BLOCKING PROBABILITIES IN B-ISDN

Sándor Molnár
Technical University of Budapest
Department of Telecommunications and Telematics
Sztoczek u. 2
H-1111 Budapest, Hungary

Søren Blaabjerg
Ellemtel
Scheelevägen 17
S-223 70 Lund
Sweden

Abstract: An appropriate model for the investigation of the call scale traffic in B-ISDN is the G/G/c loss system. The exact evaluation of this system is complicated and even for the subclass of phase-type renewal arrival processes is limited to rather small system sizes. In this paper we consider two simple approximations based on matching the mean and variance of the occupancy distribution. The first approach is the classical BPP approximation while the second is an Enropy maximization method. Based on this approximations a number of approximate blocking measures have been evaluated by a numerical example.

1 INTRODUCTION

The Poisson process has been a very successful and reasonable accurate model for analyzing traditional telephone networks. For the coming B-ISDN it is rather unlikely that the call request process of the large variety of services are all well approximated by the Poisson process. The nice property that the occupancy distribution is only influenced by the holding time distribution through its mean is only valid for the Poisson distribution, and therefore both more general arrival processes and more general holding time distributions must be taken into account i.e the G/G/c loss system must be studied.

So far one of the few approaches to obtain the exact steady state distribution of a G/G/c system is by restricting the arrival process to be of the type Markovian Arrival Process, see [Neu79] and to assume the holding time distribution to be of phase type. Then the system can be analyzed through the usual Markovian approach [Neu81] and [Han83]. However, the size of the state space very quickly exceeds any in practice manageable size even when the sparseness of the transition matrix is taken into account, thereby limiting the feasible models to those with a rather limited number of phases in the arrival process, a very limited number of phases of the holding time distribution and only a small number of servers.

By restricting the considerations to finding the mean and variance of the distribution of the number of occupied servers in an infinite server group, [Eck83] presents formulas for finding the peakedness factor (defined as the fraction variance to mean) for a general stationary arrival process and general holding time distribution. The fact that it is relatively easy to obtain the mean and variance of the distribution, but difficult to obtain the distribution itself, calls for approximations.

The objective of this paper is to present two simple approximations for the occupancy distribution based on matching the mean and the variance of the distribution to be approximated of the corresponding infinite server queue. From that we obtain a number of approximate blocking measures. The first approximation is a BPP approximation while in the second approach, we use the distribution which maximizes the entropy subject to constraints on the mean and variance. Furthermore, we demonstrate how it is possible to get closed form expressions for the peakedness in case of Coxian holding time distributions. See also [MBl94].

We do not investigate the issue of how the third and higher moments affects the occupancy distribution but refer to the work by [Hol73] in which he derives some bounds based on the Laplace transform of the occupancy distribution.

2 OCCUPANCY DISTRIBUTION IN THE G/G/∞ CASE

In this section the model under study is an infinite server group which is offered traffic from a stationary process S with arrival intensity m. All servers act independently of each other and have the same holding time distribution H which is assumed general but with finite mean $1/\mu$ and variance σ^2.

Exact Occupancy Distribution in Case of Phase Type Arrivals

Ramaswami and Neuts have studied the infinite-server queue with phase type input and general holding time distribution [RNe80]. From a linear system of differential equations they derive the time dependent occupancy distribution. By numerical integration until a sufficiently large time value the steady state distribution can be found. The advantage of the method is that the complexity is independent of the holding time distribution. However if the holding time distribution is very simple, exponential or two phase Coxian a direct Markovian steady state analysis is more efficient.

The Peakedness of the Occupancy Distribution

In [Eck83] the peakedness (variance to mean ratio) of the occupancy distribution has been studied assuming only that the arrival process is stationary. [Eck83] 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 $H_2^c(x) = \int_{-\infty}^{\infty} (1 - H(u))\,(1 - H(u-x))\,du$

According to formula (3) in [Eck83] then the peakedness Z of the complementary holding time distribution 1-H is

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

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

$$Z_{exp}(\mu) = 1 + U^*(\mu) - \frac{m}{\mu} \tag{2}$$

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

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

$$1 - H(x) = \int_{-\infty}^{\infty} e^{-xt} a(t)\, dt \quad \text{for } x > 0 \tag{3}$$

and a is a generalized function, [Eck83] obtains the following relation between the peakedness of H and the peakedness of an exponential holding time distribution with mean $1/\mu$.

$$Z[H] = 1 + 2\mu \int_{0-}^{\infty} \alpha(y)\, (Z_{exp}(y) - 1)\, dy \tag{4}$$

where

$$\alpha(y) = a(y) \int \frac{a(x)}{x+y} dx \tag{5}$$

In appendix A it is shown how expression (3), (4) and (5) are applied when the holding time distribution is Coxian.

3 APPROXIMATIONS FOR THE DISTRIBUTION

Here we shall consider two different approximation procedures both based on matching the mean and variance of the occupancy distribution in the infinite server case.

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] for more details.

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:

$$\beta = 1 - \frac{1}{Z} \qquad \alpha = M(1 - \beta) \tag{6}$$

assuming a mean holding time of 1.

3.2 The Maximum Entropy Approximation

The concept of entropy has played a significant role in various technical disciplines like thermodynamics, statistical mechanics and information theory. The idea also 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.

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:

$$H(\boldsymbol{p}) = -\sum_i p_i \ln p_i \tag{7}$$

The basic idea here is, as an approximation for the occupancy distribution, to take the one which maximizes the entropy under the constrains that; 1) it should be a proper probability distribution i.e $\sum_{i=0}^{\infty} p_i = 1$, 2) the mean should be correct i.e $\sum_{i=0}^{\infty} i p_i = E(X) = \frac{m}{\mu}$, 3) the second moment should be correct i.e. $\sum_{i=0}^{\infty} i^2 p_i = E(X^2) = \frac{m}{\mu}\left(Z(1-H) + \frac{m}{\mu}\right)$

From the theorem of Chapter 8.4.1 of [Har92] we get that the occupancy distribution which maximizes entropy is:

$$p_i = P\{X = i\} = g x_1^i x_2^{i^2} = g e^{i \ln x_1} e^{i^2 \ln x_2} \tag{8}$$

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 are:

$$\sum_{i=0}^{\infty} g x_1^i x_2^{i^2} = 1 \qquad \sum_{i=0}^{\infty} g i x_1^i x_2^{i^2} = E\{X\} \qquad \sum_{i=0}^{\infty} g i^2 x_1^i x_2^{i^2} = E\{X^2\} \tag{9}$$

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.

4 ESTIMATION OF BLOCKING PROBABILITIES

Traffic congestion from the truncation of the exact occupancy distribution

From the exact occupancy distribution in the infinite server case derived by the method of [RNe80] we can obtain a distribution for the finite capacity system by simple truncating and renormalizing it. The number of servers in the truncated system is denoted by c. From the corollary 1.10 of [Kel79] it follows that for reversible Markov processes this truncation procedure gives exact results. For other cases we use it as an approximation.

As blocking probability we use the traffic congestion defined as:

$$\text{traffic congestion} = \frac{A_{offered} - A_{carried}}{A_{offered}} \tag{10}$$

where the offered traffic $A_{offered}$ is the mean number of occupied servers in the infinite system (mean holding time = 1), and the carried traffic $A_{carried}$ is the mean of the truncated distribution.

Traffic congestion from the truncation of the BPP approximation

By applying the BPP approximation procedure and using the truncation approach the traffic congestion can be computed given by (10).

Call congestion from the truncation of the BPP approximation

From BPP approximation following Delbrouck [Del81] we have the approximate call congestion formula: $P_{\text{call cong}} = P_{\text{time cong}}\left(1 + \frac{c}{M}(Z-1)\right)$. The call congestion can differ significantly from the traffic congestion in the BPP model since the truncation immediately effects the arrival process (states above the truncation level disappear).

Traffic congestion from the truncation of the truncated normal distribution

For the case where the occupancy distribution is approximated by the truncated normal distribution the traffic congestion can also be computed from the finite capacity case obtained by truncation and renormalization.

We have the infinite server occupancy distribution $f(x) = \begin{cases} g e^{-\alpha x} e^{-\beta x^2} & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}$ implying that

$$A_{offered} = g\sqrt{\frac{\pi}{\beta}}\, e^{\frac{\alpha^2}{4\beta}} \left\{ \left(-\frac{\alpha}{2\beta}\right) \Phi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) + \frac{1}{\sqrt{2\beta}} \varphi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) \right\} \tag{11}$$

where Φ is the standard normal distribution and φ is the corresponding density function. Thereby $A_{carried}$ fulfils

$$A_{carried} = \frac{\frac{-\alpha}{2\beta}\left(\Phi\left(\frac{-\alpha}{\sqrt{2\beta}}\right) - \Phi\left(-\sqrt{2\beta}c - \frac{\alpha}{\sqrt{2\beta}}\right)\right) + \frac{1}{\sqrt{2\beta}}\left(\varphi\left(\frac{\alpha}{\sqrt{2\beta}}\right) - \varphi\left(\sqrt{2\beta}c + \frac{\alpha}{\sqrt{2\beta}}\right)\right)}{\Phi\left(\sqrt{2\beta}c + \frac{\alpha}{\sqrt{2\beta}}\right) - \Phi\left(\frac{\alpha}{\sqrt{2\beta}}\right)} \tag{12}$$

from which a closed form expression for the traffic congestion can be obtained.

Exact blocking probabilities in case of renewal input and exponential holding times

The solution to the GI/M/c and GI/M/∞ systems has been known for 50 years, see [Pal43]. Both the occupancy distribution at an arbitrary point in time and the occupancy distribution at an arbitrary arrival can be found. From these two distributions it can be concluded, see Theorem 4 of chapter 4 in [Tak62], that the traffic congestion equals the call congestion for the GI/M/c case (see [MBl94]). Even though an analytical solution exists, obtaining numerical values from the solution is challenging and in practice limited to small offered traffic and small c.

5 NUMERICAL RESULTS

We have found the exact blocking probabilities of the GI/M/c case for 8 different interarrival time distributions. Varying the squared coefficient of variation of the interarrival time (scv) from 1/3 to 9 and also varying the type of the distributions we have investigated a broad range of the arrival distributions to evaluate the approximate blocking probabilities. For smooth arrival processes the scv of 1/3 with Erlang-3 distribution and the convolution of the exponential and the deterministic distributions (M and D) are considered. In the scv of 1 case not only the Poisson arrival process (M), but also a Coxian distribution (Cox) of two branches with Erlang-2 and Erlang-3 distributions are investigated. For peaky arrival processes the scv of 3 and 9 cases with Hyperexponential balanced (Hyp.-bal) and unbalanced (Hyp.-unbal) distributions are evaluated. In all cases the offered traffic is kept equal to 10 and the number of servers is 15. Then we have evaluated four different approximative blocking probabilities.

Column 3, 4, 5 and 6 are the blocking probabilities described in the previous section. (In the column "BPP" the traffic congestions computed from the truncation of the BPP approximation are shown which in these cases equals to the call congestions. The traffic congestions based on the truncated normal distribution are shown in the column "tr. normal".)

Finally in column "Hayward" we have applied the classical Hayward approximation in which the blocking probability in a system with c servers and an offered traffic A with peakedness Z is approximated by the blocking probability in a system with c/Z servers and offered Poisson traffic A/Z.

Table 1: Approximate and exact blocking probabilities (offered traffic = 10, c = 15)

scv	type	exact	truncated	BPP	tr. normal	Hayward
1/3	Erlang-3	1.54%	1.49%	1.30%	1.67%	1.74%
1/3	M and D	1.45%	1.39%	1.29%	1.65%	1.73%
1	M	3.65%	3.65%	3.65%	3.88%	3.65%
1	Cox	3.72%	3.73%	3.65%	3.88%	3.65%
3	Hyp.-bal	7.99%	9.27%	9.86%	10.36%	8.52%
3	Hyp.-unbal	10.53%	10.90%	10.86%	11.51%	9.64%
9	Hyp.-bal	15.38%	21.86%	20.71%	23.48%	16.75%
9	Hyp.-unba	25.52%	26.64%	26.18%	29.91%	20.90%

From the Table we can conclude that all the approximations performs reasonable well. The BPP approximation underestimates in the cases with a regular arrival processes while the truncated normal distribution overestimates for peaky arrival processes.

6 CONCLUSION

In this paper we have emphasized the applicability of the concept of generalized peakedness and derived closed form expressions in the case with Coxian holding time distribution. From the generalized peakedness and the mean arrival rate, two simple two moment procedures to obtain approximate occupancy distributions of the G/G/c queue and related blocking formulas have been presented. From the numerical example and our experience in general we conclude that the overall performance is satisfactory taken into account the simplicity of the methods. Furthermore, the simplicity makes the corresponding blocking measures good candidates for link blocking measures in a network dimensioning application where the offered traffic is described by two moments.

7 ACKNOWLEDGMENT

This research was performed as a part of a joint project between Ellemtel Telecommunication System Laboratories, Sweden and the Technical University of Budapest, Hungary. The authors are grateful to Miklós Boda, Géza Gordos and Tamás Henk for their continuous support.

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. 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.

For the Coxian distributions *a(t)* in (3) can be written as a series of delta functions[1]:

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

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. By applying (5) the result of the integral is

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

Multiplying by *a(y)* and substituting the obtained $\alpha(y)$ into (4) we get the peakedness:

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

where Z_{exp} is given in (2).

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 (2) in (A.3) provides a simple closed form expression for the peakedness of any Coxian holding time distribution.

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

APPENDIX B: MATCHING MEAN AND VARIANCE OF TRUNCATED NORMAL DISTRIBUTION

The truncated normal density function has moment generating function given by

$$M(s) = \int_0^\infty e^{sx} g e^{-\alpha x} e^{-\beta x^2} dx = g\sqrt{\frac{\pi}{\beta}} e^{\frac{(s-\alpha)^2}{4\beta}} \Phi\left(\frac{s-\alpha}{\sqrt{2\beta}}\right) \tag{B.1}$$

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

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

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

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

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

[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.

[Eck83] A.E.Eckberg, "Generalized Peakedness of Teletraffic Processes", ITC-10 Session 4.4b, Montreal, 1983.

[Han83] S. Hansen, "Phase Type Distributions in Queueing Theory", Ph.D Thesis no. 41, IMSOR, Denmark, 1983.

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

[Hol73] J.M. Holtzman, "The Accuracy of the Equivalent Random Method with Renewal Input", Bell System Technical Journal, Vol. 52, pp. 1673-1679, 1973.

[Kel79] F.P. Kelly, "Reversibility and Stochastic Networks", Wiley 1979.

[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.

[Neu79] M.F. Neuts, "A Versatile Markovian Point Process", J. Appl. Prob. **16**, 1979, pp. 764-779

[Neu81] M.F. Neuts, "Matrix Geometric Solutions in Stochastic Models", John Hopkins University Press, 1981.

[Pal43] C. Palm, "Intensitätsschwankungen im Fernsprechverkehr", Ericsson Technics, 44, pp. 1-189, 1943.

[RNe80] V. Ramaswami, M.F. Neuts, "Some Explicit Formulas and Computational Methods for the Infinite-Server Queues with Phase-Type Arrivals", J. Appl. Prob. **17**, 498-514, 1980.

[Tak62] L. Takács, "Introduction to the Theory of Queues", Oxford University Press, 1962