THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 61, No. 8, October 1982
Printed in U.S.A.

Stochastic Theory of a Data-Handling System with Multiple Sources

By D. ANICK,* D. MITRA, and M. M. SONDHI

(Manuscript received December 1, 1981)

In this paper we consider a physical model in which a buffer receives messages from a finite number of statistically independent and identical information sources that asynchronously alternate between exponentially distributed periods in the 'on' and 'off' states. While on, a source transmits at a uniform rate. The buffer depletes through an output channel with a given maximum rate of transmission. This model is useful for a data-handling switch in a computer network. The equilibrium buffer distribution is described by a set of differential equations, which are analyzed herein. The mathematical results render trivial the computation of the distribution and its moments and thus also the waiting time moments. The main result explicitly gives all the system's eigenvalues. While the insertion of boundary conditions requires the solution of a matrix equation, even this step is eliminated since the matrix inverse is given in closed form. Finally, the simple expression given here for the asymptotic behavior of buffer content is insightful, for purposes of design, and numerically useful. Numerical results for a broad range of system parameters are presented graphically.

I. INTRODUCTION

1.1 Physical model

A data-handling switch receives messages from many, say N, information sources, which independently and asynchronously alternate between the 'on' and 'off' state. The on periods as well as the off periods are exponentially distributed for each source. These two distributions, while not necessarily identical, are common to all sources;

* Presently at the Department of Mathematics, University of California, Berkeley.

also, the sources are mutually independent. Without loss of generality, the unit of time is selected to be the average on period; with this unit of time, the average off period is denoted by $1/\lambda$. Again, without loss of generality, the unit of information is chosen to be the amount generated by a source in an average on period. In these units an on source transmits at the uniform rate of 1 unit of information per unit of time. Thus, when r sources are on simultaneously, the instantaneous receiving rate at the switch is r. The switch stores or buffers the incoming information that is in excess of the maximum transmission rate, c, of an output channel. (Thus, c is also the ratio of the output channel capacity to an on source's transmission rate.)

As long as the buffer is not empty, the instantaneous rate of change of the buffer content is $r - c$. Once the buffer is empty, it remains so as long as $r \leq c$. We assume that the buffer is infinite and that the following stability condition is satisfied:

$$\frac{N\lambda}{c(1 + \lambda)} < 1. \tag{1}$$

The left-hand side is the traffic intensity, ρ.

Discussions with A. G. Fraser[1] suggest that the above is a useful model for a switch in a computer network. In such an application, the output channel rate may be in the range from 5 kb/s to 56 kb/s. For one specific source type, the slow terminals, the message rate may be taken to be 300 b/s, which gives 16 ⅔ and 186 ⅔ as the extreme values of c. A representative value of λ for this source type is 0.4, as computed from the fact that $\lambda/(1 + \lambda)$ is the long term on time fraction. The stability condition is satisfied with N as large as 3.5c. Other sources, such as screen terminals and computers, will have quite different statistics. In the interests of generality, we have not placed any further restrictions on the system parameters.

We first derive in a straightforward manner the set of differential equations that governs the equilibrium buffer distribution. We obtain a set of mathematical results that renders trivial the computation of the distribution and its moments and thus also the waiting time moments. The main result explicitly gives all the system's eigenvalues. This is achieved by using the generating function method. Insertion of the boundary conditions in the differential equations requires the solution of a matrix equation, which in many cases of practical interest is dimensionally quite large. However, even this step is eliminated since the matrix inverse is given in closed form. Finally, simple expressions for the moments of the distribution and the asymptotic behavior of buffer content are obtained.

The physical model described above is related to the model in our primary reference, a powerful paper by L. Kosten.[2] Kosten's model is a limiting case of our model with $N \to \infty$, $\lambda \to 0$ in a manner so as to give a finite traffic intensity, ρ. Also, pooling the instants of commencement of the on periods of all the sources yields, by assumption, a Poisson process. The above physical model and its variants have also been proposed in other papers.[3–6]

1.2 Motivations and discussion of results

Two broad questions supply most of our motivation. In this connection we acknowledge the benefit of several discussions with our colleague A. G. Fraser. The first question concerns the right buffer size to use for a predetermined number of sources and grade of service. The other question, which is of operational significance, concerns the selection of the maximum number of sources to be allowed in the system, the reasoning being that the incremental source disproportionately affects the grade of service for all the sources.

The study of these questions requires that the number of sources in the system, N, be finite. We also examined the conventional belief that the traffic intensity is a reliable indicator of overflow probabilities.

We would like to draw the reader's attention to an important aspect of our problem, namely, numerical stability. Underlying this problem is the fact proven below that the set of linear differential equations governing the behavior of the equilibrium probabilities [see eq. (8)] has 'unstable' eigenvalues. (This is cause for calling the system of equations 'inherently unstable.') Thus, if the boundary conditions are such that any of these modes are excited, then the solution grows at an exponential rate. In the mathematical model this does not happen. However, the situation during computation is quite different. The inevitable errors, no matter how small, incurred during numerical integration are liable to excite the unstable modes and lead to solutions that blow up.

The above observations apply as well when the Laplace transforms of the equilibrium probabilities are available and are to be numerically inverted. The boundedness of solutions is, in principle, obtained by the exact cancellation of unstable factors in the numerator and denominator of the transform. Of course, a straightforward numerical inversion cannot be expected to preserve this feature.

Therefore, it appears inevitable that any method that counters the inherent instability must depend on the a priori segregation of the stable modes from the unstable modes. This in turn depends on the availability of complete information on the eigenvalues and eigenvectors—information that is generally costly to obtain. We compose the solution to eq. (8) in the form

$$\mathbf{F}(x) = \sum_{i:\mathrm{Re}\, z_i \le 0} \mathbf{A}_i e^{z_i x}, \qquad x \ge 0, \tag{2}$$

where the z_i's appearing above are a subset of the eigenvalues, and the coefficient vectors $\{\mathbf{A}_i\}$ depend on the boundary conditions and eigenvectors. While the problem of numerical stability does not arise when the solution is computed in the above form, its effectiveness depends on the efficiency of the computation of the eigenvalues and coefficient vectors. An example of the efficiency achieved is that we are able to obtain all the eigenvalues by solving only a set of quadratic equations.

We should mention that Kosten[2] is fully cognizant of the problem of numerical stability. Kosten's solution method consists of obtaining the initial conditions and then numerically integrating the differential equation while continually filtering away the component of the numerical solution that exists in the span of the eigenvectors associated with the unstable eigenvalues.

A noteworthy feature of our primary solution method is that it manages to avoid requiring the numerical solution of matrix equations. This is achieved by avoiding the direct procedure (see Section III) which requires the solution of a dense set of $\lfloor c \rfloor + 1$ linear algebraic equations.[2]* Instead, we require the solution of a typically much larger set of $N - \lfloor c \rfloor$ linear algebraic equations which, however, we obtain in closed form.

Numerical results are discussed in Section VI. We observe substantial departures from Kosten's results in cases where N is small. More generally, we observe for identical traffic intensities, rather different probabilities of overflow for different values of N. We graphically demonstrate the quite acceptable quality of an approximation to the overflow probabilities provided by a relatively simple asymptotic formula. The formula states that, for x large, the probability of buffer content exceeding x behaves as Ae^{-rx}, A some constant and

$$r \triangleq \frac{(1-\rho)(1+\lambda)}{1-c/N}. \tag{3}$$

The positive parameter r is thus, like traffic intensity, a predictor of overflow probabilities. Small values of r may be associated with high probabilities of overflow and low grades of service.

1.3 Mathematical model

If at time t the number of on sources equals i, two elementary events can take place during the next interval Δt, i.e., a new source can start or a source can turn off. Since the on and off periods are exponentially distributed, the respective probabilities are $(N - i)\lambda\Delta t$ and $i\Delta t$. Compound events have probabilities $0(\Delta t^2)$. The probability of no change is $1 - \{(N - i)\lambda + i\}\Delta t + 0(\Delta t^2)$.

* We let $\lfloor c \rfloor$ denote the integer part of c. A tacit assumption is that $c < N$, since otherwise the buffer is always empty.

Let $P_i(t, x)$, $0 \le i \le N$, $t \ge 0$, $x \ge 0$, be the probability that at time t, i sources are on and the buffer content does not exceed x. Now,

$$P_i(t + \Delta t, x) = \{N - (i - 1)\}\lambda\Delta t P_{i-1}(t, x) + (i + 1)\Delta t P_{i+1}(t, x) + [1 - \{(N - i)\lambda + i\}\Delta t]P_i\{t, x - (i - c)\Delta t\} + 0(\Delta t^2). \quad (4)$$

Passing to the limit $\Delta t \to 0$:

$$\frac{\partial P_i}{\partial t} + (i - c)\frac{\partial P_i}{\partial x} = (N - i + 1)\lambda P_{i-1} - \{(N - i)\lambda + i\}P_i + (i + 1)P_{i+1}. \quad (5)$$

We are interested only in time-independent, equilibrium probabilities,

$$F_i(x) \triangleq \begin{array}{l}\text{equilibrium probability that } i \text{ sources are on}\\ \text{and buffer content does not exceed } x.\end{array} \quad (6)$$

Therefore, we set $\partial P_i/\partial t = 0$ and obtain, for $i \in [0, N]$,

$$(i - c)\frac{dF_i}{dx} = (N - i + 1)\lambda F_{i-1} - \{(N - i)\lambda + i\}F_i + (i + 1)F_{i+1}, \quad (7)$$

where the understanding is that $F_i = 0$ if i is not in the stated interval. In matrix notation,

$$\mathbf{D}\frac{d}{dx}\mathbf{F}(x) = \mathbf{MF}(x), \qquad x \ge 0, \quad (8)$$

where $\mathbf{D} = \text{diag}\{-c, 1 - c, 2 - c, \cdots, N - c\}$ and

$$\mathbf{M} = \begin{bmatrix} -N\lambda & 1 & & & & \\ N\lambda & -\{(N-1)\lambda + 1\} & 2 & & & \\ & (N-1)\lambda & -\{(N-2)\lambda + 2\} & 3 & & \\ & & \cdot & & & \\ & & \cdot & & & \\ & & \cdot & & & \\ & & \cdot & & & \\ & & & 2\lambda & -\{\lambda + (N-1)\} & N \\ & & & & \lambda & -N \end{bmatrix}.$$

Note that the initial conditions to the differential equations are as yet unspecified. Considerations relating to their determination may be found in Section III.

Let*

$$G(x) \triangleq Pr(\text{buffer content} > x) = 1 - \mathbf{1}'\mathbf{F}(x), \qquad x \ge 0. \quad (9)$$

We refer to $G(x)$ as the 'probability of overflow beyond x', or, loosely, as just the 'probability of overflow.' Also observe that

* We let **1** denote the vector with unity for all its components and prime denote transposition.

$$F_i(\infty) = \frac{1}{(1+\lambda)^N} \binom{N}{i} \lambda^i, \qquad 0 \le i \le N, \tag{10}$$

since $F_i(\infty)$ is the probability that i out of N sources are on simultaneously. Obviously,

$$\sum_{i=0}^{N} F_i(\infty) = 1.$$

In the analysis to follow, we assume that c is not an integer. (When c is an integer, one of the differential equations in (7) degenerates to an algebraic equation that may be used to eliminate one of the unknown components of $\mathbf{F}$.)

In Ref. 2 the elements of the matrix $\mathbf{M}$ below the diagonal are identical, i.e., independent of row number, and the diagonal element is accordingly adjusted to give column sum 0, as in (8).

The work of Arthurs and Shepp considers various models related to the one considered here; the emphasis of their analysis is on obtaining Laplace transforms of the probabilities.[6] Cohen[7] obtains a broad range of results for the case $c = 1$.

II. EIGENVALUES AND EIGENVECTORS

2.1 Computing the eigenvalues

Let z be some eigenvalue of $\mathbf{D}^{-1}\mathbf{M}$ and let ϕ be the associated right eigenvector. That is,

$$z\mathbf{D}\phi = \mathbf{M}\phi. \tag{11}$$

Equation (11) is also

$$z(i-c)\phi_i = \lambda(N+1-i)\phi_{i-1} - \{(N-i)\lambda + i\}\phi_i + (i+1)\phi_{i+1}, \qquad 0 \le i \le N. \tag{12}$$

Let $\Phi(x)$ denote the generating function of ϕ, i.e.,

$$\Phi(x) \triangleq \sum_{i=0}^{N} \phi_i x^i. \tag{13}$$

By multiplying (12) by x^i and summing over i we expect to obtain an equation in $\Phi(x)$ and $\Phi'(x)$ [for example, $\sum ix^i\phi_i = x\Phi'(x)$]. In fact,

$$\frac{\Phi'(x)}{\Phi(x)} = \frac{zc - N\lambda + N\lambda x}{\lambda x^2 + (z+1-\lambda)x - 1}. \tag{14}$$

In preparing to solve the differential equation, we define r_1 and r_2 to be the distinct real roots, $r_1 > 0 > r_2$, of the quadratic in the denominator of the right-hand side, i.e.,

$$r_1 = \{-(z+1-\lambda) + \sqrt{(z+1-\lambda)^2 + 4\lambda}\}/2\lambda \tag{15a}$$

$$r_2 = \{-(z+1-\lambda) - \sqrt{(z+1-\lambda)^2 + 4\lambda}\}/2\lambda. \tag{15b}$$

Equation (14) may now be written as

$$\frac{\Phi'(x)}{\Phi(x)} = \frac{c_1}{x - r_1} + \frac{c_2}{x - r_2}, \tag{16}$$

where the residues are computed to be

$$c_2 = N - c_1 \tag{17a}$$

$$c_1 = \frac{zc - N\lambda + N\lambda r_1}{\lambda(r_1 - r_2)}. \tag{17b}$$

The solution to (16) is

$$\Phi(x) = (x - r_1)^{c_1}(x - r_2)^{N-c_1}, \tag{18}$$

where, as in the rest of the paper, we have assumed $\phi_N = 1$.

There is an observation on (18) to be made that is central to the present derivation. Observe that by its definition in (13), $\Phi(x)$ is a polynomial in x of degree N. Since r_1 and r_2 are distinct, this is possible if and only if c_1, defined in (17b), is an integer in $[0, N]$. Denoting this integer by k we get

$$\Phi(x) = (x - r_1)^k(x - r_2)^{N-k}, \quad K = 0, 1, \cdots, N. \tag{19}$$

If in (17b) we write k for c_1, use (15) to substitute expressions for r_1 and $r_1 - r_2$, rearrange, and square, then we obtain the following family of quadratics in the unknown eigenvalue z,

$$A(k)z^2 + B(k)z + C(k) = 0, \quad k = 0, 1, \cdots, N$$

where,

$$\begin{aligned} A(k) &\triangleq (N/2 - k)^2 - (N/2 - c)^2 \\ B(k) &\triangleq 2(1 - \lambda)(N/2 - k)^2 - N(1 + \lambda)(N/2 - c) \\ C(k) &\triangleq -(1 + \lambda)^2\{(N/2)^2 - (N/2 - k)^2\}. \end{aligned} \tag{20}$$

We denote by $z_1^{(k)}$ and $z_2^{(k)}$ the two roots associated with the kth quadratic.

To recapitulate, we have shown that all the roots of the above family of $N + 1$ quadratics are eigenvalues, as defined in (11). The reader will find in the following section an enumeration of the properties of the roots and eigenvalues.

We observe that the above argument takes the place of the argument "$\Phi(x)$ should be an entire function of x" employed by Kosten.[2]

2.2 Properties of the roots of the quadratics

The theorem below, which is presented in conjunction with Fig. 1, is a collection of various properties of the roots.

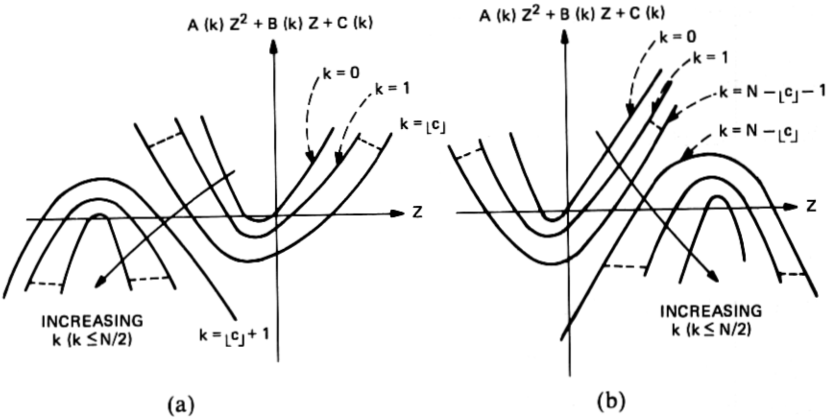

Fig. 1—Sketches of graphs of quadratics, N odd. If N is even, the quadratic for $k = N/2$ has a repeated real root. (a) $c < N/2$. (b) $N/2 < c$.

Theorem

(*i*) *The quadratics for k and k' are identical when $N/2 - k = k' - N/2$.*

(*ii*) *For each $k < N/2$ the corresponding quadratic has two real and simple roots. When N is even and $k = N/2$, the corresponding quadratic has a real repeated root.*

(*iii*) $k' < k \leq N/2 \Rightarrow A(k')z^2 + B(k')z + C(k') > A(k)z^2 + B(k)z + C(k), \quad \forall z.$

(*iv*) *The roots of the quadratic corresponding to any k are distinct from those of the quadratic corresponding to any k', provided $k' < k \leq N/2$.*

(*v*) *Ignoring the multiplicities, there are $N - \lfloor c \rfloor$ negative roots, 1 root at 0 and $\lfloor c \rfloor$ positive roots. (If the inequality in the stability condition is reversed, then there is 1 less negative root and 1 more positive root.)*

(*vi*) *The set of eigenvalues coincide exactly with the set of roots of the quadratics.*

(*vii*) *The largest negative eigenvalue is* $-(1 + \lambda - N\lambda/c)/(1 - c/N)$. □

The proof of the theorem is given in the appendix.

We will employ the following convention for the eigenvalues:

$$z_{N-\lfloor c \rfloor-1} < \cdots < z_1 < z_0 < z_N = 0 < z_{N-1} < \cdots < z_{N-\lfloor c \rfloor}. \tag{21}$$

With this convention, z_k and z_{N-k} are roots of the kth quadratic, i.e., $\{z_k, z_{N-k}\} = \{z_1^{(k)}, z_2^{(k)}\}$.

Since we shall later require the stable or negative eigenvalues, let us be explicit about their computation. Let N be first odd and employ the notation $z_1^{(k)} < z_2^{(k)}$. The stable subset is given in braces.

If $c < N/2$:

$$\left\{\begin{array}{l} z_1^{(k)} = [-B(k) - \sqrt{B^2(k) - 4A(k)C(k)}]/2A(k), \\ \qquad\qquad 0 \le k \le \lfloor c \rfloor. \\ z_{1,2}^{(k)} = [-B(k) \mp \sqrt{B^2(k) - 4A(k)C(k)}]/2A(k), \\ \qquad\qquad \lfloor c \rfloor + 1 \le k < N/2 \end{array}\right\} \quad (22a)$$

and if $N/2 < c$:

$$\{z_1^{(k)} = [-B(k) - \sqrt{B^2(k) - 4A(k)C(k)}]/2A(k), \quad 0 \le k \le N - \lfloor c \rfloor - 1.\}. \quad (22b)$$

When N is even, the only change is that the set in (22a) is augmented by $-N(1 + \lambda)/(N - 2c)$, which is one of the repeated roots of the quadratic associated with $k = N/2$.

2.3 Eigenvectors

We mention the procedure for obtaining the eigenvectors from the eigenvalues. Given an eigenvalue z, we compute in order, using (15) and (17), the quantities r_1, r_2, and k. These are used in (19) to yield the coefficients of the given polynomial and thus the eigenvector coefficients. Therefore, for the ith component of the eigenvector*

$$\phi_i = (-1)^{N-i} \sum_{j=0}^{k} \binom{k}{j} \binom{N-k}{i-j} r_1^{k-j} r_2^{N-k-i+j}, \qquad 0 \le i \le N. \quad (23)$$

Of particular subsequent interest are the eigenvectors $\boldsymbol{\phi}_N$ and $\boldsymbol{\phi}_0$, corresponding respectively to the eigenvalues $z_N = 0$ and $z_0 = (1 + \lambda - N\lambda/c)/(1 - c/N)$. The vector $\boldsymbol{\phi}_N$ may either be obtained directly or by noting that it is the appropriately normalized vector of equilibrium probabilities $\mathbf{F}(\infty)$ given in (10). Therefore,

$$\boldsymbol{\phi}_N = \frac{1}{\lambda^N} \left[1, N\lambda, \cdots, \binom{N}{i} \lambda^i, \cdots, \lambda^N \right]'$$

$$\text{and} \quad \mathbf{1}'\boldsymbol{\phi}_N = \left(\frac{1 + \lambda}{\lambda} \right)^N. \quad (24)$$

The vector $\boldsymbol{\phi}_0$ is obtained by following the procedure mentioned in the previous paragraph. We find that

$$r_1 = 1 - \frac{N}{c}, \qquad r_2 = \frac{1}{\lambda} \cdot \frac{1}{N/c - 1}, \qquad k = N,$$

$$\Phi(x) = \left\{ x + \left(\frac{N}{c} - 1 \right) \right\}^N. \quad (25)$$

* We let $\boldsymbol{\phi}_i$ *and* ϕ_i, respectively, denote the ith eigenvector and the ith component of the generic eigenvector.

Hence,

$$(\phi_0)_i = \binom{N}{i}\left(\frac{N}{c} - 1\right)^{N-i}, \qquad 0 \le i \le N \tag{26}$$

and

$$\mathbf{1}'\phi_0 = \left(\frac{N}{c}\right)^N.$$

This example serves to illustrate a noteworthy point: Recall that for a specific k we may obtain an eigenvalue z using (20); for the same z we may obtain k' from (15) and (17), as outlined in the first paragraph of this section. It may be that $k \neq k'$; however, it is always true that $|N/2 - k| = |N/2 - k'|$. This should not be surprising in view of statement (i) of the theorem, and it is due to the operations, including squaring, that allow us to go from (15) and (17) to (20).

2.4 Left eigenvectors

Sometimes, as in Section 3.3 below, not only the right, but also the left, eigenvectors ψ,

$$z\psi'\mathbf{D} = \psi'\mathbf{M} \tag{27}$$

are required to be known. It is reasonable to expect that the procedure in Section 2.1 can be repeated to obtain the generating function of the left eigenvectors, but we have not found this approach to be tractable. However, the procedure outlined below may be used.

There is a diagonal matrix τ that symmetrizes $\mathbf{M}$, i.e.,

$$\tau^{-1}\mathbf{M}\tau = (\tau^{-1}\mathbf{M}\tau)'. \tag{28}$$

In fact,

$$\tau_i = \left\{\lambda^i \binom{N}{i}\right\}^{1/2}, \qquad 0 \le i \le N, \tag{29}$$

where τ_i is the ith diagonal element of τ.

Define $\tilde{\psi}$ and $\tilde{\phi}$ to be the left and right eigenvectors obtained when $\mathbf{D}$ is replaced by $\tau^{-1}\mathbf{D}\tau$ and $\mathbf{M}$ by $\tau^{-1}\mathbf{M}\tau$. It is easy to see from the defining relations that

$$\tilde{\phi} = \tau^{-1}\phi \quad \text{and} \quad \tilde{\psi} = \tau\psi. \tag{30}$$

Now notice the important fact that since $\tau^{-1}\mathbf{M}\tau$ is symmetric, it has identical left and right eigenvectors. Hence,

$$\tau^2\psi = \phi$$

or, component-wise,

$$\boxed{\lambda^i \binom{N}{i} \psi_i = \phi_i, \qquad 0 \le i \le N.} \tag{31}$$

Thus, the left eigenvector may be obtained from the right eigenvector.

III. THE SOLUTION

The solution to the differential equations in (8) with $\mathbf{F}(0) = \mathbf{f}$, can be written as

$$\mathbf{F}(x) = \sum_{i=0}^{N} e^{z_i x} \frac{\boldsymbol{\psi}_i' \mathbf{D} \mathbf{f}}{\boldsymbol{\phi}_i' \mathbf{D} \boldsymbol{\psi}_i} \boldsymbol{\phi}_i. \tag{32}$$

It is clear, however, that as only bounded solutions are allowed,

$$\mathbf{F}(x) = \mathbf{F}(\infty) + \sum_{i=0}^{N-\lfloor c \rfloor -1} e^{z_i x} \frac{\boldsymbol{\psi}_i' \mathbf{D} \mathbf{f}}{\boldsymbol{\phi}_i' \mathbf{D} \boldsymbol{\psi}_i} \boldsymbol{\phi}_i, \tag{33}$$

where, according to our convention in (21), the z_i's appearing in the above expression are all negative. The term $\mathbf{F}(\infty)$ in (33) is identical to the $i = N$ term in (32). Recall that $\mathbf{F}(\infty)$ is already known, as shown in eq. (10). With appropriate identification, (33) also may be written as

$$\mathbf{F}(x) = \mathbf{F}(\infty) + \sum_{i=0}^{N-\lfloor c \rfloor -1} e^{z_i x} a_i \boldsymbol{\phi}_i. \tag{34}$$

Our primary solution method developed below in Sections 3.1 and 3.2 depends on the explicit solution of the coefficients a_i in the form appearing in (34). In Section 3.3 we give a second, contrasting, method in which the initial condition vector $\mathbf{f}$ is numerically solved and substituted in (33).

3.1 A key property of the solution at the boundary x = 0

If the number of sources on at any time exceeds c, then the buffer content increases and the buffer cannot stay empty. It follows that

$$F_i(0) = 0, \qquad \lfloor c \rfloor + 1 \le i \le N. \tag{35}$$

By supplementing (35) with the tri-diagonal structure of the matrix $\mathbf{D}^{-1}\mathbf{M}$ we may make further deductions regarding the behavior of $\mathbf{F}(x)$ when x is small. Observe that an application of $\mathbf{D}^{-1}\mathbf{M}$ on $\mathbf{F}(0)$ will diminish by 1 the number of trailing elements that are zero, and that each additional application will have the same effect until $(\mathbf{D}^{-1}\mathbf{M})^{N-\lfloor c \rfloor -1}\mathbf{F}(0)$ has only its last component equal to zero and $(\mathbf{D}^{-1}\mathbf{M})^{N-\lfloor c \rfloor}\mathbf{F}(0)$ has none. Thus,

$$\{(\mathbf{D}^{-1}\mathbf{M})^{j}\mathbf{F}(0)\}_i = 0, \qquad \lfloor c \rfloor + 1 + j \le i. \tag{36}$$

Now recollect that on account of the governing differential equations for $\mathbf{F}(\cdot)$ in (8),

$$\mathbf{F}^{(j)}(0) = (\mathbf{D}^{-1}\mathbf{M})^{j}\mathbf{F}(0). \tag{37}$$

Thus, from (36) we find that

$$F_i^{(j)}(0) = 0, \qquad \lfloor c \rfloor + 1 + j \le i, \tag{38}$$

and, in particular, the following relation, which we shall find most useful:

$$F_N^{(j)}(0) = 0, \qquad j = 0, 1, \cdots, N - \lfloor c \rfloor - 1. \tag{39}$$

Thus, not only is the event "all sources are on and buffer is empty" of probability zero, as already is known from (35), but the growth of the probability is also slow when the buffer content is small.

3.2 Procedure for obtaining the solution

We proceed to obtain the coefficients $\{a_i\}$ in the solution expression (34). Recall that by convention $\{\phi_i\}_N = 1$, so that from (34) we find that

$$F_N(x) = \left(\frac{\lambda}{1+\lambda}\right)^N + \sum_{i=0}^{N-\lfloor c \rfloor -1} a_i e^{z_i x}, \qquad x \geq 0. \tag{40}$$

The above, taken with (39), implies the following set of equations:

$$\sum_{j=0}^{N-\lfloor c \rfloor -1} (z_j)^i a_j = -\left(\frac{\lambda}{1+\lambda}\right)^N \delta_{oi}, \qquad 0 \leq i \leq N - \lfloor c \rfloor - 1. \tag{41}$$

Equation (41) in matrix form is

$$\mathbf{Va} = -\left(\frac{\lambda}{1+\lambda}\right)^N \mathbf{e}, \tag{42}$$

where $V_{ij} = (z_j)^i$, $\mathbf{a} = (a_0, a_1, \cdots, a_{N-\lfloor c \rfloor -1})'$ and $\mathbf{e} = (1, 0, \cdots, 0)'$.

The key observation is that $\mathbf{V}$ is a Vandermonde matrix. Well known results on such matrices allow us to solve (42) explicitly. Note that $\mathbf{V}$ is nonsingular because the eigenvalues $\{z_i\}$ are distinct,[8] as previously established in the theorem (see Section 2.2). Therefore,

$$|\mathbf{V}| = \prod_{0 \leq i \leq j \leq N - \lfloor c \rfloor -1} (z_i - z_j). \tag{43}$$

This formula, applied to the minors, which also are related to Vandermonde matrices, gives

$$\boxed{a_j = -\left(\frac{\lambda}{1+\lambda}\right)^N \prod_{\substack{i=0 \\ i \neq j}}^{N-\lfloor c \rfloor -1} \frac{z_i}{z_i - z_j}, \qquad 0 \leq j \leq N - \lfloor c \rfloor - 1.} \tag{44}$$

To summarize, the above procedure for obtaining the equilibrium probabilities $\mathbf{F}(x)$ and the probability of overflow $G(x)$ is based on using the expressions

$$\mathbf{F}(x) = \mathbf{F}(\infty) + \sum_{i=0}^{N-\lfloor c \rfloor -1} e^{z_i x} a_i \boldsymbol{\phi}_i \tag{45}$$

and

$$G(x) = -\sum_{i=0}^{N-\lfloor c \rfloor -1} e^{z_i x} a_i (\mathbf{1}' \boldsymbol{\phi}_i). \tag{46}$$

Only the stable eigenvalues appear in the above forms and they are explicitly given in (22); $\{\phi_i\}$ is obtained from the generating function in (18) and the coefficients $\{a_i\}$ appear in (44). Note that $\mathbf{1}'\boldsymbol{\phi} = \Phi(1)$, so that $\{\phi_i\}$ need not be computed explicitly to compute $G(x)$. $\mathbf{F}(\infty)$ is given in (10).

3.3 An alternative procedure for obtaining the solution

Under certain conditions the following procedure may be considered a viable alternative to the one described above. Here we compute $\mathbf{F}(0) = \mathbf{f}$ and use it in the solution series (33).

Recall from (35) that only the leading $\lfloor c \rfloor + 1$ elements of $\mathbf{f}$ are nonzero and need to be computed. Precisely the same number of equations are forthcoming by requiring of the initial conditions that the $\lfloor c \rfloor$ unstable modes are not excited (compare with Kosten's "illegal eigenvalues"[2]) and that $\mathbf{F}(\infty)$ is normalized, i.e.,

$$\left.\begin{aligned} \psi_i'\mathbf{D}\mathbf{f} &= 0 \qquad i = N-1, N-2, \cdots, N-\lfloor c \rfloor \\ \psi_0'\mathbf{D}\mathbf{f} &= -c + \frac{N\lambda}{1+\lambda} \end{aligned}\right\}. \tag{47}$$

In matrix form,

$$\left.\begin{aligned} \mathbf{f}' &= \{\mathbf{f}_1', \mathbf{0}'\} \\ \text{and } \mathbf{A}\mathbf{f}_1 &= \left(-c + \frac{N\lambda}{1+\lambda}\right)\mathbf{e} \end{aligned}\right\}, \tag{48}$$

where the coefficients in (47) multiplying the unknown $\mathbf{f}$ have been arranged to form the matrix $\mathbf{A}$ in (48). We note parenthetically that $\mathbf{A}$ is not sparse. The $(\lfloor c \rfloor + 1)$-dimensional matrix equation in (48) has to be solved.

Recall that our primary procedure in Section 3.2 and the one shown above require solutions of matrix equations. However, in the former case we were able to explicitly obtain the solution even though the dimension there, $N - \lfloor c \rfloor$, is typically much greater. In the absence of an explicit inverse for A, we expect the above procedure to be useful only for small c.

IV. ASYMPTOTICS

4.1 Probability of overflow

Here we examine the behavior of $G(x)$, the probability of overflow beyond x, for large values of x. The asymptotic formulas obtained are useful for the following reasons: As can be seen from the numerical results in Section VI, they often describe the system behavior rather well in all but the regions of lesser importance, where x is small; also, the analytic formulas are simple, even though they contain the essential information.

Since the form of the solution in (34) is a sum of exponential terms, the departure of $\mathbf{F}(x)$ from $\mathbf{F}(\infty)$ will be dominated by the exponential with the largest exponent. Hence,*

$$\mathbf{F}(x) - \mathbf{F}(\infty) \sim a_0 \phi_0 e^{-rx} \tag{49}$$

and

$$G(x) \sim -a_0(\mathbf{1}'\phi_0)e^{-rx}, \tag{50}$$

where $-r(=z_0)$ is the largest negative eigenvalue of $\mathbf{D}^{-1}\mathbf{M}$. Using statement (vii) of the theorem, we find that

$$r = \frac{(1 + \lambda - N\lambda/c)}{1 - c/N} = \frac{(1 + \lambda)(1 - \rho)}{1 - c/N} \tag{51}$$

where in the latter, more suggestive form, ρ is the traffic intensity. The coefficient a_0 and the eigenvector ϕ_0 are given in (45) and (26). Collecting terms, we find that

$$\boxed{G(x) \sim \rho^N \left\{ \prod_{i=1}^{N-\lfloor c \rfloor -1} \frac{z_i}{z_i + r} \right\} e^{-rx}.} \tag{52}$$

In the event that the alternative procedure given in Section 3.3 is used, then the following asymptotic formula is more relevant.

$$G(x) \sim Ae$$

where

$$A = \left(\frac{N - c}{N - 2c + c/\rho} \right)^N \left[\frac{\rho}{c(1 - \rho)} - \frac{1}{N - c} \right] \sum_{i=0}^{\lfloor c \rfloor} \frac{(c - i)f_i}{\lambda^i (N/c - 1)^i}. \tag{53}$$

4.2 Limiting equations for an infinite number of sources

Here we bridge some of the results presented in this paper and Kosten's results. We will show that some of Kosten's important expressions are obtained by passing to the appropriate limit, namely,

$$N \to \infty, \qquad \lambda \to 0 \quad \text{and} \quad N\lambda = \bar{\lambda}. \tag{54}$$

(Our notation is close to Kosten's with one notable exception: $\bar{\lambda}$ is Kosten's λ.) The results obtained in the limit may be interpreted to be applicable when the number of sources grows large and the fraction of time that each source is on decreases in such a manner that the traffic intensity approaches the fixed constant $\bar{\lambda}/c$.

In this section we follow Kosten and normalize the generating

---

* By $a(x) \sim b(x)$ we mean that $a(x)/b(x) \to 1$ as $x \to \infty$.

function to yield $\phi_0 = 1$ (rather than, as in the rest of the paper, $\phi_N = 1$), so that

$$\Phi(x) = \left(1 - \frac{x}{r_1}\right)^k \left(1 - \frac{x}{r_2}\right)^{N-k}, \tag{55}$$

where r_1 and r_2 are as in (15). It may be shown that, in the limit (54)

$$r_1 \to 1/(z+1) \tag{56}$$

$$r_2 \to -N(z+1)/\bar{\lambda}; \tag{57}$$

and, furthermore, on using (55),

$$\Phi(x) \to e^{x\bar{\lambda}/(z+1)}\{1 - x(1+z)\}^k, \tag{58}$$

which is Kosten's expression for the generating function.

To establish another correspondence, observe that on substituting (17b), the expression for r_1 and r_2 in (15), we obtain

$$k = \frac{zc - N\lambda + N\lambda[-(z+1-\lambda) + \sqrt{(z+1-\lambda)^2 + 4\lambda}]/2\lambda}{\sqrt{(z+1-\lambda)^2 + 4\lambda}}. \tag{59}$$

The limit of the right-hand side may be obtained and found to give Kosten's key equation

$$k = z(cz + c - \bar{\lambda})/(z+1)^2. \tag{60}$$

The correspondence in the eigenvector components is particularly illuminating, showing that

$$\phi_i = \lambda^i \sum_{j=0}^{k} \binom{k}{j}\binom{N-k}{i-j} (-\lambda)^{-j}(r_1)^{i-2j},$$

and, from (23), that

$$\phi_i \to \bar{\lambda}^i \sum_{j=0}^{k} \binom{k}{j} \frac{1}{(i-j)!} (-\bar{\lambda})^{-j} \left(\frac{1}{1+z}\right)^{i-2j}. \tag{61}$$

A final correspondence concerns the important rate parameter r appearing in the asymptotic formulas in Section 4.1. We find that

$$r = \frac{(1+\lambda)(1-\rho)}{1 - c/N} \to 1 - \bar{\rho}, \tag{62}$$

where $\bar{\rho}$ is simply the limiting traffic intensity $\bar{\lambda}/c$.

We should also mention that certain key results of this paper, such as those pertaining to our primary solution method in Sections 3.1 and 3.2, have no parallel in Ref. 2.

V. MOMENTS

We give below expressions for the moments of the equilibrium buffer content, which allows for their easy computation once the elements of the exponential series solution in (34) are available.

We observe that the nth moment

$$E(x^n) = \int_0^\infty x^n d\{\mathbf{1}'\mathbf{F}(x)\} = n \int_0^\infty x^{n-1} G(x) dx. \tag{63}$$

Since

$$G(x) = -\sum_{i=0}^{N-\lfloor c \rfloor -1} e^{z_i x} a_i (\mathbf{1}'\boldsymbol{\phi}_i),$$

as we saw in eq. (46), we obtain

$$E(x^n) = \frac{n!}{(-1)^{n+1}} \sum_{i=0}^{N-\lfloor c \rfloor -1} \frac{a_i(\mathbf{1}'\boldsymbol{\phi}_i)}{z_i^n}. \tag{64}$$

If the alternative procedure given in Section 3.3 is used to obtain the solution, then (64) may again be used with the identification

$$a_i = \frac{\boldsymbol{\psi}_i'\mathbf{D}\mathbf{f}}{\boldsymbol{\psi}_i'\mathbf{D}\boldsymbol{\phi}_i}. \tag{65}$$

VI. NUMERICAL RESULTS

In Figs. 2 through 5 we have held the source statistic λ to a constant

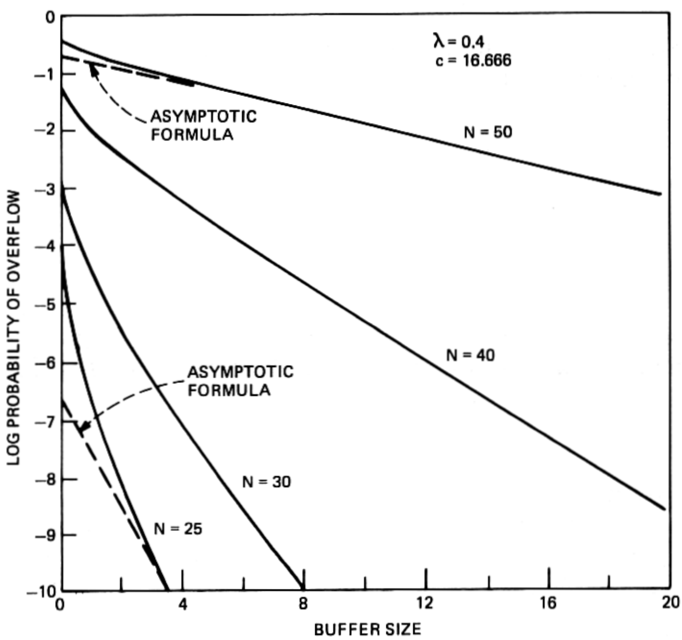

Fig. 2—Probability of overflow vs buffer size with λ and c constant. For $N = 25, 30, 40$, and 50, the traffic intensity ρ is 0.43, 0.51, 0.69, and 0.86, respectively.

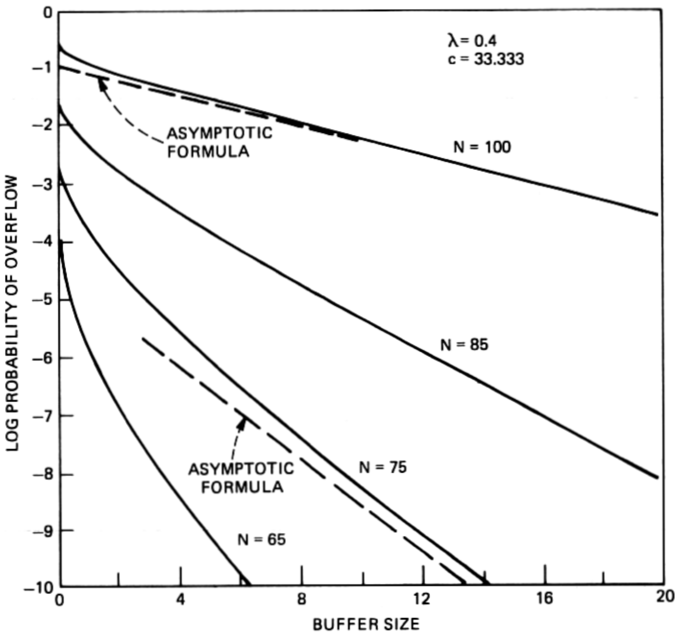

Fig. 3—Probability of overflow vs buffer size with λ and c constant. For $N = 65$, 75, 85, and 100, the traffic intensity ρ is 0.56, 0.65, 0.73, and 0.83, respectively.

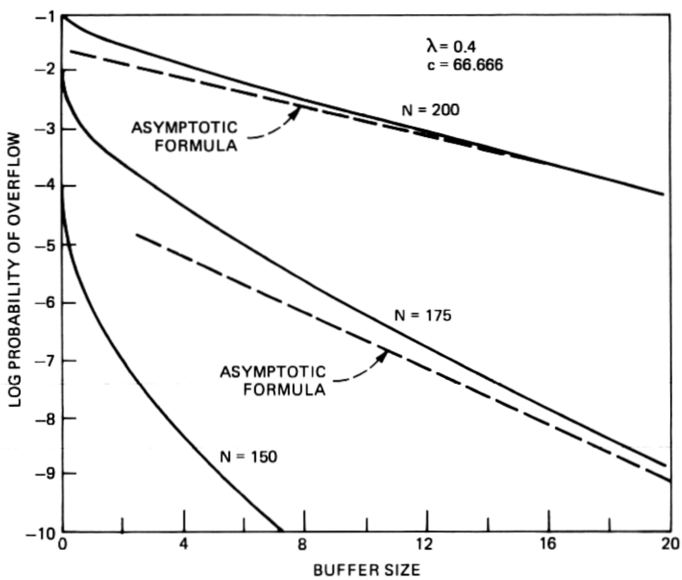

Fig. 4—Probability of overflow vs buffer size with λ and c constant. For $N = 150$, 175, and 200, the traffic intensity ρ is 0.64, 0.75, and 0.86, respectively.

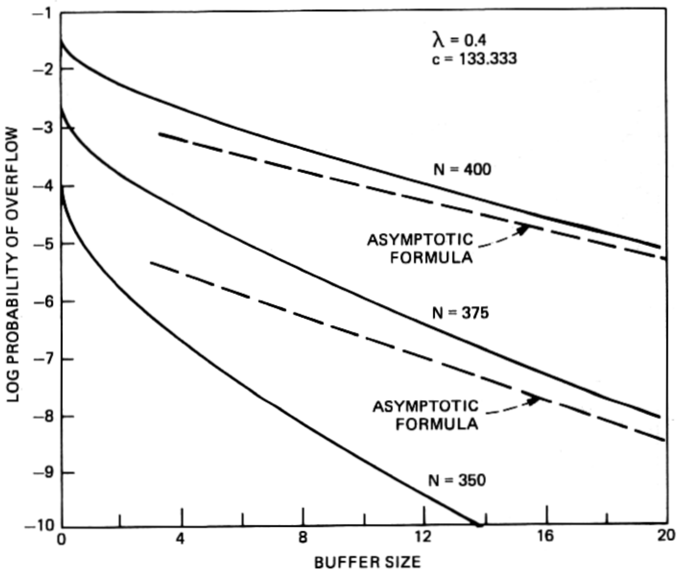

Fig. 5—Probability of overflow vs buffer size with λ and c constant. For $N = 350$, 375, and 400, the traffic intensity $\rho = 0.75$, 0.80, and 0.86, respectively.

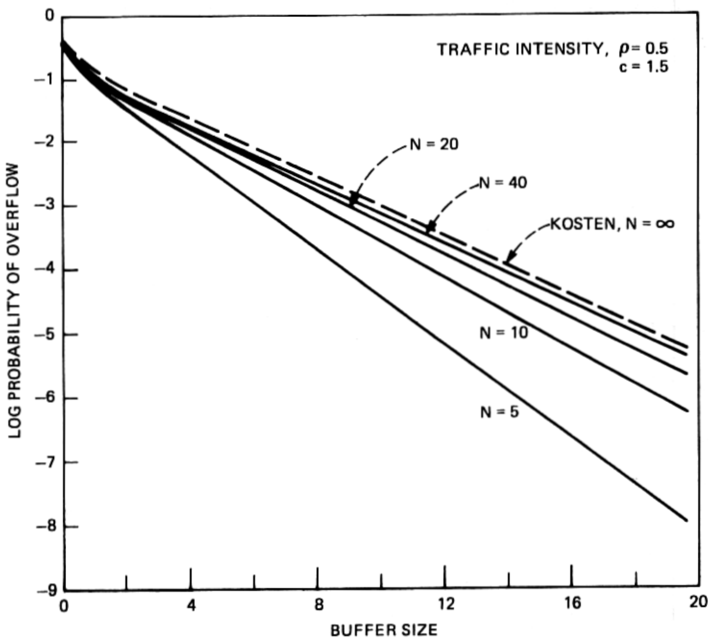

Fig. 6—Probability of overflow vs buffer size with traffic intensity ρ and c constant. For $N = 5$, 10, 20, and 40, the parameter λ is 0.18, 0.08, 0.04, and 0.02, respectively. The curve for $N = \infty$ is from Ref. 2.

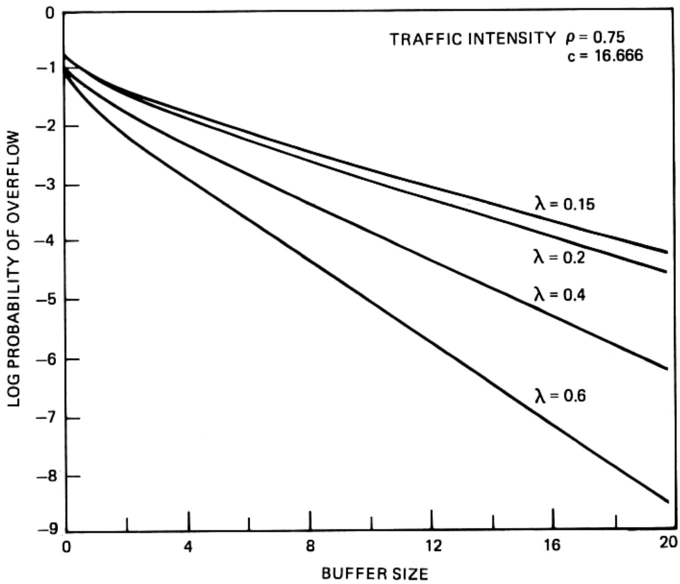

Fig. 7—Probability of overflow vs buffer size with traffic intensity ρ and c constant. For $\lambda = 0.6, 0.4, 0.2$, and 0.15. The number of sources $N = 33, 44, 75$, and 96, respectively.

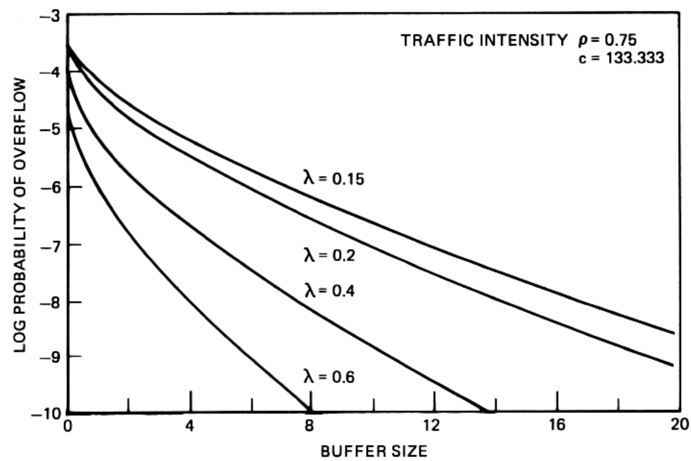

Fig. 8—Probability of overflow vs buffer size with traffic intensity ρ and c constant. For $\lambda = 0.6, 0.4, 0.2$, and 0.15, the number of sources $N = 267, 350, 600$, and 767, respectively.

value, 0.4. Each figure has a distinctive value of c, the ratio of output to input transmission rates. The four values of c are chosen for the cases where an on source transmits at 300 b/s and the output channel rates are 5 kb/s, 10 kb/s, 20 kb/s, and 40 kb/s. These figures show

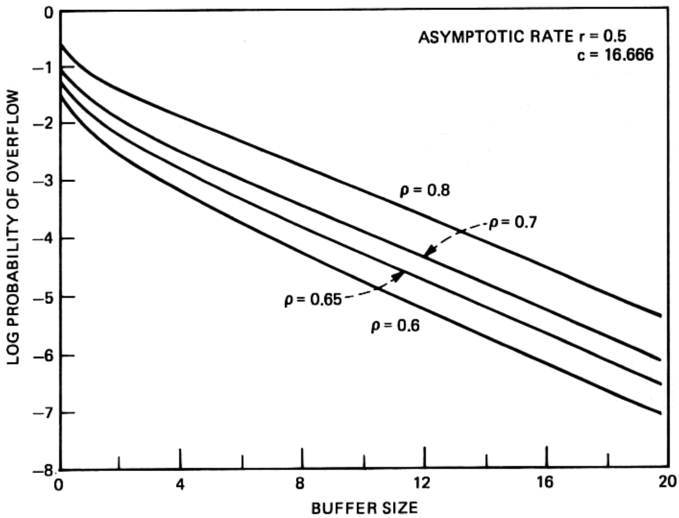

Fig. 9—Probability of overflow vs buffer size for constant asymptotic rate r [see eq. (51)] and c. For $\rho = 0.6$, 0.65, 0.7, and 0.8, the number of sources $N = 127$, 85, 63, and 41, respectively.

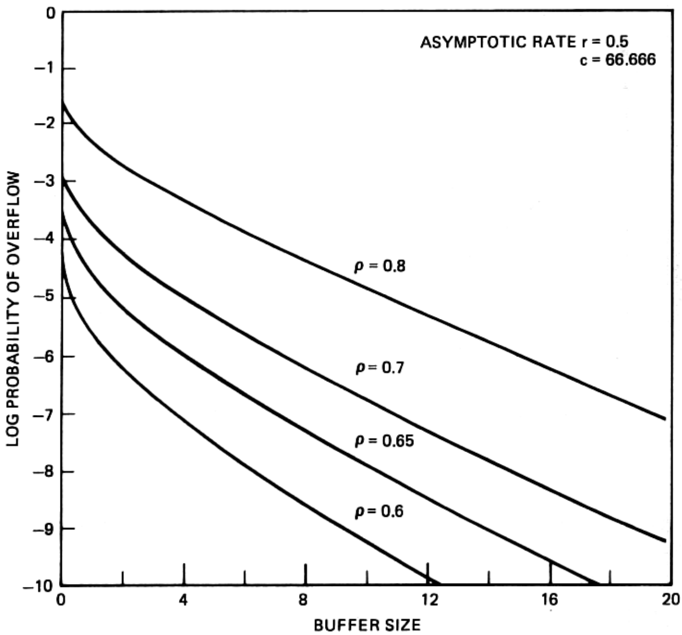

Fig. 10—Probability of overflow vs buffer size for constant asymptotic rate r and c. For $\rho = 0.6$, 0.65, 0.7, and 0.8, the number of sources $N = 507$, 338, 253, and 164, respectively.

rather clearly the effect of incremental sources. For example, we see from Fig. 2 that for a fixed grade of service as given by the probability of overflow = 10^{-5}, a 33-percent increase in the number of sources from 30 to 40 requires about a 300-percent increase in buffer size from 2 to 8. Recall from the discussion in Section 1.1 that the unit of information, and thus of the buffer as well, is the amount generated by one source in the average on period.

Observe in Figs. 2 through 5 the generally acceptable quality of the approximation to the probability of overflow provided by the asymptotic formula in eq. (52).

In each of Figs. 6 through 8 we have a constant traffic intensity, ρ, and a constant ratio of transmission rates, c. Note in particular that any two curves will have different source statistics λ and different N. The figures in this series illustrate the difference between the model considered in this paper and Kosten's limiting model. The figures also demonstrate rather emphatically the limitations of using only the traffic intensity as a predictor of overflow behavior. For example, in Fig. 7 we see that for constant traffic intensity and buffer size = 20, probability of overflow varies from about 10^{-4} to about 10^{-9}, depending on λ.

In Figs. 9 and 10 we examine the proposition that the rate parameter

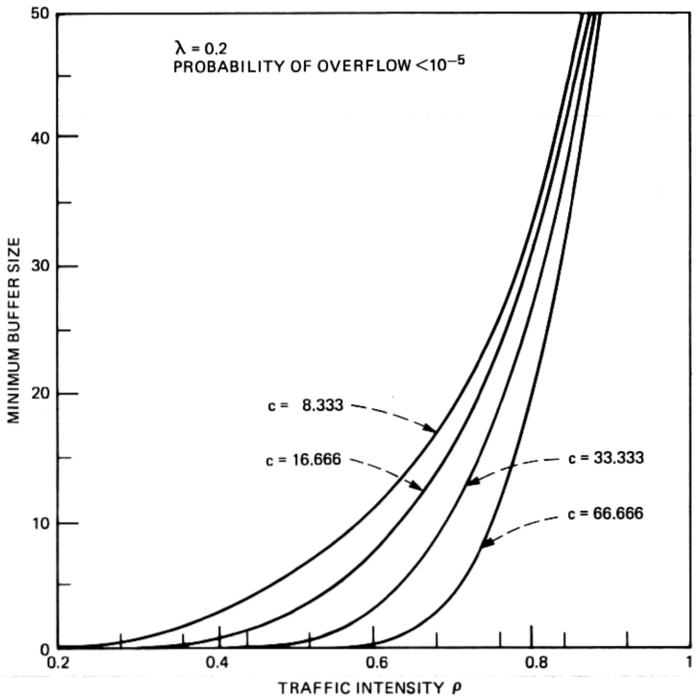

Fig. 11—Minimum buffer size required to satisfy traffic intensity ρ with constrained probability of overflow. The constant is λ.

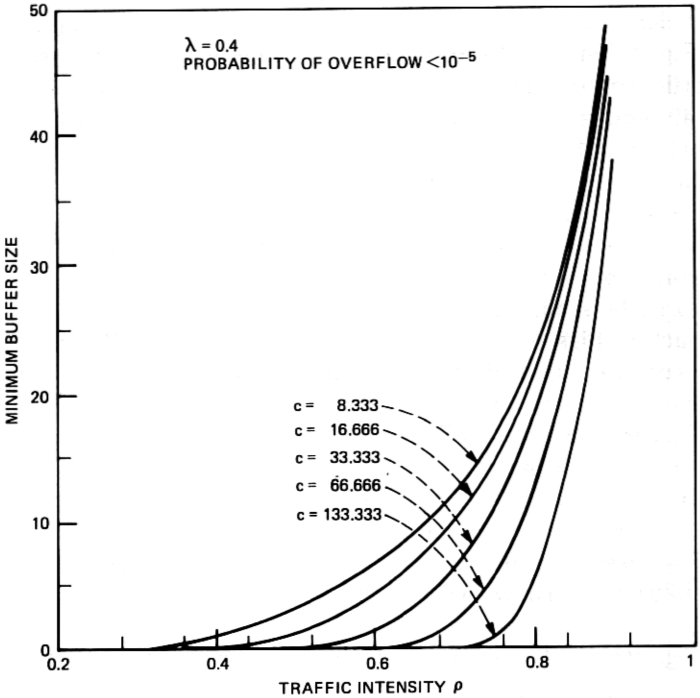

Fig. 12—Minimum buffer size required to satisfy traffic intensity ρ with constrained probability of overflow. The constant is λ.

r, which gives the slopes of the curves obtained from the asymptotic formula, is a useful single index of overflow behavior. Equations (3) and (51) give r. The strength and limitations of the index are contrasted in Figs. 9 and 10.

Figures 11 and 12 are motivated by the design problem in which it is required to estimate the buffer size needed to meet various traffic conditions with a specified grade of service and fixed source statistics. The slackening requirements in the buffer size with increasing c, which are observed in both figures, denote the economies of scale that stem from using higher capacity output channels. In comparing Figs. 11 and 12 we observe that for the same traffic intensity, buffer requirements are less stringent when λ is greater, i.e., when the source type is on for higher fractions of time.

REFERENCES

1. A. G. Fraser, private communication.
2. L. Kosten, "Stochastic Theory of a Multi-Entry Buffer (1)", Delft Progress Report, Vol. *1*, 1974, pp. 10–18.
3. H. Rudin, "Buffered Packet-Switching: A Queue with Clustered Arrivals", Int. Switching Symp. Rec., M.I.T., June 1972.

4. O. Hashida and M. Fujika, "Queueing Models for Buffer Memory in Store-and-Forward Systems", Int. Teletraffic Congress, Stockholm, June 1973, pp. 323/1–323/7.
5. W. W. Chu and L. C. Liang, "Buffer Behavior for Mixed Input Traffic and Single Constant Output Rate", IEEE Trans. on Commun., Vol. *COM-20*, No. 2 (April 1972), p. 230–35.
6. E. Arthurs and L. Shepp, unpublished work, 1979.
7. J. W. Cohen, "Superimposed Renewal Processes and Storage with Gradual Input," Stochastic Processes and Their Applications, Vol. *2* (January 1974), pp. 31–58.
8. R. Bellman, *Introduction to Matrix Analysis*, 2d ed., New York: McGraw-Hill, 1970, p. 193.

APPENDIX

Proof of Theorem

(*i*) Observe that $A(k)$, $B(k)$, and $C(k)$, as given in eq. (20), depend on k only through $(N/2 - k)^2$.

(*ii*) Roots are real if $B^2(k) - 4A(k)C(k) \geq 0$, and distinct as well if the expression is positive. View the expression as a function of $(N/2 - k)^2$. Observe that it is a quadratic in $(N/2 - k)^2$ and, furthermore, that it is concave since the leading coefficient of $(N/2 - k)^4$ is $4(1 - \lambda)^2 - 4(1 + \lambda)^2 < 0$. Thus, the minimum of $B^2(k) - 4A(k)C(k)$ for $(N/2 - k)^2$ in $[0, (N/2)^2]$ is at one of the corner points where the respective values are 0 and positive.

(*iii*) We claim that

$$\{A(k') - A(k)\}z^2 + \{B(k') - B(k)\}z + \{C(k') - C(k)\} > 0, \qquad \forall z. \tag{66}$$

To prove eq. (66) we need to observe that

$$A(k') - A(k) > 0 \tag{67}$$

and

$$\{B(k') - B(k)\}^2 - 4\{A(k') - A(k)\}\{C(k') - C(k)\} < 0. \tag{68}$$

(*iv*) It follows immediately from (*iii*) that the graphs of the quadratics are nonintersecting. See Fig. 1.

(*v*) We first consider $c < N/2$ and later its opposite. Consider in turn $k = 0$, $0 < k \leq c$, and $c < k \leq N/2$.

When $k = 0$, it turns out that $C(k) = 0$, so that the quadratic reduces to

$$A(0)z\{z + B(0)/A(0)\}. \tag{69}$$

Hence

$$z_1^{(0)} = -\frac{B(0)}{A(0)} = \frac{-(1 + \lambda - N\lambda/c)}{1 - c/N} \tag{70}$$

and

$$z_2^{(0)} = 0. \tag{71}$$

Observe that the stability condition implies that $z_1^{(0)} < 0$; reversal of the inequality in the stability condition gives $z_1^{(0)} > 0$.

Now consider $0 < k < c$. Observe that $A(k) > 0$ and $C(k) < 0$, so that $z_1^{(k)} < 0 < z_2^{(k)}$.

Finally, consider $c < k < N/2$. It is easy to see that $A(k) < 0$ and $C(k) < 0$. We show below that $B(k) < 0$, from which it will follow that $z_1^{(k)} < 0$ and $z_2^{(k)} < 0$. If $\lambda \geq 1$ then the defining expression for $B(k)$ in eq. (20) shows that $B(k) < 0$. If $\lambda < 1$ then the following equivalent expression shows that $B(k) < 0$:

$$B(k) = 2(1-\lambda)\left[\left(\frac{N}{2}-k\right)^2 - \left(\frac{N}{2}-c\right)^2\right] - 2c\left(\frac{N}{2}-c\right)\left(1-\lambda+\frac{N\lambda}{c}\right). \quad (72)$$

The above completes the considerations related to $c < N/2$.

For the opposite situation where $N/2 < c$, the case of $k = 0$ is unchanged. For $0 < k < N - c$, it is easy to see that $A(k) > 0$ and $C(k) < 0$, so that $z_1^{(k)} < 0 < z_2^{(k)}$.

Now consider $N - c < k < N/2$. We show below that $B(k) > 0$ (note the contrast with the situation for $c < N/2$), which when taken with $A(k) < 0$ and $C(k) < 0$, which are easy to see, gives $0 < z_1^{(k)} < z_2^{(k)}$. If $\lambda \leq 1$ then the defining expression for $B(k)$ in eq. (20) shows that $B(k) > 0$. Assume now that $1 < \lambda$. Then,

$$\begin{aligned} B(k) &= N(1+\lambda)\left(c-\frac{N}{2}\right) + 2(1-\lambda)\left(\frac{N}{2}-k\right)^2 \\ &> N(1+\lambda)\left(c-\frac{N}{2}\right) + 2(1-\lambda)\left(c-\frac{N}{2}\right)^2 \\ &= 2c\left(c-\frac{N}{2}\right)\left(1-\lambda+\frac{N\lambda}{c}\right) > 0. \end{aligned} \quad (73)$$

(*vi*) From our derivation of the quadratics it is clear that all roots are eigenvalues of $\mathbf{D}^{-1}\mathbf{M}$. As we have isolated exactly $(N+1)$ distinct values for the roots, there cannot be an eigenvalue that is not one of the roots.

(*vii*) It follows from (*iii*) and (*v*) that the largest negative eigenvalue is a root of the quadratic corresponding to $k = 0$. The negative root in this case is $-(1+\lambda-N\lambda/c)/(1-c/N)$, as shown in eq. (70).