Analysis of Shared Finite Storage in a Computer Network Node Environment Under General Traffic Conditions

FAROUK KAMOUN, MEMBER, IEEE, AND LEONARD KLEINROCK, FELLOW, IEEE

Abstract—**Nodal storage limitations in a store and forward computer network lead to blocking; this results in degradation of network performance due to the loss or retransmission of blocked messages. In this paper, we consider several schemes for sharing a pool of buffers among a set of communication channels emanating from a given node in a network environment so as to make effective use of storage in a variety of applications. Five sharing schemes are examined, analyzed, and displayed in a fashion which permits one to establish the tradeoffs among blocking probability, utilization, throughput, and delay. The key to the analysis lies in the observation that the equilibrium joint probability distribution of the buffer occupancy obeys the well-known product form solution for networks of queues. The study indicates advantages and pitfalls of each of the sharing schemes. We observe, in general, that sharing with appropriate restrictions on the contention for space is very much desirable.**

I. INTRODUCTION

QUEUEING models for computer networks often assume infinite storage at the switching nodes. Such an assumption is questionable, especially in view of the storage capacity issues which have been observed in the ARPANET. Furthermore, storage becomes critical in the context of large computer networks [1]-[3]. As a result, a storage constraint must be introduced in realistic network models. This we do for a single node in this paper. The application of the results derived here to the performance analysis of a class of symmetrical networks can be found in [1] and [3].

In store-and-forward (S/F) computer nets, the outgoing channels of a node share a certain number (say B) of buffers (S/F) buffers). If no feedback is considered (i.e., no retransmission of rejected messages), the S/F function of a node may be modeled as a set of $M/M/1$ queueing systems (one for each channel) which share a finite waiting room[1] under some scheme [1]. The purpose of this paper is to analyze and compare a few existing and/or intuitive storage sharing schemes. The first (and simplest) is the *complete partitioning (CP)* scheme where actually no sharing is provided, but where the entire finite storage (waiting room) is permanently partitioned among the (say) R servers. At the other extreme is the second scheme, *complete sharing (CS)*, which is such that an arriving customer is accepted if any storage space is available, independent of the server to which it is directed. We find that CS succeeds in achieving a better performance (smaller probability of blocking) than CP under normal traffic conditions and for fairly balanced input systems. However, for highly asymmetrical message input rates (λ_i $i = 1, \cdots, R$) and equal service rates, CS tends to heavily favor servers with higher input rates, even though they may be close to saturation (input rate close to service rate). The failure to recognize servers at or near saturation results in most of the space being occupied by customers waiting for those servers, to the detriment of the others. Moreover, even with perfectly balanced arrival rates (i.e., $\lambda_i = \lambda$, $i = 1, \cdots, R$), under overload conditions, CS fails (where CP succeeds) in securing a full utilization of all the R servers. The above considerations suggest that contention for space must be limited in some way. In order to avoid the possible utilization of the entire space by any particular output channel, we impose a limit on the number of buffers to be allocated at any time to any server. This idea is incorporated in our third scheme: *sharing with maximum queue lengths (SMXQ)*. Of course, the sum of those maxima must be greater than the total space if some sharing is to be provided. SMXQ still does not guarantee a full utilization of the servers under heavy traffic conditions. This deficiency motivates the fourth scheme: *sharing with a minimum allocation (SMA)* scheme. With SMA, a minimum number of buffers is always reserved for each server and, in addition, a common pool of buffers is to be shared among all servers, with no further constraints on the queue size. With SMA, the shared area tends to be unfairly utilized as mentioned earlier; hence, we have the fifth and final scheme: *sharing with a maximum queue and minimum allocation (SMQMA)*. A schematic representation of the first four schemes is given in Fig. 1, as well the constraint set for $R = 2$ servers. The constraint set shows the feasible regions for the four schemes in an $n_1 \times n_2$ plane where n_i = number of buffers available to server i. Note that the feasible set is the largest for the CS scheme where all points ○, □, △ are included and the smallest for the CP scheme where only the points ○ are included. In between, we find the feasible sets for the SMA and SMXQ schemes; these two schemes are equivalent when $R = 2$.

Rich and Schwartz [4] studied a scheme very similar to SMA except that the *entire* common storage is dynamically allocated to one server at a time. Drukey [5] analyzed the CS scheme with the assumption that all the ρ_i's are equal; for the general case of different ρ_i's, he restricts his study to two channels.

Paper approved by the Editor for Computer Communication of the IEEE Communications Society for publication without oral presentation. Manuscript received October 10, 1976; revised October 22, 1979. This work was supported by the Advanced Research Projects Agency of the Department of Defense under Contract MDA 903-77-C-0272.

F. Kamoun is with the Department of Informatique, Faculte des Sciences, Tunis, Tunisia.

L. Kleinrock is with the Department of Computer Science, University of California, Los Angeles, CA 90024.

[1] The waiting room accounts for all the switching node buffers, including those occupied by messages in transmission (i.e., in service).

0090-6778/80/0700-0992$00.75

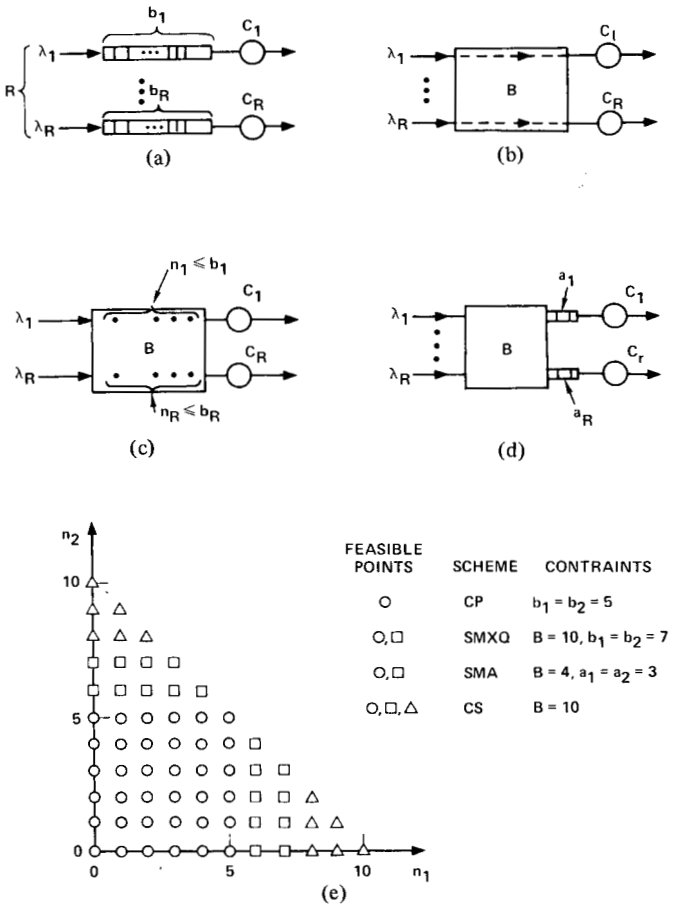


Fig. 1. Storage sharing schemes: (a) CP; (b) CS; (c) SMXQ; (d) SMA; and (e) set of constraints, $R = 2$.

Irland [6] studied the SMXQ scheme. He obtained a z-transform expression for the constant term in the expression of the joint queue length distribution. An explicit expression for that term was derived only for the special case of two servers. With the two-server environment, he also performs a numerical search for the optimal value of the constraint on the queue length. Lam [7] also tackled the storage constraint problem for both a single node and a network environment. His model assumes a complete sharing (CS) scheme and, moreover, it accounts for more nodal functions such as time-out, acknowledgment, and retransmission. These added features make the model and the results fairly different. However, no comparisons with other sharing schemes were attempted.

Problems of this sort are frequently encountered in telephony and are referred to as "graded" systems [8]. The main interest there, however, is in sharing (extra) lines as opposed to storage. In this paper, we intend to characterize the five storage schemes under steady-state conditions; namely, we derive expressions for the probabilities of blocking, the average time in system, and the throughput. A comparison of the sharing schemes is also provided. The key to the analysis lies in the observation that the equilibrium joint probability distribution for the buffer occupancy obeys the well-known product form solution for networks of queues (see [9]-[13] and the bibliographies therein). The results of the analysis are presented and displayed in a fashion which permits one to establish the tradeoffs among blocking probability, utilization, throughput, and delay. We conclude that no one scheme is always optimal. The selection of a specific scheme depends upon the particular operational environment. This study establishes the importance of storage on the nodal performance. It also shows that, in general, sharing with some restriction on the contention for space is more advantageous than no sharing, especially when little storage is available. A summary of results obtained for the case of equal ρ_i's was reported by the authors in [14].

II. MODEL AND GENERAL SOLUTION

We consider R $M/M/1$ queueing systems which share a finite storage capacity of size B under one of the above schemes. Queueing system $i (i = 1, \cdots, R)$ is characterized by a Poisson input stream at a rate λ_i and an exponential service time of mean $1/\mu C_i$; C_i is the channel capacity (bits/s) and $1/\mu$ is the average number of bits per message. Customers to be served by server i are referred to as type or class i customers. Arriving customers not admitted to the queue (because of the sharing scheme) depart without service. Accepted (nonrejected) customers of type i are served by server i on a first-come first-served basis.

The sharing of space introduces dependencies among the R queueing systems. The entire system is a birth-death process [12], whose state can be simply described by the vector $\boldsymbol{n} = (n_1, \cdots, n_R)$ where n_i is a nonnegative random variable denoting the number of type-i customers. The basic equation which describes the behavior of the system of queues in steady state obeys the well-known product form solution for a network of queues, i.e.,

$$P(n_1, n_2, \cdots, n_R) = \begin{cases} P(\boldsymbol{n}) = C_x \rho_1^{n_1} \rho_2^{n_2} \cdots \rho_R^{n_R}, & \text{for } \boldsymbol{n} \in F_x \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\rho_i = \lambda_i/\mu C_i$. The subscript x indicates the scheme referred to, i.e., $x \in \{a, b, c, d, e\}$ where a stands for CP, b for CS, c for SMXQ, d for SMA, and e for SMQMA (see Fig. 1). F_x represents the set of possible system states. The proof for (1) comes from the fact that it satisfies the system's balance equations for all the schemes considered [1].

In what follows, we first characterize C_x for each of the five sharing schemes; then, from the joint probability distribution, we obtain the probability of blocking, the throughput, and the average delay.

C_x is simply the probability of an empty system, i.e., $C_x = P(\mathbf{0})$. When there is no confusion as to the scheme under investigation, we use the notation P_0 instead of C_x or $P(\mathbf{0})$. C_x can be computed by requiring that the probabilities sum to one, i.e.,

$$C_x^{-1} = \sum_{\boldsymbol{n} \in F_x} \rho_1^{n_1} \rho_2^{n_2} \cdots \rho_R^{n_R}. \tag{2}$$

III. COMPLETE PARTITIONING (CP)

CP is a degenerate case where actually all the R queueing systems are independent. The basic equations describing the

behavior of any of the queues are well known (see, for example, [12]). Furthermore, each of those systems is equivalent to CS with only one type of customer, and so we present the results for CP in the following section as the special case of CS for $R = 1$. We note that

$$F_a = \{\boldsymbol{n} \mid 0 \leqslant n_i \leqslant b_i, \quad i = 1, \cdots, R\}$$

where b_i is the number of buffers reserved for type-i customers.

IV. COMPLETE SHARING (CS)

We now combine the buffer space into a common pool [see Fig. 1(b)], whose size will be denoted by B. Empty space is allocated on a FCFS basis regardless of the type of arriving customer. In what follows, we analyze the general case of arbitrary ρ_i ($i = 1, \cdots, R$) and then we apply our results to the special case of equal ρ_i's.

A. General Case: Arbitrary ρ_i

In this section, the ρ_i's are arbitrary. The set of feasible system states is

$$F_b = \left\{ \boldsymbol{n} \mid \sum_{i=1}^{R} n_i \leqslant B \right\}. \tag{3}$$

Let us define $G(K)$ as

$$G(K) = \sum_{\substack{0 \leqslant n_i \leqslant K \\ \Sigma_i n_i = K}} {\rho_1}^{n_1} \cdots {\rho_R}^{n_R}. \tag{4}$$

From (2) and (3), we see that

$${C_b}^{-1} = {P_0}^{-1} = \sum_{K=0}^{B} G(K). \tag{5}$$

Several efficient algorithms exist to compute $G(K)$ [15]-[17]. Moreover, if all the ρ_i's are different,[2] we can use the generating function approach to derive a *closed form expression* for $G(K)$ which leads to faster computational algorithms and exhibits the interrelationships among the system variables. In this study, however, the numerical applications deal mostly with the equal ρ_i case [14]. Briefly, let

$$g(t) = \prod_{i=1}^{R} \frac{1}{(1-\rho_i t)} = \prod_{i=1}^{R} (1 + \rho_i t + {\rho_i}^2 t^2 + \cdots). \tag{6}$$

From this expansion and from (4), we recognize that $G(K)$ is the coefficient of t^K. A partial fraction expansion of the first product yields

$$G(K) = \sum_{i=1}^{R} A_i {\rho_i}^K \qquad K = 0, 1, 2, \cdots \tag{7}$$

[2] This condition ($\rho_i \neq \rho_j$) will be assumed throughout the rest of this section.

where

$$A_i = \prod_{\substack{k=1 \\ k \neq i}}^{R} \frac{1}{(1-\rho_k/\rho_i)}.$$

Then from (2), (5), and (7), we get

$${C_b}^{-1} = {P_0}^{-1} = \sum_{i=1}^{R} A_i \frac{1-{\rho_i}^{B+1}}{1-\rho_i}. \tag{8}$$

Equations (1) and (8) completely characterize our system in the steady state. Note that for $R = 1$, $A_1 = 1$; if we then let $B = b_i$ in (8), we obtain the expression of ${P_0}^{-1}$ (or ${C_a}^{-1}$) for the CP scheme. Now we proceed to derive the steady-state distribution and expressions of other variables of interest.

Distribution of the Total Number in System; Probability of Blocking: Let n be the number in system and $P_n = P_r[\Sigma n_i = n]$ be its corresponding distribution; then

$$P_n = P_0 G(n) = G(n) / \sum_{K=0}^{B} G(K)$$

for $0 \leqslant n \leqslant B$ and zero otherwise. Due to Poisson arrivals, the probability of blocking PB is simply

$$PB = P_0 G(B). \tag{9}$$

Marginal Distributions and Averages: A marginal distribution is defined as the probability distribution of a given class of customers. Here we derive the probability that there are at least j type i customers in the system $P_r[n_i \geqslant j]$. That probability is equal to the sum of $P(\boldsymbol{n})$ for $\boldsymbol{n} \in F_b$ and such that $n_i \geqslant j$; after some algebra, we find (for $0 \leqslant j \leqslant B$)

$$P_r[n_i \geqslant j] = P_0 {\rho_i}^j \sum_{n=0}^{B-j} G(n). \tag{10}$$

Then the expression for the average number of type i customers is

$$\bar{n}_i = \frac{\rho_i}{1-\rho_i} \frac{\sum_{n=0}^{B-1} (1-{\rho_i}^{B-n}) G(n)}{\sum_{n=0}^{B} G(n)}. \tag{11}$$

Let λ_i' be the average rate of nonrejected type-i customers, i.e., the throughput of server i, and let PB_i be the probability of blocking for type-i customers; then for all schemes considered in this paper,

$$\lambda_i' = (1 - PB_i)\lambda_i. \tag{12}$$

For CS, we have $PB_i = PB$ for all $i = 1, 2, \cdots, R$.

If we let T_i denote the average time in system (queue and server) of nonblocked type-i customers, then from Little's result

$$T_i = \bar{n}_i/\lambda_i' = \bar{n}_i/(1 - PB_i)\lambda_i. \tag{13}$$

A similar approach is used throughout this study whereby the marginal distribution of the number of type i customers is determined; then an expression for $\bar{n}_i$ (hence, T_i) is derived. Derivations of marginal distributions and averages will be omitted in this paper; they can be found in [1].

This terminates the characterization of CS in the general case. Of interest is the study of its behavior under some special limiting conditions of storage and traffic.

B. Limiting Behavior

We consider two cases: first, when B goes to infinity, and second, when all arrival rates increase uniformly toward infinity. In the first case, for the existence of a steady state, it is necessary that $\rho_i < 1 (i = 1, \cdots, R)$. With this condition, in the limit ($B = \infty$) the system becomes equivalent to R independent $M/M/1$ queues [1].

We now let all input rates increase proportionally toward infinity. In [1], we show that our system becomes equivalent to a closed network of R queues and B customers. Let

$$\lambda_i = \eta\lambda_i^0 \qquad i = 1, \cdots, R \tag{14}$$

where the scale factor η is a positive real variable and λ_i^0 is a constant. The service rates (μC_i) are maintained constant. Equation (14) is also equivalent to saying that $\rho_i = \eta\rho_i^0$ with $\rho_i^0 = \lambda_i^0/\mu C_i$ constant.

From the above definition and (7),

$$G(K) = \eta^k G^0(K) \tag{15}$$

where

$$G^0(K) = \sum_{i=1}^{R} A_i(\rho_i^0)^K.$$

Note that A_i is invariant to the rate increase; hence, $G^0(K)$ is also independent of η. From the above definitions, we determine the limiting throughput

$$\lim_{\eta \to \infty} (1 - PB)\eta\lambda_i^0 = \frac{G^0(B-1)}{G^0(B)}\lambda_i^0. \tag{16}$$

Also, $P_r[n_i = 0] = 1 - G^0(B-1)\rho_i^0/G^0(B)$; hence, there is a nonzero probability that server i is idle (i.e., not fully utilized) even with infinite input rates. This is not the case with CP, since for $R = 1$, $G^0(K) = (\rho_i^0)^K$; hence, $P_r[n_i = 0] = 0$.

The numerical example below illustrates the general and limiting behavior of this system with respect to η. In this example, we assume that $R = 4$, $B = 20$, $\rho_1^0 = 0.1$, $\rho_2^0 = 0.4$, $\rho_3^0 = 0.6$, $\rho_4^0 = 0.9$, and we let $\rho_i = \eta\rho_i^0$.

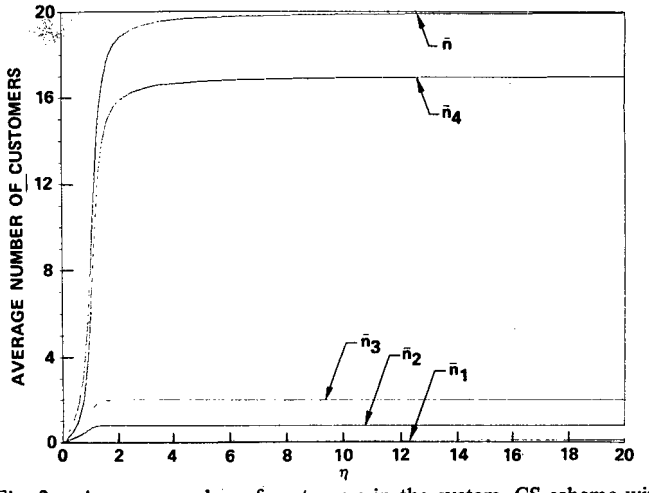


Fig. 2. Average number of customers in the system, CS scheme with asymmetric input rates.

The utilization of server i is $\rho_i' = (1 - PB)\rho_i = (1 - PB)\eta\rho_i^0$. The limiting value of $\eta(1 - PB) = 1.111$; hence, $\eta \to \infty \Rightarrow \rho_1' = 0.1111, \rho_2' = 0.4444$, $\rho_3' = 0.6666$, and $\rho_4' = 0.9999$. Note that server 4 reaches saturation, whereas the others are far from it. The average total limiting utilization is $\bar{\rho} = \frac{1}{R}\Sigma\rho_i' = 0.555$ instead of 1 which could be obtained with CP.

Fig. 2 shows the behavior of the average number of type-i customers in the system ($i = 1, \cdots, 4$) with respect to η. Also represented is the average total number in system $\bar{n}$. The limiting values for the averages are (evaluated at $\eta = 20$)

$$\bar{n}_1 \to 0.125, \bar{n}_2 \to 0.800, \bar{n}_3 \to 1.99, \bar{n}_4 \to 17.02,$$

and

$$\bar{n} \to 19.94 \approx B = 20.$$

Note that for large η, most of the buffers are, on the average, used by type 4 customers. Also, a sharp increase of $\bar{n}_4$ (from 4.8 to 14) occurs when η varies from 0.95 to 1.5. The value of $\eta = 1.111$ corresponds to the saturation of server 4 (i.e., $\rho_4 \approx 1$) if there were no limitation in buffer storage, and at that point the queue size becomes infinite. This explains the sharp increase in $\bar{n}_4$.

In summary, we conclude that with asymmetrical utilizations $\{\rho_i^0\}$, CS tends to favor the server with the highest utilization even though it has reached saturation. Furthermore, the other servers are left with very little space to share, and therefore, they often go idle. These considerations motivate the schemes studied in the rest of this paper. Before we proceed, let us apply the general results obtained in this section to the case where all ρ_i's are equal.

C. Special Case: Equal ρ_i's

This section deals with the case where all the ρ_i's are equal and we let ρ be the common value. As a result, a simpler expression is obtained for $G(K)$, and thus for the other variables and distributions.

$G(K)$ is the well-known expression obtained for networks of queues [12].

$$G(K) = \binom{K+R-1}{R-1}\rho^K. \tag{17}$$

Also, the average time in system of the nonrejected type-i customers is

$$T_i = \frac{1/\mu C_i}{1-\rho} \frac{\sum_{K=0}^{B-1} \binom{K+R-1}{R-1}\rho^K - \binom{B+R-1}{R}\rho^B}{\sum_{K=0}^{B-1} \binom{K+R-1}{R-1}\rho^K}$$

$$i = 1, \cdots, R. \tag{18}$$

Of interest are the two cases when $\rho = 1$ and $\rho \to \infty$.

$\rho = 1$:

$$\left\{ \begin{array}{ll} P_0^{-1} = \binom{B+R}{R} & \lambda_i' = \dfrac{R}{B+R}\mu C_i \\ P(n_1, \cdots, n_R) = P_0 \quad \forall \boldsymbol{n} \in F_b & \\ PB = \dfrac{R}{R+B} & T_i = \dfrac{B+R}{R+1}\dfrac{1}{\mu C_i} \\ \bar{n}_i = \dfrac{B}{R+1} \quad i = 1, R; & \bar{n} = \dfrac{RB}{R+1} \end{array} \right\} \tag{19}$$

Note that all states $\boldsymbol{n}$ have equal probability P_0.

The expression for PB may be rewritten as $1/(1 + B/R)$, which is exactly the same as for a single $M/M/1$ queue with B/R buffers. This means that for $\rho = 1$, CS and CP (with $b_i = B/R$) lead to the same probability of blocking. This fact is illustrated in the figures below.

$\rho \to \infty$:

The service rates (μC_i) are assumed to be constant. The limits are

$$\left\{ \begin{array}{ll} P_0 \to 0, & PB \to 1 \\ \bar{n}_i \to B/R, & \bar{n} \to B \\ \lambda_i' \to \dfrac{B}{B+R-1}\mu C_i, & i = 1, \cdots, R \\ T_i \to \dfrac{B+R-1}{R}\dfrac{1}{\mu C_i}, & i = 1, \cdots, R. \end{array} \right. \tag{20}$$

As noted earlier, infinite input rates do not lead to full utilization of the servers (except for $R = 1$), but only to a fraction $B/(B + R - 1)$ of the capacity.

The illustration of the behavior of the probability of blocking, the utilization, and the delay with respect to the load $\rho = \lambda/\mu C$ and for several values of B can be found in [1].

This concludes the analysis of the complete sharing (CS) scheme. Let us now compare it with the complete partitioning (CP) scheme.

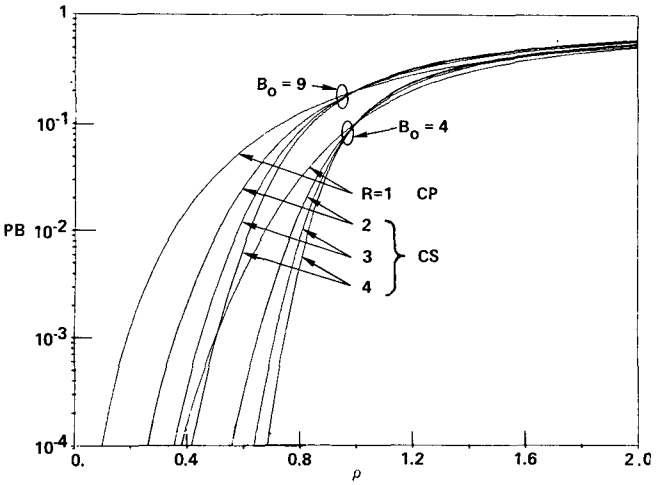


Fig. 3. Comparison of CP and CS: blocking.

D. Comparison of CP and CS

In this section, we assume that all ρ_i's are equal (to ρ) and that each server "contributes" B_0 buffers, i.e., $b_i = B_0$, $i = 1, \cdots, R$ [see Fig. 1(a)]; therefore, $B = RB_0$.

With the above conditions, the behavior of CP (for any of its queues) is identical to CS with $R = 1$.

Fig. 3 illustrates the behavior of the probability of blocking PB with respect to ρ for a set of values of R, ($R = 1, \cdots, 4$). $R = 1$ corresponds to CP; $R = 2, 3, 4$ corresponds to the merging of 2, 3, 4 single queues. Note that all the curves meet at $\rho = 1$ where, from (19), $PB = 1/(1 + B_0)$. Note also that for $0 \leqslant \rho < 1$, CS leads to a smaller PB, and, hence, a better performance than CP. This improvement is quite considerable for small values of B_0 and increases with R. However, for $\rho > 1$, CP shows a better performance (smaller PB) than CS.

Fig. 4 shows the respective channel utilizations $\rho(1 - PB)$ (normalized throughputs $\lambda'/\mu C$). Note the loss in limiting throughput ($\rho \to \infty$) with CS for small values of B_0.

Finally, Fig. 5 shows the respective average delays. We note, of course, that the average message delay for the nonblocked traffic increases as more buffers are provided, i.e., as R increases.

The better performance of CP for $\rho > 1$ intuitively indicates that some buffers should be permanently allocated to each server. This idea is incorporated in scheme 4, SMA. Moreover, we observed earlier that very unbalanced input rates lead (on the average) to uneven usage of the storage space. This remark motivates the next scheme, SMXQ.

V. SHARING WITH MAXIMUM QUEUE LENGTHS (SMXQ)

Like CS, SMXQ allows the sharing of a pool of B buffers with a further constraint imposed on the number of buffers to be allocated to any server, and at any time. Let b_i be the maximum number of buffers that can be used by type-i customers; the set of feasible states becomes

$$F_c = \left\{ \boldsymbol{n} \mid 0 \leqslant \sum_{i=1}^{R} n_i \leqslant B, \right.$$

$$\left. 0 \leqslant n_i \leqslant b_i; \quad i = 1, \cdots, R \right\}$$

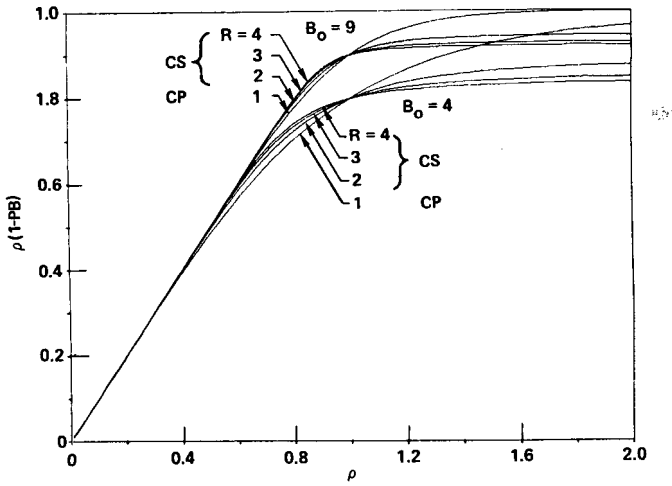


Fig. 4. Comparison of CP and CS: utilization.

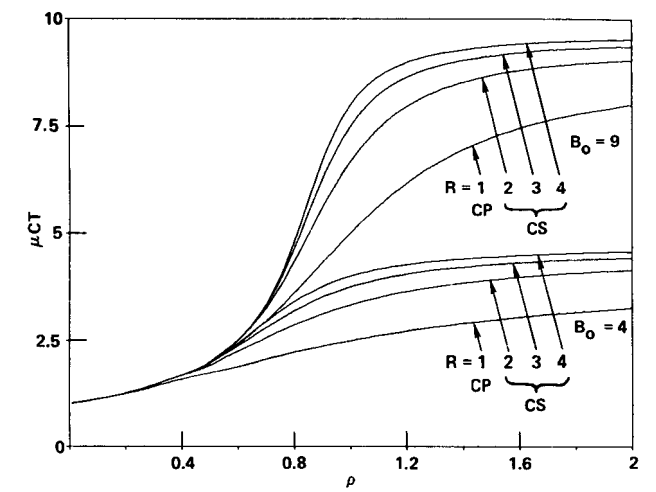


Fig. 5. Comparison of CP and CS: delay.

The evaluation of C_c is much more complicated here because of the added constraint on n_i. In what follows, we again consider the two cases of different and equal ρ_i's.

A. General Case

In this section, the ρ_i's are all different. We first evaluate C_c (also denoted by P_0). From the above considerations,

$$C_c^{-1} = P_0^{-1} = \sum_{K=0}^{B} Q(K) \tag{21}$$

where

$$Q(K) = \sum_{\substack{\sum_{i=1}^{R} n_i = K \\ 0 \leq n_i \leq b_i}} \rho_1^{n_1} \cdots \rho_R^{n_R}. \tag{22}$$

Note that the difference between $Q(K)$ and $G(K)$, (4), comes from the added constraint $n_i \leq b_i$.

In order to find $Q(K)$, we use a method similar to the generating function approach. Let $f(t)$ be defined as

$$f(t) = \prod_{i=1}^{R} (1 + \rho_i t + (\rho_i t)^2 + \cdots + (\rho_i t)^{b_i}) = \prod_{i=1}^{R} \frac{1 - (\rho_i t)^{1+b_i}}{1 - \rho_i t}. \tag{23}$$

From the expansion of the first product, we recognize $Q(k)$ as the coefficient of t^k.

Let

$$h(t) = \prod_{i=1}^{R} \frac{1}{1 - (\rho_i t)^{1+b_i}}.$$

Then, recalling (6), we have $g(t) = f(t)h(t)$; this leads to an equation relating Q to G in terms of C_i. C_i is computed from the partial fraction expansion of $h(t)$.

$$\sum_{K=0}^{\infty} G(K)t^K = \left(\sum_{K=0}^{\sum b_i} Q(K)t^K \right) \left[\sum_{i=1}^{R} C_i (1 + (\rho_i t)^{1+b_i} + \cdots + (\rho_i t)^{k(1+b_i)} + \cdots) \right] \tag{24}$$

where

$$C_i = \prod_{\substack{j=1 \\ j \neq i}}^{R} \frac{1}{1 - (\rho_j/\rho_i)^{1+b_j}}. \tag{25}$$

Equating the terms of equal degrees in t in (24), we arrive at a relation between G, Q, and the C_i's. This relation is quite complicated and requires the ordering of the b_i's. However, if we restrict our consideration, either to the case where $b_i \geq B/2$ or where $b_i = b$ for all i, then we obtain the simple relations below.

1) We assume that each queue is allowed to occupy more than half of the entire space, i.e.,

$$b_i \geq B/2 \qquad \forall i = 1, \cdots, R. \tag{26}$$

Then regardless of the assumption of different ρ_i's, we find

$$G(K) = Q(K) + \sum_{\substack{i=1 \\ i \text{ s.t. } b_i < K}}^{R} \rho_i^{1+b_i} Q(K - b_i - 1). \tag{27}$$

2) Now assume $b_i = b$ for $i = 1, \cdots, R$. As with $G(K)$, let us define

$$L(K) = \sum_{\sum n_i = K} \rho_1^{(1+b)n_1} \cdots \rho_R^{(1+b)n_R}. \tag{28}$$

Similarly

$$L(K) = \sum_{i=1}^{R} C_i \rho_i^{(1+b)K} \tag{29}$$

and

$$h(t) = \sum_{K \geq 0} L(K) t^{(1+b)K}. \tag{30}$$

Substituting this into (24), we arrive at

$$G(\alpha(b+1)+k) = \sum_{i=0}^{\alpha} Q((\alpha - i)(b+1)+k) L(i) \tag{31}$$

where

$$0 \leq k \leq b$$

and

$$0 \leq \alpha(b+1) + k \leq B.$$

Note that $G(0) = Q(0) = L(0) = 1$, and that (31) allows the sequential computation of the sequence $Q(K)$ for K varying from 1 to B.

Note that if $b \geq B/2$, then (31) becomes

$$\begin{cases} G(k) = Q(k) & \text{for } 0 \leq k \leq b \\ G(b+1+k) = Q(b+1+k) + Q(k)L(1) & \\ & 0 \leq k \leq B-b-1 \end{cases}$$

and from (28),

$$L(1) = \sum_{i=1}^{R} \rho_i^{(1+b)};$$

thus, the combination of the last three equations gives (27) with $b_i = b$.

In what follows, we *restrict* our considerations to the case where $b_i \geq B/2$ unless specified otherwise. As a result, and from (21) and (27), we arrive at

$$C_c^{-1} = P_0^{-1} = \sum_{K=0}^{B} G(K) - \sum_{i=1}^{R} \rho_i^{1+b_i} \sum_{K=0}^{B-b_i-1} G(K). \tag{32}$$

Note that if $b_i = B$ for all i, then SMXQ becomes CS, and the above equation reduces[3] to (5).

Similarly to Section IV, we now assume that all the ρ_i's are different ($\rho_i \neq \rho_j \; \forall \; i \neq j$); then, using (7) and (8), we arrive at

$$C_c^{-1} = P_0^{-1} = \sum_{i=1}^{R} A_i \frac{1-\rho_i^{B+1}}{1-\rho_i} - \sum_{\substack{i=1 \\ i \text{ s.t. } 0 \leq b_i < B}}^{R} \rho_i^{1+b_i} \sum_{j=1}^{R} A_j \frac{1-\rho_j^{B-b_i}}{1-\rho_j}. \tag{33}$$

[3] By convention, we set $\sum_{a}^{a-1} \triangleq 0, \forall \; a$ integer.

Equation (1) and either (32) or (33) completely characterize the queueing system under SMXQ and the condition of (26). Similarly to Section IV, we proceed with the derivation of distributions and average quantities of interest.

Distribution of the Total Number in System–Probability of Blocking: Let n be the total number in system; then, for $0 \leq n \leq B$, $P_n \triangleq P_r[\Sigma_i n_i = n] = P_0 Q(n)$; and then from (27),

$$P_n = P_0 \left[G(n) - \sum_{\substack{i=1 \\ i \text{ s.t. } b_i < n}}^{R} \rho_i^{1+b_i} G(n - b_i - 1) \right]. \tag{34}$$

Let us now derive the probability of blocking of type-i customers PB_i. Recall that type-i customers are blocked if, upon arrival, the entire space is full or if the number of type-i customers is equal to b_i. Since arrivals are Poisson, then

$$PB_i = P_r[\Sigma_j n_j = B \quad \text{or} \quad n_i = b_i] \tag{35}$$

which is also

$$PB_i = P_r[\Sigma n_j = B] + \sum_{K=0}^{B-b_i-1} P_r \left[n_i = b_i \text{ and } \sum_j n_j = b_i + K \right].$$

Because of (26),

$$n_i = b_i \geq B/2 \quad \text{and} \quad \Sigma_j n_j = b_i + K \Rightarrow n_j \leq b_j \quad \forall_j \neq i;$$

hence,

$$PB_i = P_r[\Sigma_i n_i = B] + P_0 \rho_i^{b_i} \sum_{K=0}^{B-b_i-1} G_i(K) \tag{36}$$

where

$$G_i(K) = \sum_{\substack{\Sigma \\ j \neq i}} {}_{n_j = K} \rho_1^{n_1} \cdots \rho_{i-1}^{n_{i-1}} \rho_{i+1}^{n_{i+1}} \cdots \rho_R^{n_R}. \tag{37}$$

Note that $G_i(K)$ is similar to $G(K)$, except that we have deleted ρ_i (i.e., type-i customers).

We now proceed with the derivation of the marginal distribution and average number and delay of type-i customers.

Marginal Distribution and Limiting Behavior: Let $k \leq b_i$; then

$$P_r[n_i = k] = P_0 \sum_{\substack{n \in F_c \\ n_i = k}} \rho_1^{n_1} \cdots \rho_R^{n_R}$$

$$= P_0 \rho_i^k \sum_{\substack{0 \leq \Sigma n_j \leq B-k \\ j \neq i \\ 0 \leq n_j \leq b_j}} \left(\prod_{\substack{j=1 \\ j \neq i}}^{R} \rho_j^{n_j} \right).$$

The summation above is similar to that of P_0^{-1} except that B is now $B - k$ and the component n_i is deleted. Also note that $b_j \geq (B-k)/2 \; \forall \; j \neq i$ and, hence, the above summation is given by (32) where B is replaced by $B - k$ and $G(K)$ by $G_i(K)$ [as defined in (37)]. Thus,

$$P_r[n_i = k] = P_0 \rho_i^k \left[\sum_{K=0}^{B-k} G_i(K) - \sum_{\substack{j=1 \\ j \neq i}}^{R} \rho_j^{1+b_j} \sum_{K=0}^{B-k-b_j-1} G_i(K) \right]. \tag{38}$$

The above summations can be further reduced to expressions similar to the one in (33). From (38), we may derive the average number, delay, and utilization of type-i customers.

This terminates the characterization of the system as operated with SMXQ and with the assumption of $b_i \geq B/2$ and different ρ_i's. Next we study the case of equal ρ_i's; we leave numerical applications to Section VIII. Before we proceed, let us note that in a similar environment as that of the numerical example in Section IV-B, we obtain in the limit of $\eta \to \infty$ and for $b_i = b = 10$,

$$\rho_1' = 0.1985, \rho_2 = 0.7904, \rho_3 = 0.9958, \rho_4' = 0.9999.$$

(These values correspond to $\eta = 20$.) Therefore, the average utilization $\bar{\rho} = 0.746$, which represents an improvement over the value obtained with CS, $\bar{\rho} = 0.555$. As for the limiting average numbers of customers, we find (evaluated at $\eta = 20$), $\bar{n}_1 = 0.242, \bar{n}_2 = 2.49, \bar{n}_3 = 8.01, \bar{n}_4 = 9.16$.

B. Special Case: Equal ρ_i's

As in Section IV-C, let $\rho_i = \rho \; \forall \; i$; then $G(K)$ is given by (17). Also, we assume that all the b_i's are equal to b and that $b \geq B/2$. Therefore, from (32), P_0^{-1} becomes

$$P_0^{-1} = \sum_{K=0}^{B} \binom{K+R-1}{R-1} \rho^K - R\rho^{b+1} \sum_{K=0}^{B-b-1} \binom{K+R-1}{R-1} \rho^K. \tag{39}$$

From (37) and similarly to $G(K)$, we have

$$G_i(K) = \binom{K+R-2}{R-2} \rho^K. \tag{40}$$

From (34), (36), and (40), we derive the probability of blocking PB which is independent of the customer's type:

$$PB_i = PB = P_0 \rho^B \left[\binom{B+R-1}{R-1} - R \binom{B-b+R-2}{R-1} \right] + P_0 \rho^b \sum_{K=0}^{B-b-1} \binom{K+R-2}{R-2} \rho^K. \tag{41}$$

The rest of the expressions, $P_r[n_i = k]$, $\bar{n}_i$, λ_i', T_i, follow in the same way as before. This terminates the analysis of the SMXQ schemes; further expressions of the above variables (P_0, PB) at $\rho = 1$ and $\rho \to \infty$ can be found in [1]. We note, in particular, that if $\rho \to \infty$, then the utilization of any server $\rho(1 - PB)$ does not reach one except for $R = 1$ and $R = 2$ (assuming that $b < B$ for $R = 2$). For $R > 2$ (and $\rho \to \infty$), SMXQ still does not provide a full utilization of the server. Our next scheme is motivated by this deficiency.

VI. SHARING WITH MINIMUM ALLOCATION (SMA)

Similarly to CS, SMA allows the sharing of a pool of B buffers and, in addition, a_i buffers are permanently allocated to type-i customers, $i = 1, \cdots, R$ (see Fig. 1). As a result, the set of feasible states becomes

$$F_d = \left\{ n \;\middle|\; \sum_{i=1}^{R} \sup\{0, n_i - a_i\} \leq B, \quad 0 \leq n_i \leq B + a_i \quad i = 1, \cdots, R \right\}.$$

Following the same steps as earlier, we first consider the general case of different ρ_i's.

A. General Case

In order to evaluate C_d (also denoted by P_0), we partition the set F_d into disjoint subsets which lead to known summations. Let $\mathcal{R}$ be the set of customer types,

$$\mathcal{R} = \{1, 2, \cdots, R\}$$

and let $\mathcal{X}$ be the set of all subsets of $\mathcal{R}$,

$$\mathcal{X} = \{X_m \mid X_m \subset \mathcal{R}, \quad 1 \leq m \leq 2^R\}.$$

The set $\mathcal{X}$ contains 2^R elements; among them are the set $\mathcal{R}$ itself and the empty set. We then associate with each subset X_m a subset of F_d, namely, S_m defined as

$$S_m = \left\{ n \in F_d \mid n_i \begin{cases} \geq a_i & i \in X_m \\ < a_i & \text{otherwise} \end{cases} \right\}.$$

Obviously,

$$S_m \cap S_n = \phi \quad \text{for } m \neq n \quad \text{and } F_d = \bigcup_{m=1}^{2^R} S_m.$$

Therefore,

$$P_0^{-1} = \sum_{n \in F_d} \left(\prod_{i=1}^{R} \rho_i^{n_i} \right) = \sum_{m=1}^{2^R} \sum_{n \in S_m} \left(\prod_{i=1}^{R} \rho_i^{n_i} \right) \triangleq \sum_{m=1}^{2^R} H_m(a, B). \tag{42}$$

$H_m(a, B)$ is defined as the summation of the products of the ρ_i's over all states in S_m; a is the vector $(a_1, a_2, \cdots, a_R)$. From the definition of S_m, we may easily compute $H_m(a, B)$ (also denoted H_m); let us define the generating function $C_m(k)$ such that

$$C_m(K) = \sum_{\substack{\sum_{i\in X_m} n_i = K \\ 0 \leq n_i \leq K}} \left(\prod_{i\in X_m} \rho_i^{n_i} \right). \tag{43}$$

$C_m(K)$ is similar to $G(K)$ given in (4); thus, it can be computed in the same way.

$$H_m(a, B) \triangleq H_m = \prod_{i\notin X_m} \frac{1-\rho_i^{a_i}}{1-\rho_i} \prod_{i\in X_m} \rho_i^{a_i} \sum_{K=0}^{B} C_m(K). \tag{44}$$

Note that if $a_i = 0$ for all i, then $X = \{R\}$, $C_m(K) = G(K)$, and the above equations reduce to the description of CS. Also, the summation of $C_m(K)$ in (44) is set to 1 if $X_m = \phi$.

If we now assume that all ρ_i's are different, then from (7),

$$C_m(K) = \sum_{i\in X_m} A_{i,m} \rho_i^K \tag{45}$$

where

$$A_{i,m} = \prod_{\substack{k\in X_m \\ k\neq i}} \frac{1}{1-\rho_k/\rho_i}. \tag{46}$$

Using a similar summation as in (8), we arrive at

$$H_m = \prod_{i\notin X_m} \frac{1-\rho_i^{a_i}}{1-\rho_i} \prod_{i\in X_m} \rho_i^{a_i} \sum_{i\in X_m} A_{i,m} \frac{1-\rho_i^{B+1}}{1-\rho_i}. \tag{47}$$

Equations (1) and (44) or (47) completely characterize our system. We now proceed with the derivation of distributions and averages of the variables of interest.

Distribution of Total Number of Customers in Shared Area—Probability of Blocking: Let n_s be the total number of customers in the shared area, i.e.,

$$n_s \triangleq \sum_{i=1}^{R} \sup\{0, n_i - a_i\}. \tag{48}$$

Then the distribution of n_s is, for $k \leq B$, equal to the sum of probabilities of states n which satisfy (48) for $n_s = k$. In order to evaluate that summation, we use the same methodology as for the determination of P_0. Let $F_d(k) \subset F_d$ and $S_m(k) \subset S_m$ be defined as

$$F_d(k) = \left\{ n \,\middle|\, \sum_{i=1}^{R} \sup\{0, n_i - a_i\} = k, \quad 0 \leq n_i \leq k + a_i \right\}$$

$$S_m(k) = \left\{ n \,\middle|\, \begin{array}{ll} a_i \leq n_i \leq a_i + k & i \in X_m, \quad \sum_{i\in X_m} (n_i - a_i) = k \\ n_i < a_i & i \notin X_m \end{array} \right\}.$$

It is obvious that $F_d(k) = \bigcup_m S_m(k)$ and, consequently,

$$P_r[n_s = k] = P_0 \sum_{m=1}^{2^R} \sum_{n\in S_m(k)} \left(\prod_{i=1}^{R} \rho_i^{n_i} \right) \triangleq P_0 \sum_{m=1}^{2^R} h_m(k)$$

where $h_m(k)$ is the summation over all states in $S_m(k)$. Then, as above,

$$h_m(k) = \prod_{i\notin X_m} \frac{1-\rho_i^{a_i}}{1-\rho_i} \left(\prod_{i\in X_m} \rho_i^{a_i} \right) C_m(k). \tag{49}$$

For the probability of blocking of type-r customers PB_r, we have

$$PB_r = P_r[n_s = B \quad \text{and} \quad n_r \geq a_r]. \tag{50}$$

PB_r can be computed in a fashion similar to $P_r[n_s = B]$ with the restriction that the subset X_m must contain r, i.e.,

$$PB_r = P_0 \sum_{\substack{m \\ m \mid r\in X_m}} h_m(B). \tag{51}$$

There are 2^{R-1} such sets which can be obtained as follows. Let $\mathcal{R}' = \mathcal{R} - \{r\}$ and $X' = \{X_n' \quad n = 1, \cdots, 2^{R-1}\}$ be the set of subsets of $\mathcal{R}'$; then $X_n = X_n' \cup \{r\}$ is such a subset of $\mathcal{R}$ which contains r.

Marginal Distribution and Average Number and Time in System: Below, we give the expression of the marginal distribution of type-r customers or, more precisely, $P_r[n_r \geq j]$. The methodology is similar to that used to find P_0.

$$P_r[n_r \geq j] = \begin{cases} P_0 \rho_r^j \sum_{m=1}^{2^R} H_m[a', B] & \forall j < a_r, a' = (a_1, \cdots, a_r - j, \cdots, a_R) \\ P_0 \rho_r^{j-a_r} \sum_{m \mid r\in X_m} H_m(a, B - j + a_r) & \forall j \text{ s.t. } a_r \leq j \leq B. \end{cases} \tag{52}$$

From the above equation, we may obtain $\bar{n}_r$ and, hence, T_r.

Let us now apply the above results to the special case of uniform utilization and allocations.

B. Special Case: $\rho_i = \rho$, $a_i = a$

The assumption of equal ρ_i's and a_i's leads to much simpler expressions for the variables above. First, if p is the size of the

subset $X_m (p = | X_m |)$, then from (43)

$$C_m(K) = \binom{K+p-1}{p-1} \rho^K. \tag{53}$$

Note that if $p = 0$, then

$$C_m(K) = \begin{cases} 0 & K > 0 \\ 1 & K = 0. \end{cases} \tag{54}$$

Also, from (44),

$$H_m = \left(\frac{1-\rho^a}{1-\rho}\right)^{R-p} \rho^{pa} \sum_{K=0}^{B} \binom{K+p-1}{p-1} \rho^K. \tag{55}$$

Note that $C_m(K)$ and H_m depend only on the size p of the set X_m. The number of sets X_m of size p is equal to $\binom{R}{p}$; thus, from (44) and (55),

$$P_0^{-1} = \sum_{p=0}^{R} \binom{R}{P} \left(\frac{1-\rho^a}{1-\rho}\right)^{R-p} \cdot \rho^{pa} \sum_{K=0}^{B} \binom{K+p-1}{p-1} \rho^K. \tag{56}$$

Similarly, we derive the expression for PB_r (51). Recall that we only account for sets X_m which contain r and, hence, $p \geq 1$.

$$PB_r = P_0 \sum_{p=1}^{R} \binom{R-1}{p-1} \left(\frac{1-\rho^a}{1-\rho}\right)^{R-p} \cdot \rho^{pa} \binom{B+p-1}{p-1} \rho^B \qquad r = 1, \cdots, R. \tag{57}$$

Note that in the above expressions, a was assumed to be greater than zero; *if* $a = 0$, then all subsets X_m are empty except one: $X_m = \mathcal{R}$ whose size is equal to R. Moreover, with $a = 0$, SMA reduces to CS. If $B = 0$, SMA reduces to CP.

Let us now derive the marginal distribution of the number of type-r customers. Equation (52) provides $P_r[n_i \geq j]$ for $j < a_r$; the terms $H_m[\boldsymbol{a}', B]$ can be evaluated as in (56) except that $a_r' = a_r - j$. Therefore, we must distinguish the sets X_m which contain r from those which do not. If $p = | X_m |$, then $\binom{R-1}{p-1}$ such sets contain r and $\binom{R-1}{p}$ do not. As a consequence,

$$P_r[n_r \geq j] = P_0 \sum_{p=1}^{R} \binom{R-1}{p-1} \left(\frac{1-\rho^a}{1-\rho}\right)^{R-p} \cdot \rho^{pa} \sum_{K=0}^{B} \binom{K+p-1}{p-1} \rho^K + P_0 \sum_{p=0}^{R-1} \binom{R-1}{p} \left(\frac{1-\rho^a}{1-\rho}\right)^{R-p-1} \times \frac{1-\rho^{a-j}}{1-\rho} \rho^{pa+j} \sum_{K=0}^{B} \binom{K+p-1}{p-1} \rho^K \qquad \text{for } j < a_r. \tag{58}$$

For the case where $j \geq a_r$, and using the same procedure as above, we find

$$P_r[n_r \geq j] = P_0 \sum_{p=1}^{R} \binom{R-1}{p-1} \left(\frac{1-\rho^a}{1-\rho}\right)^{R-p} \cdot \rho^{(p-1)a+j} \sum_{K=0}^{B-j+a} \binom{K+p-1}{p-1} \rho^K. \tag{59}$$

The calculation of $\bar{n}_r$, λ_r', T_r follows from the above considerations.

Of further interest is the limiting behavior when ρ goes to infinity. Indeed, we find for a nondegenerate SMA, i.e., $a \neq 0$, that $\rho \to \infty \Rightarrow P_0 \to 0$, $PB_r \to 1$, $\rho(1 - PB_r) \to 1$. Hence, as expected, the minimum allocation of at least one buffer per channel allows a full utilization of the channels in the limit.

With SMA, the shared area is prone to be unfairly utilized in the case of unbalanced traffic rates. We accommodate for this deficiency in our next and final scheme.

VII. SHARING WITH MAXIMUM QUEUE LENGTH AND MINIMUM ALLOCATIONS: SMQMA

In addition to SMA, SMQMA (or scheme e) imposes a constraint on the maximum number of buffers from the shared pool to be allocated to any server at any time. Let b_i be that constraint with respect to server i. As a result, the set of feasible states becomes

$$F_e = \{n \in F_d \mid 0 \leq \sup\{0, n_i - a_i\} \leq b_i \qquad i = 1, \cdots, R\}.$$

Equivalently,

$$F_e = \left\{ \boldsymbol{n} \mid 0 \leq \sum_{i=1}^{R} \sup\{0, n_i - a_i\} \leq B, \quad 0 \leq n_i \leq a_i + b_i \right\}.$$

We proceed as earlier with the evaluation of C_e, which we also denote by P_0.

General Case: The same procedure as in Section VI can be utilized to partition the set F_e into disjoint subsets which then leads to known summations. Those subsets are

$$S_m{}^e = \left\{ \boldsymbol{n} \;\middle|\; \sum_{i \in X_m} (n_i - a_i) \leq B, \quad a_i \leq n_i \leq b_i + a_i \quad i \in X_m; \quad n_i < a_i \quad i \notin X_m \right\}.$$

Consequently,

$$C_e{}^{-1} = P_0{}^{-1} = \sum_{m=1}^{2^R} H_m{}^e \tag{60}$$

with

$$H_m{}^e = \left(\prod_{i \notin X_m} \frac{1-\rho_i{}^{a_i}}{1-\rho_i}\right)\left(\prod_{i \in X_m} \rho_i{}^{a_i}\right) \sum_{K=0}^{B} Q(K) \tag{61}$$

where $Q(K)$ is as defined in (22).

As a consequence, the computation of P_0^{-1} follows as in Sections V [for $Q(K)$] and VI. This remark holds true for the computation of the summations which appear in the analysis of this scheme. As a result, we need carry out the study of this scheme no further.

VIII. FURTHER NUMERICAL RESULTS AND COMPARISONS

In this section, we intend to compare our first four sharing schemes: CP, CS, SMXQ, SMA under the assumption of equal ρ_i's. Although SMQMA appears to be an excellent sharing scheme which has the ability to avoid the deficiencies of the other four schemes, we do not include it in the comparison since a rather involved and detailed numerical evaluation is required and then the overall study of optimizing its many parameters must be carried out; this comparison is currently under study and is the subject of a forthcoming paper. Before we proceed, let us recall that if B is the total number of buffers ($B = RB_0$) and b is the maximum queue size (for any queue) when using an SMXQ scheme, then

1) if $b = B$, SMXQ is equivalent to CS;

2) if $b = B_0$, SMXQ is equivalent to CP; and

3) if $R = 2$, then SMXQ is equivalent to SMA with a minimum allocation per queue equal to $B - b$.

Thus, the study of SMXQ with $R = 2$ and a variable b will allow us to cover the four sharing schemes to be considered here.

In the numerical example below, we assume that $R = 2$, $B = 6$, and that b satisfies $B/2 \leqslant b \leqslant B$ [see (26)], i.e., $b = 3, 4, 5, 6$. From our previous considerations, we know that $b = 3$ leads to CP, $b = 4$ and $b = 5$ lead to nondegenerate SMXQ and SMA, and $b = 6$ leads to CS.

Figs. 6, 7, and 8, respectively, show the probability of blocking PB, the channel utilization $\rho(1 - PB)$, and the normalized average message delay μCT, obtained with the four schemes. With respect to blocking and utilization, the optimal b (i.e., the optimal scheme) is a function of ρ. We note that for small values of ρ, $b = 6$ (i.e., CS) is optimal; as ρ increases, $b = 5$, then $b = 4$ (i.e., SMXQ, SMA) becomes optimal, and, finally, for a larger ρ, $b = 3$ (i.e., CP) becomes optimal. However, the average delay is an increasing function of b, thereby showing a tradeoff between the probability of blocking and the system delay. The selection of a particular scheme must account for these two variables, as well as the load on the system.

IX. SUMMARY

In this study, we considered various schemes for sharing a pool of buffers among a set of communication channels in a computer communication network environment. Five sharing schemes were examined, and the results of the analysis were presented and displayed in a fashion which permits one to establish the tradeoffs among blocking probability, utilization, throughput, and delay.

We have shown that, in general, sharing with some restrictions on the contention for space is certainly more advantageous than nonsharing, especially when little storage is available.

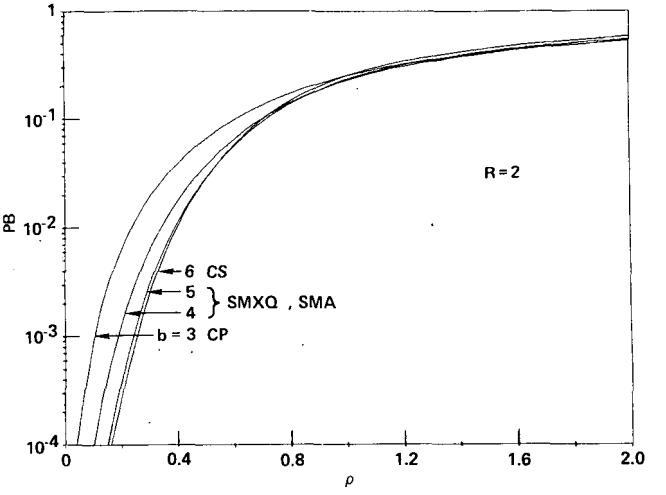


Fig. 6. Comparison of the four schemes: blocking.

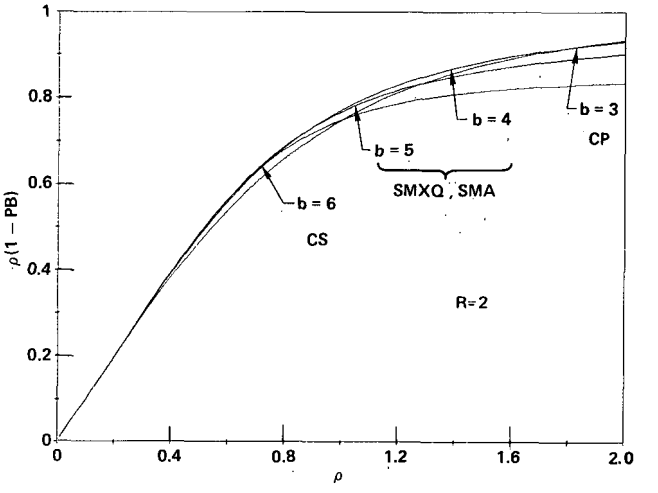


Fig. 7. Comparison of the four schemes: utilization.

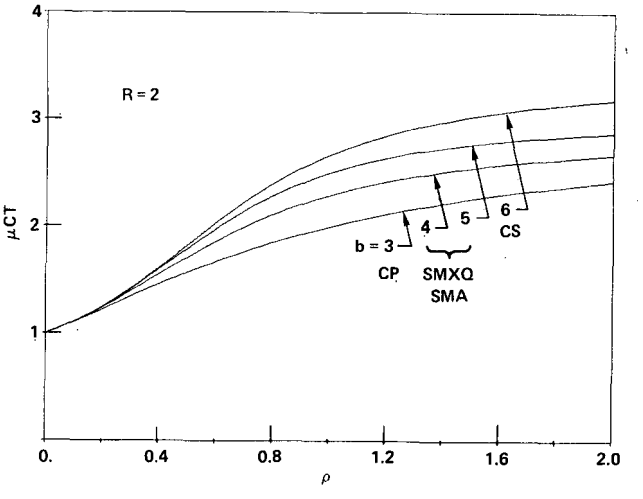


Fig. 8. Comparison of the four schemes: delay.

REFERENCES

[1] F. Kamoun, "Design considerations for large computer communication networks," Dep. Comput. Sci., School Eng. and Appl. Sci., Univ. California, Los Angeles, UCLA-ENG-7642, Apr. 1976.

[2] L. Kleinrock and F. Kamoun, "Hierarchical routing for large networks, performance evaluation and optimization," *Comput. Networks*, vol. 1, pp. 155–174, Jan. 1977.

[3] ——, "Stochastic performance evaluation of hierarchical routing for large networks," *Comput. Networks*, vol. 3, pp. 337–353, Nov. 1979.

[4] M.A. Rich and M. Schwartz, "Buffer sharing in computer-communication network nodes," in *Proc. ICC*, San Francisco, CA, June 1957, pp. 33-17–33-20.

[5] D.L. Drukey, "Finite buffers for purists," TRW Inc., Redondo Beach, CA, TRW Syst. Group Rep. 75.6400-10-97, 1975.

[6] M. Irland, "Queueing analysis for a buffer allocation scheme for a packet switch," in *Proc. NTC*, vol. 1, Nov. 1975, pp. 24-8–24-13.

[7] S. Lam, "Store-and-forward buffer requirements in a packet switching network," *IEEE Trans. Commun.*, vol. COM-24, pp. 394–403, Apr. 1976.

[8] R. Syski, *Introduction to Congestion in Telephone Systems*. Edinburgh and London: Oliver and Boyd, 1960.

[9] J.R. Jackson, "Networks of waiting lines," *Oper. Res.*, vol. 5, pp. 518–521, 1957.

[10] W.J. Gordon and G.F. Newell, "Closed queueing systems with exponential servers," *Oper. Res.*, vol. 15, pp. 254–265, 1957.

[11] F. Baskett, K. Chandy, R. Muntz, and F. Palacios, "Open, closed, and mixed networks of queues with different classes of customers," *J. Ass. Comput. Mach.*, vol. 22, pp. 248–260, Apr. 1975.

[12] L. Kleinrock, *Queueing Systems, Vol. I: Theory*. New York: Wiley-Interscience, 1975.

[13] J.W.N. Wong, "Queueing network models for computer systems," School Eng. Appl. Sci., Univ. California, Los Angeles, UCLA-ENG-7579, Oct. 1975.

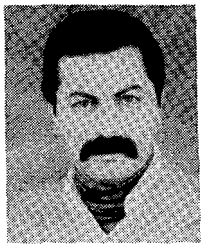
[14] F. Kamoun, and L. Kleinrock, "Analysis of shared storage in a computer network environment," in *Proc. 9th HICSS*, Honolulu, HI, Jan. 1976, pp. 80–92.

[15] J. Buzen, "Queueing network models of multiprogramming," Ph.D. dissertation, Div. Eng. Appl. Sci., Harvard Univ., Cambridge, MA, 1971.

[16] A. C. Williams and R. A. Bhandiwad, "A generating function approach to queueing network analysis of multiprogrammed computers," *Networks*, vol. 6, pp. 1–22, Jan. 1976.

[17] F. R. Moore, "Computational model of a close queueing network with exponential servers," *IBM J. Res. Develop.*, vol. 16, pp. 567–572, Nov. 1972.

★



Farouk Kamoun (M'78) received the degree in engineering from the Ecole Superieure d'Electricite Paris, France, in 1970, and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles, in 1972 and 1976, respectively.

From 1973 to 1976 he was with the University of California, Los Angeles, where he participated in the ARPA Network Project as a Postgraduate Research Engineer and did research on design considerations for large computer communication networks. In 1976 he joined the Faculte des Sciences de Tunis, Tunisa, where he is a Professor of Computer Science and Chairman of the Department of Computer Science.

★

Leonard Kleinrock (S'55–M'64–SM'71–F'73), for a photograph and biography, see p. 574 of the April 1980 issue of this TRANSACTIONS.