shared among the traffic types on a link. The partitioning results from the original fixpoint method (which does not take into account any variability measure) and from the extended method (which uses the peakedness of the traffic as described in the previous section, and we used peakedness of 1 to the logical subnetwork 1 and peakedness of 15 to the logical subnetwork 2) are shown in Table 1.

Table 1: Capacity partitioning (Capacity to log. subnetw. 1 - Capacity to log. subnetw. 2)

Link	Fixpoint	Fixpoint-BPP
1-2	22.5-22.5	15.8-29.2
2-3	22.5-22.5	13.8-31.2
3-4	22.5-22.5	13.8-31.2
4-1	22.5-22.5	15.8-29.2

The results show that the logical subnetwork 2, which carries rather bursty traffic, requires more capacities (and the logical subnetwork 1, which carries smooth traffic, requires smaller capacities) on the links compared to the results of the original fixpoint method which equally partitioned the links between the two logical subnetworks. The optimal partitioning corresponding to the maximum of the carried traffic can be obtained by the above link partitioning and indicates the importance of the variability measure.

5 CONCLUSION

In this paper we have presented a two-parameter traffic characterization consists of the mean and the variability of the arrival of connection requests. We have suggested the generalized peakedness as a variability measure by which an accurate description of traffic demands can be obtained. It has been demonstrated that the proposed characterization of traffic provides the network designer and manager with a precision tool to solve network dimensioning problems retaining the algorithmic feasibility and practical applicability. Finally, the contribution of the paper can be summarized as follows:

1. proposal for a two-parameter traffic characterization using the concept of the generalized peakedness
2. a new closed form expression for the generalized peakedness in case of Coxian holding time distributions
3. a traffic congestion measure based on the Bernoulli-Poisson-Pascal (BPP) approximation and using the generalized peakedness
4. a traffic congestion measure based on a new Maximum Entropy method and using the generalized peakedness
5. a new network dimensioning algorithm based on the fixpoint method and using the proposed traffic measures.

ACKNOWLEDGEMENTS

The research was supported by Ellemtel Telecommunication System Laboratories.

APPENDIX A: DERIVING THE PEAKEDNESS FOR COXIAN HOLDING TIME

In this section we derive the peakedness formula for cases with Coxian holding time distributions.

Consider a Coxian distribution represented by a weighted sum of generalized Erlang distributions as shown in Fig. A1. For simplicity reasons we assume that $\lambda_i \neq \lambda_j$ for $i \neq j$ but the general case can be included with only minor changes in the expressions.

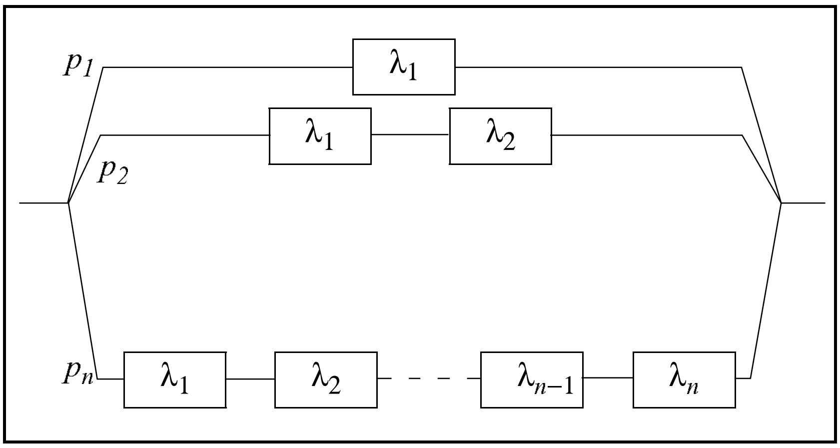

Fig. A1: Coxian distribution represented as a weighted sum of generalized Erlang distributions

For the Coxian distributions *a(t)* in (5) can be written as a series of delta functions[1]:

$$a(t) = \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \left(\sum_{j=i}^{n} p_j \prod_{k=i+1}^{j} \frac{\lambda_k}{\lambda_k - \lambda_i} \right) \delta_{\lambda_i} \qquad \text{(A.1)}$$

where λ_i ($\lambda_i \neq \lambda_j$ for $i \neq j$) and p_i, $1 \leq i \leq n$, are the intensities and branch probabilities of an n-branch Coxian distribution, respectively. (See Fig. A1.) By applying (7) the result of the integral is

$$b(y) = \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \left(\frac{1}{y + \lambda_i} \right) \left(\sum_{j=i}^{n} p_j \prod_{k=i+1}^{j} \frac{\lambda_k}{\lambda_k - \lambda_i} \right) \qquad \text{(A.2)}$$

Multiplying by *a(y)* and substituting the obtained $\alpha(y)$ into (6) we get the peakedness:

$$Z = 1 + 2\mu \sum_{l=1}^{n} p_l \sum_{i=1}^{l} \left(\prod_{j=1}^{i-1} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \left(\prod_{k=i+1}^{l} \frac{\lambda_k}{\lambda_k - \lambda_i} \right) b(\lambda_l)\,(Z_{exp}(\lambda_l) - 1) \qquad \text{(A.3)}$$

where Z_{exp} is given in (4).

The importance of (A.3) should be realized. For any stationary arrival process for which the Laplace transform of the renewal function is available, which includes e.g. renewal processes, Markov renewal processes and doubly stochastic processes, insertion of formula (4) in (A.3) provides a simple closed form expression for the peakedness of any Coxian holding time distribution.

APPENDIX B: MATCHING MEAN AND VARIANCE

The truncated normal density function is given as:

$$f(x) = \begin{cases} g e^{-\alpha x} e^{-\beta x^2} & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases} \qquad \text{(B.1)}$$

and has moment generating function given by

$$M(s) = \int_0^{\infty} e^{sx} g e^{-\alpha x} e^{-\beta x^2} dx = g \sqrt{\frac{\pi}{\beta}} e^{\frac{(s-\alpha)^2}{4\beta}} \Phi\left(\frac{s-\alpha}{\sqrt{2\beta}} \right) \qquad \text{(B.2)}$$

Thereby the system of equations which needs to be solved to match mean and variance and still have a proper probability distribution is:

1. When $\lambda_i = \lambda_j$ for some i and j, derivatives of delta functions appear in the expression for $a(t)$.

$$g\sqrt{\frac{\pi}{\beta}}e^{\frac{\alpha^2}{4\beta}}\Phi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) = 1$$

$$g\sqrt{\frac{\pi}{\beta}}e^{\frac{\alpha^2}{4\beta}}\left\{\left(-\frac{\alpha}{2\beta}\right)\Phi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) + \frac{1}{\sqrt{2\beta}}\varphi\left(\frac{-\alpha}{\sqrt{2\beta}}\right)\right\} = E\{X\}$$

$$g\sqrt{\frac{\pi}{\beta}}e^{\frac{\alpha^2}{4\beta}}\left\{\left(\frac{\alpha^2}{4\beta^2} + \frac{1}{2\beta}\right)\Phi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) - \frac{\alpha}{(2\beta)^{\frac{3}{2}}}\varphi\left(\frac{-\alpha}{\sqrt{2\beta}}\right)\right\} = E\{X^2\} \qquad \text{(B.3)}$$

Dividing the normalization equation up into the two moment equations, and multiplying the results with $\sqrt{2\beta}$ and 2β respectively, yields:

$$G(-x) - x = \sqrt{2\beta}E\{X\} \quad \text{and} \quad x^2 + 1 - xG(-x) = 2\beta E\{X^2\}$$

where $x = \frac{\alpha}{\sqrt{2\beta}}$ and where $G(x) = \frac{\varphi(x)}{\Phi(x)}$. Squaring the left hand equation and dividing the result into the right hand equation yields the following equation with x as the only unknown.

$$\frac{x^2 + 1 - xG(-x)}{(G(-x) - x)^2} = \frac{E\{X^2\}}{E\{X\}^2}$$

The function $x \to \frac{x^2 + 1 - xG(-x)}{(G(-x) - x)^2}$ is increasing with range]1,2[which can be seen by applying the inequality $\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \frac{15}{x^7} \le \frac{\Phi(-x)}{\varphi(x)} \le \frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5}$ valid for $x > 0$.

This implies that for occupancy distributions with $0 < Z < E\{X\}$ a unique truncated normal distribution with same mean and variance exists.

REFERENCES

[Cox62] D. R. Cox, "Renewal Theory", Methuen, 1962.

[CLe66] D. R. Cox, P. A. W. Lewis, "The Statistical Analysis of Series of Events", Methuen, 1966.

[Ben65] V.E. Benes, "Mathematical Theory of Interconnecting Networks", Academic Press, 1965.

[Del81] L.E.N. Delbrouck, "A Unified Approximate Evaluation of Congestion Functions for Smooth and Peaky Traffic", IEEE Trans. on Comm. Vol 29, Feb.1981, pp. 85-91.

[DPK87] Dziong, Piero, Koerner "An Adaptive Call Routing Strategy for Circuit switched Networks", Seventh Nordic Teletraffic Seminar, 1987.

[Eck82] A.E.Eckberg, "Generalized Peakedness of Teletraffic Processes", ITC-10 Session 4.4b, Montreal, 1982.

[FBH94] A. Faragó et. al. "Resource separation - an Efficient Tool for Optimizing ATM Network Configuration", NETWORKS'94, Sept. 1994.

[Gir90] A. Girard: Routing and Dimensioning in Circuit-Switched Networks, Addison-Wesley 1990.

[Har92] P.G. Harrison, "Performance Modelling of Communications Networks and Computer Architectures", Addison Wesley 1992.

[Kel91] F. P. Kelly, "Loss Networks", The Annals of Applied Probability, 1(1991/3), pp. 319-378.

[Kou86] D.D. Kourvatsos, "Maximum Entropy and the G/G/1/N queue", Acta Informatica, Vol. 23, pp. 545-465, 1986.

[MBl94] S. Molnár, S. Blaabjerg, "On some simple blocking approximations for the G/G/c queue", COST 242 TD(94)001, Barcelona, Spain, Feb. 10-11, 1994.

[Mol94] S. Molnár, "Network dimensioning based on refined blocking measures", Internal Report, Version 2, Department of Telecommunication and Telematics, Technical University of Budapest, Budapest, Hungary, May, 1994.