Optimal design parameters in capacity assignment for a broadband network

Chung-Ju Chang
Yann-Ju Chu

Indexing terms: Broadband networks, Integrated services, Queueing

Abstract: A capacity assignment strategy for a broadband network is proposed and studied in this paper. This broadband network has a large channel capacity in its elements. It usually has two kinds of customers: one is queueable narrowband (NB), and the other is blockable wideband (WB) which needs real-time delivery. Besides the characteristics of these two types of traffic, since the improvement in the blocking probability of WB is more significant than that in the mean waiting time of NB, the capacity assignment strategy favours the WB traffic by providing bit-rate compression, reserving some channels for WB customers with non pre-emption by NB customers, and permitting NB customers to use free WB channels only when the NB queue exceeds a certain length. An analytic model is developed and the solutions are presented. The results show that, via proper choice of design parameters, the proposed strategy can have superior performance.

1 Introduction

Models of multiserver queue with many heterogeneous customers have many applications and have been widely researched [1–10]. These queues are usually divided into many categories by considering the assignment strategy of servers. The assignment strategy takes into account the characteristics and requirements of the customers, such as the customer's required number of servers, the reaction of a blocked customer, and the release mechanisms of servers. The reaction of a blocked customer here denotes the response that the blocked customer has to being queued or discarded, while the server release mechanism specifies whether servers allocated to the same customer end service asynchronously or synchronously. The approach to analysis is to solve the state equations numerically if the buffer is finite, or to find the generating functions or a matrix-geometric solution if the buffer is infinite.

In the application of telecommunications, the broadband packet-switched network is a recent example. Owing to the progress of technologies such as microelectronics, photoelectronics and communications protocols, the broadband network is developing rapidly. It has large channel capacity in all its elements, such as remote multiplexing node (RMN), access node (AN), local exchange node (LEN) and transit exchange node (TEN) [12], and has wideband (WB) customers which require more channels and narrowband (NB) customers which need fewer channels. In the existing broadband network, the NB and WB traffics are usually, for example, data and slow-motion video, respectively. The slow-motion video applications are videophone and teleconferencing. The blocking probability and delay requirements for these two traffics are quite different. The WB traffic needs real-time delivery but is able to tolerate a certain degree of packet loss due to blocking, while the NB traffic is delayable but is unable to put up with packet loss. In considering the assignment of channel capacity, the objective is to make the blocking probability of WB and the delay of NB as small as possible. And in order to handle the information traffics efficiently, an optimal capacity assignment strategy for the network is thus required [1–4, 8–10]. The reasons for this are highlighted briefly in [1]. The authors studied a model of broadband channel carrying a mixture of NB and WB traffics in which WB traffic is nonqueueable while NB traffic may wait in an infinite-sized buffer. Two access control strategies, namely pre-emptive priority (PP) and bit-rate compression (BC), were analysed by the moment generating functions approach. This is a moveable boundary problem. By comparing the two strategies it is shown that the PP strategy is not suitable when WB traffic is high since no bit-rate compression is provided, and the BC strategy is not beneficial when NB traffic is high since NB customers are not allowed to use free WB bandwidth. A combined access strategy according to the offered load is thus recommended.

However, as stated in [1], this combined adaptive access scenario, needing a complicated load estimator, is not feasible. And, in the PP strategy, the priority pre-emption of NB seems impractical for implementation and is inappropriate for the integrity of packets; in the BC strategy, the prohibitive usage of the free WB bandwidth is inefficient for heavy NB traffic. Hence, in this paper, a new capacity assignment strategy is proposed and studied. In order to satisfy the characteristics and requirements of WB traffic, this strategy favours the WB customer. It permanently allocates a *reserved number of channels* for WB customers which are not allowed to be used by NB customers, and sets a certain *constraint bound* on the NB customers by which they cannot use the WB channels until the number of NB customers in the waiting queue exceeds this bound. In addition, the WB customer will be compressed as an NB customer and will use the NB channel when, generally speaking, the WB channels are full and some NB channels are free. This bit-rate compression is considered here since it is still a realistic assumption if two kinds of WB bit-rate coders

Paper 8196I, received 5th April 1990

The authors are with the Institute of Communication Engineering and Center for Telecommunications Research, National Chiao Tung University, Hsinchu, Taiwan 30039, Republic of China

are provided. The details of this strategy will be described in the next section. As such, a larger permanently reserved amount m of channels for WB, and a larger constraint bound c for NB, will result in the blocking probability of WB being too low but the delay of NB being too high. On the other hand, smaller m and c will result in the blocking probability of WB being too high but the delay of NB being too low. Consequently, an optimal pair of m and c exists such that the system performance is superior at all offered load conditions. Furthermore, this strategy can be implemented without the complicated load estimator and adaptive switchover which would cause a disturbance in system operation and a decline in system performance. It also avoids the priority pre-emption of NB customers. The results show that this scheme, via proper choice of design parameters, can improve the system performance more than the recommended combined adaptive access control strategy in [1] under all traffic conditions.

2 Model

Consider a broadband network element in which there are devices with a capacity of r basic channels. The channel capacity can be the bandwidth of the multiplexer or the input port of the switching network. As shown in Fig. 1, two types of customer are offered. The NB customer, termed the type 1 customer, requests one basic channel, and operates with delay. Its queueing buffer is assumed to be infinite. The WB customer, termed the type 2 customer, requests b basic channels, $1 < b \leqslant r$, which are seized and released simultaneously. These aggregated b basic channels form one WB channel. The WB customer is assumed to operate with loss. These two kinds of traffic are originated independently by a Poisson process with mean rate λ_i; and their service times are exponentially distributed with mean $1/\mu_i$, where $i = 1, 2$ denotes the type of traffic. The number of basic channels among the r channel capacities assigned to NB traffic is r_1 and that to WB traffic is $r_2 b$. Clearly $r = r_1 + r_2 b$. For r_1 NB channels, the NB traffic has nonpre-emptive priority; for r_2 WB channels, there are m WB channels permanently reserved for WB traffic and $(r_2 - m)b$ basic channels for which the WB traffic has nonpre-emptive priority. When the basic channels used by the NB traffic total r_1, the NB queue length is longer than c and there is a free basic channel (this channel must belong to WB traffic) available, the NB customer at the head of the queue can enter the idle WB bandwidth to receive service. In this case, the number of NB customers allowed to enter the WB bandwidth cannot make the remaining number of customers waiting in the queue less than c; and $(r_2 - m)b$ basic WB channels is the maximum number that can be used by NB customers. On the other hand, when the basic channels used by WB traffic reach $r_2 b$ and there is a free basic channel (this channel must belong to NB traffic) available, the WB customer is assumed to be compressed as an NB customer. The required bandwidth is one basic channel, and the mean service time is $1/\mu_1$. The number of channels mb permanently reserved for WB traffic (or say m WB channels), and the constraint bound c over which the NB traffic is then allowed to use free WB channels, are two design parameters in the proposed capacity assignment for the broadband network.

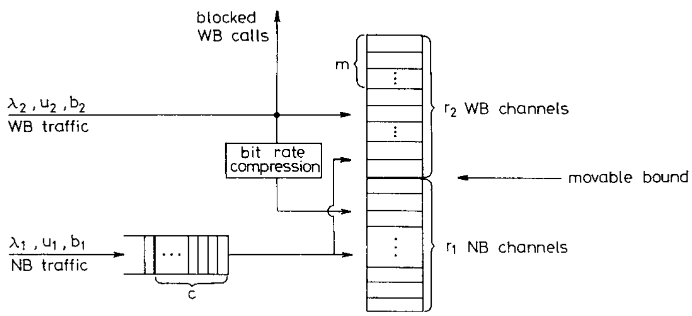

Fig. 1 *Proposed bandwidth allocation strategy*

3 Analysis

Let the system state at time t be $(n_1(t), n_2(t))$, where $n_i(t)$, $i = 1, 2$, represents the number of type i customers present in the system at time t. Because we have assumed that the input process is Poisson and the service time is exponentially distributed, the process $(n_1(t), n_2(t))$ forms a two-dimensional Markov chain with joint probabilities

$$\pi_{k,j} = \lim_{t \to \infty} P_r[n_1(t) = k, n_2(t) = j]$$

The state transition equations, at steady state, can easily be obtained by considering that the flow out of a given state equals the flow into it. In the derivation, the state spaces are categorised into four regions as shown in Fig. 2. Their state transition equations are derived in the following.

3.1 Region 1: $\{(n_1, n_2) \mid 0 \leqslant n_1 < r_1 + c, 0 \leqslant n_2 \leqslant r_2\}$

3.1.1 $0 \leqslant n_1 < r_1 + c$, $0 \leqslant n_2 < r_2$: Type 1 and type 2 customers will use bandwidth resources of their own because the NB queue length is always shorter than c and there is WB channel available. Thus

$$\begin{aligned}
[\lambda_1 + \lambda_2 + \min(n_1, r_1)\mu_1 + n_2\mu_2]\pi_{n_1,n_2} &= \lambda_1\pi_{n_1-1,n_2} + \lambda_2\pi_{n_1,n_2-1} \\
&\quad + \min(n_1+1, r_1)\mu_1\pi_{n_1+1,n_2} \\
&\quad + (n_2+1)\mu_2\pi_{n_1,n_2-1} \qquad (1)
\end{aligned}$$

3.1.2 $r_1 \leqslant n_1 < r_1 + c$, $n_2 = r_2$: The arriving WB customer will be blocked since no empty WB and NB channels are available. Thus

$$[\lambda_1 + r_1\mu_1 + r_2\mu_2]\pi_{n_1, r_2} = \lambda_1\pi_{n_1-1, r_2} + \lambda_2\delta_{n_1, r_1}\pi_{r_1-1, r_2} + \lambda_2\pi_{n_1, r_2-1} + r_1\mu_1\pi_{n_1+1, r_2} \quad (2)$$

where δ_{ij} is the Kronecker function.

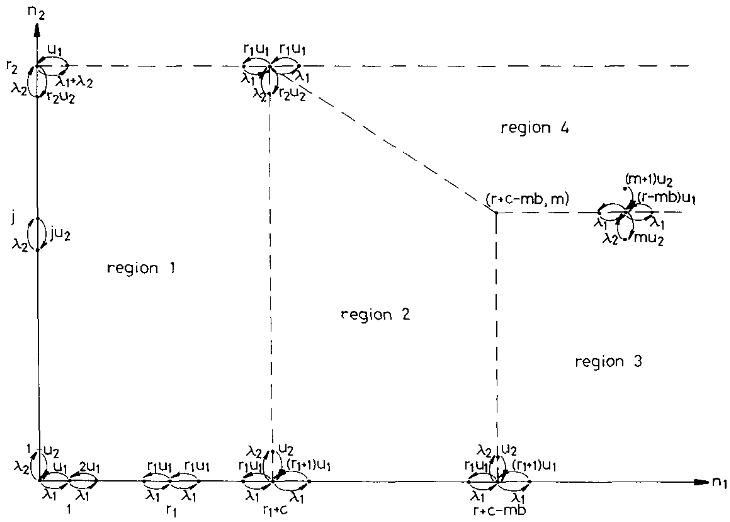

Fig. 2 *State transition diagram*

3.1.3 $0 \leqslant n_1 < r_1$, $n_2 = r_2$: Data compression can be performed on WB customers. At this time, the compressed WB customer will be regarded as an NB, and the number of NB customers is increased by one. Then

$$[\lambda_1 + \lambda_2 + n_1\mu_1 + r_2\mu_2]\pi_{n_1, r_2} = \lambda_1\pi_{n_1-1, r_2} + \lambda_2[\pi_{n_1-1, r_2} + \pi_{n_1, r_2-1}] + (n_1 + 1)\mu_1\pi_{n_1+1, r_2} \quad (3)$$

3.2 Region 2: $\{(n_1, n_2) \mid r_1 + c \leqslant n_1 < r + c - mb,\ 0 \leqslant n_2 b < r + c - n_1\}$

In this region, the queue length of NB customers is always maintained at c since the remaining WB bandwidths, $r_2 b - n_2 b$, are greater than the number of those NB customers, $n_1 - r_1 - c$, that are allowed to use the WB bandwidth. The number of NB customers that can accept service is $n_1 - c$ rather than $r - n_2 b$.

3.2.1 $r_1 + c \leqslant n_1 < r + c - mb$, $0 \leqslant n_2 b \leqslant r + c - n_1 - b$: There is at least a WB channel available. Thus an arriving WB will get service. Then

$$[\lambda_1 + \lambda_2 + (n_1 - c)\mu_1 + n_2\mu_2]\pi_{n_1, n_2} = \lambda_1\pi_{n_1-1, n_2} + \lambda_2\pi_{n_1, n_2-1} + (n_1 - c + 1)\mu_1\pi_{n_1+1, n_2} + (n_2 + 1)\mu_2\pi_{n_1, n_2+1} \quad (4)$$

3.2.2 $r_1 + c < n_1 < r + c - mb$, $r + c - n_1 - b < n_2 b \leqslant r + c - n_1$: The remaining WB bandwidths are greater than zero but less than b basic channels. Therefore they are available for arriving NB customers but not WB customers. Thus we have

$$[\lambda_1 + (n_1 - c)\mu_1 + n_2\mu_2]\pi_{n_1, n_2} = \lambda_1\pi_{n_1-1, n_2} + \lambda_2\pi_{n_1, n_2-1} + [n_1 - c + (1 - \delta_{n_1, r+c-n_2 b})]\mu_1\pi_{n_1+1, n_2} + (n_2 + 1)\mu_2\pi_{n_1, n_2+1} \quad (5)$$

3.3 Region 3: $\{(n_1, n_2) \mid n_1 \geqslant r + c - mb,\ 0 \leqslant n_2 \leqslant m\}$

In this region, all bandwidth resources except m WB channels are fully used by NB customers. An arriving WB customer will be accepted if n_2 is not equal to m. Thus

$$[\lambda_1 + \lambda_2(1 - \delta_{n_2, m}) + (r - mb)\mu_1 + n_2\mu_2]\pi_{n_1, n_2} = \lambda_1\pi_{n_1-1, n_2} + \lambda_2\pi_{n_1, n_2-1} + (r - mb)\mu_1\pi_{n_1+1, n_2} + (n_2 + 1)\mu_2\pi_{n_1, n_2+1} \quad (6)$$

3.4 Region 4: $\{(n_1, n_2) \mid n_1 > r - n_2 b + c,\ m < n_2 \leqslant r_2\}$

All channels are used up, so the arriving WB customer will be blocked and an arriving NB customer will be queued. Then

$$[\lambda_1 + (r - n_2 b)\mu_1 + n_2\mu_2]\pi_{n_1, n_2} = \lambda_1\pi_{n_1-1, n_2} + (r - n_2 b)\mu_1\pi_{n_1+1, n_2} + (n_2 + 1)\mu_2\pi_{n_1, n_2+1} \quad (7)$$

To solve these equations, the moment generating function approach is used [13, 14]. Define the generating functions of π_{n_1, n_2} as

$$G_{n_2}(z) = \sum_{n_1=M}^{\infty} \pi_{n_1, n_2} z^{n_1 - M} \quad n_2 = 0, 1, \ldots, r_2 \quad (8)$$

where $M = r + c - mb$. Applying eqn. 8 to eqns. 6 and 7, we get the following equations respectively:

$$[\lambda_1 z(1 - z) + (1 - \delta_{n_2, m})\lambda_2 z - (r - mb)\mu_1(1 - z) + n_2\mu_2 z]G_{n_2}(z) = \lambda_1 z\pi_{M-1, n_2} + \lambda_2 z G_{n_2-1}(z) - (r - mb)\mu_1\pi_{M, n_2} + (n_2 + 1)\mu_2 z G_{n_2+1}(z) \quad \text{if } 0 \leqslant n_2 \leqslant m \quad (9)$$

$$[\lambda_1 z(1-z) - (r - n_2 b)\mu_1(1-z) + n_2 \mu_2 z]G_{n_2}(z)$$
$$= \lambda_1 z\pi_{M-1,n_2} - (r - n_2 b)\mu_1 \pi_{M,n_2}$$
$$+ (n_2+1)\mu_2 z G_{n_2+1}(z) \quad \text{if } m+1 \leqslant n_2 \leqslant r_2 \quad (10)$$

The $G_{r_2+1}(z) = 0$. In order to make these equations easier to handle, we rewrite them as

$$a_{n_2}(z)G_{n_2}(z) = \lambda_1 z\pi_{M-1,n_2} + \lambda_2 z G_{n_2-1}(z) - (r - mb)\mu_1\pi_{M,n_2} + (n_2+1)\mu_2 z G_{n_2+1}(z) \quad \text{if } 0 \leqslant n_2 \leqslant m \quad (11)$$

$$a_{n_2}(z)G_{n_2}(z) = \lambda_1 z\pi_{M-1,n_2} - (r - n_2 b)\mu_1\pi_{M,n_2} + (n_2+1)\mu_2 z G_{n_2+1}(z) \quad \text{if } m+1 \leqslant n_2 \leqslant r_2 \quad (12)$$

where

$$a_{n_2}(z) = \begin{cases} \lambda_1 z(1-z) + (1-\delta_{n_2,m})\lambda_2 z - (r-mb)\mu_1(1-z) + n_2\mu_2 z & \text{if } 0 \leqslant n_2 \leqslant m \\ \lambda_1 z(1-z) - (r - n_2 b) \times \mu_1(1-z) + n_2\mu_2 z & \text{if } m+1 \leqslant n_2 \leqslant r_2 \end{cases} \quad (13)$$

Eqns. 11 and 12 form a set of $(r_2 + 1)$ linear equations in $G_0(z)$, $G_1(z)$, ..., $G_{r_2}(z)$. In Appendix 7.1 it is shown that the solution of these linear equations can be given by

$$G_j(z) = e_j(z)G_0(z) + f_j(z) \quad \text{if } j = 1, 2, \ldots, r_2 \quad (14a)$$

$$G_0(z) = \frac{-r_1\mu_1\pi_{M,r_2} + \lambda_1 z\pi_{M-1,r_2} - a_{r_2}(z)f_{r_2}(z)}{a_{r_2}(z)e_{r_2}(z)} \quad (14b)$$

where $e_{r_2}(z)$ and $f_{r_2}(z)$ are known polynomials defined in Appendix 7.1. There are unknown probabilities π_{M-1,r_2} and π_{M,r_2} in $G_0(Z)$. They can be solved by using eqns. 1–7. It is found that there are $(r + c - mb)(r_2 + 1)$ equations but $(r + c - mb + 1)(r_2 + 1)$ unknown probabilities which are $\pi_{0,0}, \pi_{0,1}, \ldots, \pi_{0,r_2}, \pi_{1,0}, \ldots, \pi_{r+c-mb,r_2}$ in eqns. 1–7. Thus $r_2 + 1$ additional equations are needed. We can prove that the denominator of $G_0(z)$ has r_2 zeros in the range (0, 1). This is shown in Appendix 7.2. Hence there are r_2 equations which can be obtained by equating the numerator of $G_0(z)$ to zero at each root z_i, $i = 1, 2, \ldots, r_2$, of its denominator. Also, the summability-to-one criterion

$$\sum_{n_1=0}^{M-1}\sum_{n_2=0}^{r_2}\pi_{n_1,n_2} + \sum_{n_2=0}^{r_2} G_{n_2}(1) = 1$$

provides one further boundary condition. Once π_{n_1,n_2} is obtained, the performance measures can be readily evaluated. The blocking probability of a type 2 customer, PB_2, is

$$PB_2 = \left[\sum_{n_1=r_1}^{\infty}\pi_{n_1,r_2} + \sum_{n_2=m}^{r_2-1}\sum_{n_1=\xi}^{\infty}\pi_{n_1,n_2}\right] \quad (15a)$$

and is computed as

$$PB_2 = \left[\sum_{n_2=m}^{r_2} G_{n_2}(1) + \sum_{n_1=r_1}^{M-1}\pi_{n_1,r_2} + \sum_{n_2=m}^{r_2-1}\sum_{n_1=\xi}^{M-1}\pi_{n_1,n_2}\right] \quad (15b)$$

where $\xi = r - n_2 b - b + c + 1$. It should be noted that the sample space of PB_2 is exclusive of compressed component.

The average time delay of type 1 traffic can be obtained by using Little's formula

$$E(w_1) = \frac{E(n_1)}{\lambda_{eff}} \quad (16)$$

The $E(n_1)$ is the average number of type 1 packets consisting of the original type 1 and the compressed type 2 packets, and λ_{eff} is the effective arrival rate of a packet to group 1, including both type 1 and the compressed type 2. It can be easily seen that

$$E(n_1) = \sum_{n_2=0}^{r_2}\sum_{n_1=0}^{M-1} n_1\pi_{n_1,n_2} + \sum_{n_2=0}^{r_2}[G'_{n_2}(1) + MG_{n_2}(1)] \quad (17)$$

$$\lambda_{eff} = \lambda_1 + \lambda_2\left(\sum_{n_1=0}^{r_1-1}\pi_{n_1,r_2}\right) \quad (18)$$

4 Numerical examples

The objective of channel capacity assignment for the broadband network is to make both the average waiting time of NB traffic and the blocking probability of WB traffic as small as possible. Therefore we first use the combined performance measure $P = (1 - PB_2)/\mu_1 E(w_1)$ as the criterion for system performance. Maximisation of the measure of effectiveness P amounts to minimising the blocking probability of WB traffic and the delay of NB traffic. P is illustrated versus the varied type 1 traffic load λ_1 at the fixed type 2 traffic load λ_2. We have the following example systems with $r = 12$, $b = 2$, $r_1 = 4$, $r_2 = 4$, $\mu_1 = \mu_2 = 1$ and $\lambda_2 = 3$.

As discussed in Section 1, m and c are used to protect the WB customer from being blocked. If m and c are relatively large, the bandwidth resources of the system may be wasted if WB traffic is low. This will result in poor system utilisation and a long mean waiting time for NB. But if m and c are made smaller, the service quality of WB deteriorates. Thus we wish to find an optimal m and c to maximise the combined performance measure. The combined measure P versus the NB arrival rate λ_1 for different pairs of m and c is shown in Fig. 3. We observe that the optimal pair of m and c exists for various loads. When λ_1 is small, the blocking probability of type 2 dominates the performance measure P. Thus the optimal m is larger. But as λ_1 increases, the queueing delay of type 1 increases. A larger m would restrict NB customers from using the WB channel, thereby make the performance worse. This is why the curve $m = 3$, $c = 1$ falls so steeply at high NB traffic. This reasoning can also apply to the BC strategy. In fact, BC is the special case with $m = 4$ in our example. On the other hand, a smaller m can give a better utilisation of the channel, thus reducing the delay. Here we choose m and c in such a way that the P is optimal over all load conditions. Thus in Fig. 4 the curve of $m = 2$ and $c = 1$ is selected to compare with the BC and PP strategies. It can be found that the performance is improved at all traffic loads shown.

However, with respect to the blocking probability of WB and the mean waiting time of NB, we consider that the improvement for the former is more significant than that for the latter. The performance measure P described above, where equal weight is given to the blocking probability of WB and the mean waiting time of NB, is not a satisfactory criterion for system performance. Thus Figs. 5 and 6 show the blocking probability of WB and the average waiting time of NB with different m and c, respectively. The PP and BC strategies are also shown for comparison. These figures clearly indicate that many pairs of m and c exist which have almost the same mean waiting time of NB but still have a superior blocking probability of WB compared with the PP strategy; or they can have almost the same blocking probability of WB but still have superior mean waiting time compared

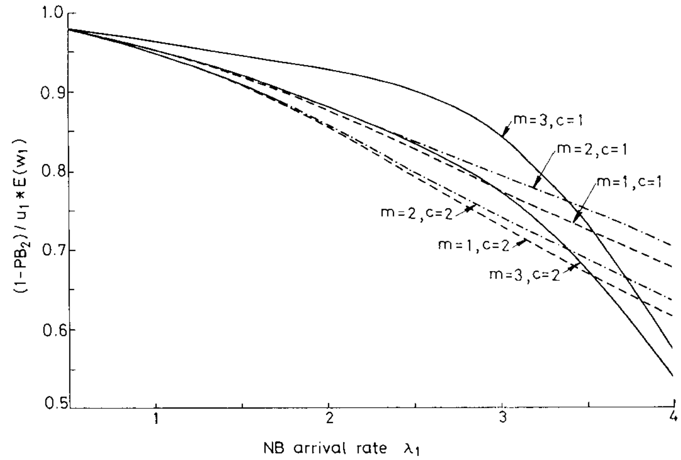

Fig. 3 *P versus λ_1 for various m and c, given $\lambda_2 = 3.0$*

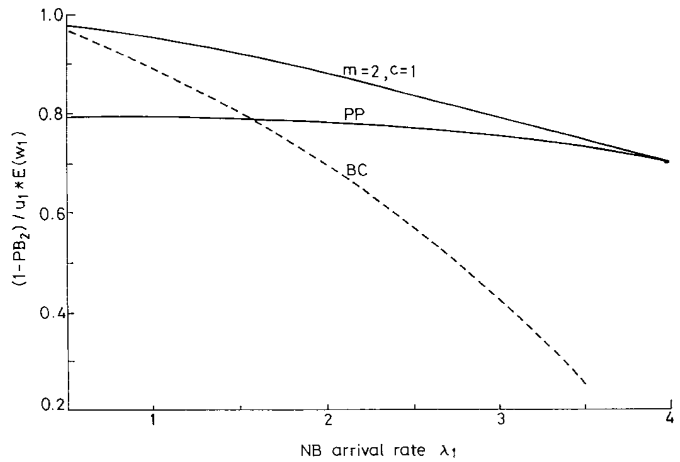

Fig. 4 *P versus λ_1 for optimal m and c, given $\lambda_2 = 3.0$*

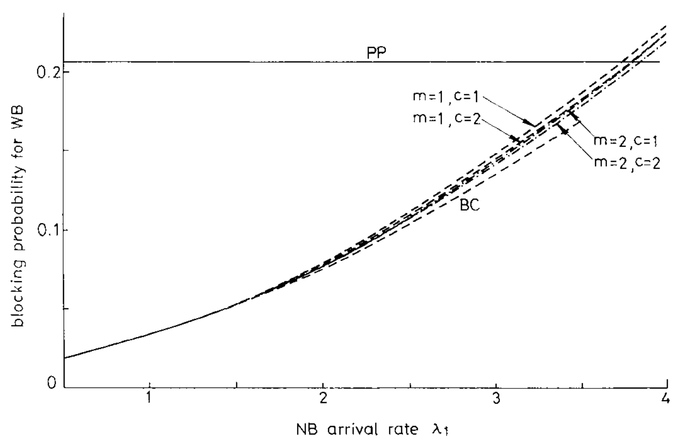

Fig. 5 *Blocking probability of WB traffic versus λ_1 for various m and c, given $\lambda_2 = 3.0$*

with the BC strategy. We can also observe that, as m is fixed, the blocking probability of WB for $c = 2$ is reduced by about 2.08% compared with that for $c = 1$ at $\lambda_1 = 3.5$, $\lambda_2 = 3.0$; this is a significant improvement. The mean

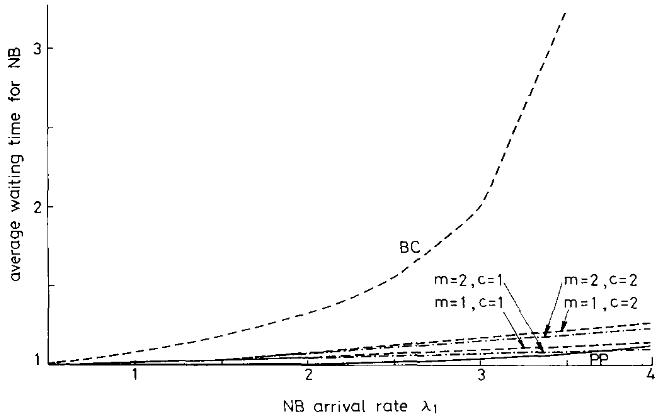

Fig. 6 *Mean waiting time of NB traffic versus λ_1 for various m and c, given $\lambda_2 = 3.0$*

waiting time for NB is raised about 9.84%, which means 0.107 unit time of increment; this is a trivial improvement. Thus the parameter c makes the system more satisfactory. Hence we believe that the strategy presented here is more promising than the PP or BC strategies.

5 Conclusions

A capacity assignment strategy is proposed which makes more efficient use of channels. We analyse the performance of this strategy by using the moment generating function approach. With optimal choice of design parameters m and c in this capacity assignment, the performance criterion P is higher than with the PP and BC strategies at all offered load conditions. With regard to the blocking probability of WB and the mean waiting time of NB, respectively, there are many pairs of m and c which have superior performance. Furthermore, this strategy avoids using the complicated load estimator, and also prevents NB customers from inappropriate priority pre-emption by WB customers.

6 References

1 KRAIMECHE, B., and SCHWARTZ, M.: 'Analysis of traffic access control strategies in integrated service networks', *IEEE Trans.*, 1985, **COM-33**, pp. 1085–1093

2 KRAIMECHE, B., and SCHWARTZ, M.: 'Bandwidth allocation strategies in wide-band integrated network', *IEEE J, Sel. Areas Commun.*, 1986, **SAC-6**, pp. 869–878

3 KRAIMECHE, B., and SCHWARTZ, M.: 'Traffic access control strategies in integrated digital network'. Proceedings of IEEE INFOCOM'84, San Diego, April 1984, pp. 230–235

4 GIMPELSON, L.A.: 'Analysis of mixtures of wide- and narrow-band traffic', *IEEE Trans.*, 1965, **COM-13**, (3), pp. 258–266

5 GREEN, L.: 'A queueing system in which customers require a random number of servers', *Oper. Res.*, 1980, **28**, (6), pp. 1335–1346

6 GREEN, L.: 'Comparing operating characteristics of queueing in which customers require a random number of servers', *Manage. Sci.*, 1981, **27**, (1), pp. 65–74

7 GREEN, L.: 'A multiple dispatch queueing model of police patrol operations', *Manage. Sci.*, 1984, **30**, (6), pp. 653–664

8 KAUFMAN, J.S.: 'Blocking in a shared resource environment', *IEEE Trans.*, 1981, **COM-29**, (10), pp. 1474–1481

9 SERRES, Y.D., and MASON, L.G.: 'A multiserver queue with narrow- and wide-band customers and wide-band restricted access', *IEEE Trans.*, 1988, **COM-36**, pp. 675–684

10 LIAO, K.Q., and MASON, L.G.: 'An approximate performance model for a multislot integrated services system', *IEEE Trans.*, 1989, **COM-37**, (3), pp. 211–221

11 CHANG, C.J., and CHU, Y.J.: 'Design considerations for bandwidth allocation in broadband network'. Technical report, Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, October 1989

12 T1S1.5/89-001 'T1S1 technical sub-committee broadband aspect of ISDN'. Baseline document (draft), November 1989

13 GROSS, D., and HARRIS, C.M.: 'Fundamental of queueing theory' (Wiley, 1974)

14 DICKSON, L.E.: 'First course in the theory of equation' (Wiley, 1926)

7 Appendix

7.1 Derivation of eqn. 14

We can write eqns. 11 and 12 in matrix form as

$$\boldsymbol{M}(z)\boldsymbol{G}(z) = \boldsymbol{C}(z)G_0(z) + \boldsymbol{b}(z) \tag{19}$$

where

$$\boldsymbol{M}(z) = \begin{bmatrix} -\mu_2 z & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ a_1(z) & -2\mu_2 z & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ -\lambda_2 z & a_2(z) & -3\mu_2 z & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_2 z & a_m(z) & -(m+1)\mu_2 z & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & a_{m+1}(z) & -(m+2)\mu_2 z & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -r_2\mu_2 z \end{bmatrix}$$

$$\boldsymbol{G}(z) = [G_1(z), G_2(z), \ldots, G_{r_2}(z)]^T$$

$$\boldsymbol{C}(z) = [-a_0(z), \lambda_2 z, 0, \ldots, 0]^T$$

$$
b(z) = \begin{bmatrix} -(r - mb)\mu_2 \pi_{M,0} + \lambda_1 z \pi_{M-1,0} \\ -(r - mb)\mu_2 \pi_{M,1} + \lambda_1 z \pi_{M-1,1} \\ \vdots \\ -[r - (m+1)b]\mu_1 \pi_{M,m+1} + \lambda_1 z \pi_{M-1,m+1} \\ \vdots \\ -[r - (r_2 - 1)b]\mu_1 \pi_{M,r_2-1} + \lambda_1 z \pi_{M-1,r_2-1} \end{bmatrix}
$$

Applying Cramer's rule [14] to eqn. 19 we get

$$
\begin{aligned}
G_j(z) &= \frac{\det [m_c^j(z)]}{\det [M(z)]} G_0(z) + \frac{\det [m_b^j(z)]}{\det [M(z)]} \\
&= e_j(z) G_0(z) + f_j(z), \quad j = 1, 2, \ldots, r_2 \qquad (20)
\end{aligned}
$$

where $\det [M(z)] = (-1)^{r_2} r_2! (\mu_2 z)^{r_2}$; $m_c^j(z)$, $m_b^j(z)$ are matrix $M(z)$ with its jth column replaced by $C(z)$, $b(z)$, respectively; and $e_j(z) = \det [m_c^j(z)]/\det [M(z)]$, $f_j(z) = \det [m_b^j(z)]/\det [M(z)]$. Taking $G_{r_2}(z)$ into the last equation ($n_2 = r_2$) of eqn. 12, we have

$$
\begin{aligned}
& a_{r_2}(z)[e_{r_2}(z) G_0(z) + f_{r_2}(z)] \\
&\quad = -r_1 \mu_1 \pi_{M,r_2} + \lambda_1 z \pi_{M-1,r_2} \qquad (21)
\end{aligned}
$$

Thus expression 14 is derived.

7.2 *Proof of the denominator of $G_0(z)$ having r_2 zeros in the range (0, 1)*

From eqn. 14*b*, the denominator of $G_0(z)$ is given by

$$
\begin{aligned}
a_{r_2}(z) e_{r_2}(z) &= [\lambda_1 z(1 - z) - r_1 \mu_1 (1 - z) + r_2 \mu_2 z] \\
&\quad \times \left[\frac{\det [m_c^{r_2}(z)]}{(-1)^{r_2} r_2! (\mu_2 z)^{r_2}} \right] \qquad (22)
\end{aligned}
$$

Because we now consider its zeros only, the term $(-1)^{r_2} r_2! (\mu_2 z)^{r_2}$ can be ignored without affecting the result. Thus we have

$$
\begin{aligned}
& [\lambda_1 z(1 - z) - r_1 \mu_1 (1 - z) + r_2 \mu_2 z][\det m_c^{r_2}(z)] \\
&\quad = a_{r_2}(z)[\det m_c^{r_2}(z)] \qquad (23)
\end{aligned}
$$

Define the determinants

$$
\rho_i = \det \begin{bmatrix}
a_0(z) & -\mu_2 z & 0 & \cdots & 0 & 0 & 0 & \cdots & & 0 \\
-\lambda_2 z & a_1(z) & -2\mu_2 z & \cdots & 0 & 0 & 0 & \cdots & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \vdots \\
0 & 0 & 0 & \cdots & -\lambda_2 z & a_m(z) & -(m+1)\mu_2 z & 0 & 0 \ \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & a_{m+1}(z) & -(m+2)\mu_2 z & 0 \ \cdots & 0 \\
\vdots & & & & & & & & & \vdots \\
0 & \cdots & & & & & & & & a_i(z)
\end{bmatrix}
$$

where $0 \leqslant i \leqslant r_2$. These determinants have the following relations:

$$
\rho_i = \begin{cases} a_i(z)\rho_{i-1}(z) - i\lambda_2 \mu_2 z^2 \rho_{i-2}(z) & \text{if } 0 \leqslant i \leqslant m \\ a_i(z)\rho_{i-1}(z) & \text{if } m + 1 \leqslant i \leqslant r_2 \end{cases} \qquad (24)
$$

with $\rho_{-1}(z) = 1$ and $\rho_{-2}(z) = 0$. Also define $M_c^i(z)$ as the matrix formed by taking the first $i \times i$ elements of $M(z)$ with the last column replaced by $C(z)$. The $\det [M_c^i(z)]$ has the following characteristics:

$$
\det [M_c^i(z)] = (-1)^i \rho_{i-1}(z) \qquad (25)
$$

$$
\det [M_c^{r_2}(z)] = \det [m_c^{r_2}(z)] \qquad (26)
$$

The sequence of the polynomial $\rho_0, \rho_1, \ldots, \rho_{m-1}$ forms a sturm sequence [14], and ρ_m can be proved to have m distinct roots in the interval (0, 1) [1, 11]. Furthermore, because $a_i(z) = \lambda_1 z(1 - z) - (r - ib)\mu_1(1 - z) + i\mu_2 z$ for $i \geqslant m + 1$, we can find $a_i(0) = -(r - \mu_2 b)\mu_1 < 0$ and $a_i(1) = i\mu_2 > 0$. Thus there is exactly one root of $a_i(z)$ for $i > m$ in the range (0, 1). This implies that

$\rho_{m+1} = a_{m+1}(z)\rho_m$ has $m + 1$ roots in the range (0, 1)

$\rho_{m+2} = a_{m+2}(z)\rho_{m+1}$ has $m + 2$ roots in the range (0, 1)

$\vdots$

$\rho_{r_2} = a_{r_2}(z)\rho_{r_2-1}$ has r_2 roots in the range (0, 1)

Then eqn. 23 has r_2 roots in the range (0, 1).