A Decomposition Approximation for the Analysis of Voice/Data Integration

Sayeed Ghani and Mischa Schwartz, *Life Fellow, IEEE*

***Abstract*— A closed-form decomposition approximation for finding the data performance in voice/data queuing systems is presented. The approximation is based on Courtois' decomposition/aggregation techniques and is applied to Senet hybrid multiplexing, movable boundary frame allocation schemes. The approximation is applied to both infinite and finite buffer systems. In the former case the approximation is valid only in the underload region and serves as an upperbound for the mean data queuing delay. In the finite buffer case it is valid for the whole data traffic range and is shown to improve as the number of channels increase, and deteriorates as the buffer size increases. For finite buffer systems upper and lower bounds for the decomposition approximation have also been derived. It is found that the lower bound is tight in the underload and low traffic region of the overload. In these same regions the decomposition approximation serves as a tight upperbound.**

I. Introduction

IN THE ANALYSIS of integrated voice/data systems it is well known that exact solutions are computationally intractable for large systems. Hence, it is of interest to obtain approximation techniques for the analysis of such systems. This study presents an approximation technique which is applicable to such systems if the holding times of the traffic are widely different. In voice/data systems with no speech interpolation the ratio of voice to data holding times is typically of the order of 10^4 [1]. Similarly for video/voice systems this ratio can be large. Thus such integrated systems are suitable for a decomposition analysis.

The voice/data process is generally approximated by a two-dimensional Markov chain whose equilibrium stationary distribution is desired. The solution technique for such a problem falls broadly into two categories, exact and approximate techniques. Exact techniques include matrix methods and moment generating functions. The latter approach requires finding the roots of a polynomial within the unit circle. For large problems this poses convergence problems since the roots are generally close to one. Also finding roots of a polynomial is generally an ill-conditioned problem [2], hence the approach should be avoided. Exact matrix techniques give rise to very large matrices since the state space can be very large, and hence become computationally very expensive.

Approximation techniques include the fluid-flow and diffusion approximations, which have been applied to infinite data buffer systems [1], [3]. These approximations are inherently limited to the heavy traffic regions. In this paper we propose an approximate technique based on a decomposition technique originally developed by Simon and Ando [11]. The method was later applied by Courtois [4] to nearly completely decomposable matrices. The method developed by Courtois is essentially a matrix technique, but we show that when applied to typical voice/data integration it gives rise to closed form solutions. This is a significant advantage over previous exact and approximate methods mentioned above, all of which give solutions based on numerical techniques, such as finding roots of a polynomial or evaluating the eigenvalues and eigenvectors of matrices.

The decomposition method essentially approximates the steady state behavior of such systems by 'decomposing' them into long and short term behavior. Thus it converts a multidimensional Markov chain into a hierarchy of groups of states, such that the interaction *between* the groups is small compared to interaction *within* groups. Thus in obtaining the probability distribution of the system, the short term equilibrium distribution of a group is approximated by ignoring its interaction with other groups.

In order for the decomposition approximation to be accurate, it is necessary that each group of states should achieve equilibrium in isolation. In integrated voice/data systems, as we shall see later, if infinite size buffers are assumed, then this is only true in the *underload region,* which we shall define later as the region of flow-controlled data traffic. In the *overload region,* the data traffic region beyond the under load region, one or more of the subgroups becomes unstable in isolation, and hence the decomposition method for infinite buffers fails.

Thus for infinite data buffers the decomposition approximation is proposed as an underload approximation, a region where other approximations such as fluid flow give zero waiting time. The importance of the underload region stems from the fact that delays in this region are small compared to the overload region, where the delay is known to be proportional to α, the ratio of voice holding time to that of data. Thus for flow-controlled systems, this would be the desirable region of operation.

In order for the technique to be applicable for the whole data traffic region we consider finite data buffers, which is of

Paper approved by J. W. Wong, the Editor for Wide Area Networks of the IEEE Communications Society. Manuscript received May 11, 1989; revised September 4, 1991 and February 6, 1992. This work was supported by the National Science Foundation under Grant CDR-84-21402 and by the Office of Naval Research under Grant N00014-85-K0371. This paper was presented in part at the Proceedings Indo-U.S. Workshop on Systems and Signal Processing, Bangalore, India, January 1988.

S. Ghani is with NYNEX Science and Technology, Inc., White Plains, NY 10604 USA.

M. Schwartz is with the Department of Electrical Engineering and Center for Telecommunications Research, Columbia University, New York, NY 10027 USA.

IEEE Log Number 9401838.

course what real systems have. We show that in this case the decomposition approximation is a good approximation for the whole data traffic region, and is particularly good at very low and very high data loads.

Section II presents a description of the voice/data integration model known as the movable boundary scheme or Senet. In Section III we apply the Courtois decomposition technique to Senet, and present a closed form solution for the equilibrium stationary distribution and data waiting time. Section IV shows the results of Section III applied to larger systems and compared to exact and simulation results. In Section V we develop a technique for finding bounds on the decomposition error.

II. Voice/Data Integration

Since the early seventy's a great deal of research has been undertaken for the development of various architectures for the multiplexing of voice and data in an integrated voice/data networks. One of the proposed architectures for such a switch is Senet (slotted envelope network) [5]. It is based on the concept of a movable boundary scheme proposed by K. Kummerle of IBM Research, Zurich, in 1974 [6]. In a voice/data network such a multiplexor would exist at each node of the network. In the switching scheme we present below, the general problem of evaluating end to end voice/data performance is a very difficult one. This is due to the dependence of data performance on the voice traffic; and due to the node to node interdependence of circuit switched voice traffic. Thus it becomes necessary for the approximate solutions found for single node problems to have the following properties: that they are easily extendible to include more classes of traffic, such as video; and are also extendible to multimode problems. While keeping this eventual goal in mind, we presently restrict our analysis to a single node, with two classes of traffic voice and data. We show that the solution to this problem can be expressed quite accurately in closed form, and that the technique used is easily extendible to include more classes of traffic.

A. Senet Model and Analysis

The Senet frame structure is shown in Fig. 1. Time is divided into equal sized frames of period T seconds, and each frame consists of N TDM slots, each of size b bits. Thus, each slot represents a channel with a bit rate of b/T bits/second. Senet employs what is known as a movable boundary frame allocation, in which N_d channels are reserved for data transmission and the rest $N_v = N - N_d$ are shared between voice and data with voice having preemptive priority over data. Data packets preempted by voice are placed back at the head of the data queue. This technique can offer considerable advantage over a fixed boundary scheme in which data is limited to N_d channels. For the same data traffic load the movable boundary scheme gives reduced queuing delays with no reduction in voice performance. No good approximation techniques exist, however, for determining Senet performance in a closed form solution, and which may be useful for design purposes. This is the purpose of this study.

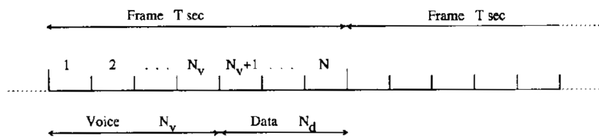

Fig. 1. Senet frame structure.

The multiplexor is hybrid, that is voice channels are circuit switched (with no buffering) whereas data is packet switched with a finite buffer. For the sake of simplicity we assume that the total number of data packets allowed in the system, i.e., those being served plus those in the buffer, is equal to M.[1] We assume that the frame size is small compared to the packet lengths and the duration of a circuit switched call. Voice call and packet arrivals are Poisson at average rates of λ_v calls/s and λ_d packets/s, respectively. Arriving voice calls and data packets are transmitted on individual channels at a bit rate of b/T. The holding time of a call and the time required to transmit a packet are assumed to be exponential with an average of $1/\mu_v$ and $1/\mu_d$ s, respectively. The data and voice utilizations, defined as $\rho_d = \lambda_d/\mu_d$ and $\rho_v = \lambda_v/\mu_v$, give us the average number of channels occupied by voice and data traffic, respectively. Using these definitions one can in turn define two regions of operation of the data traffic [3]. The underload region is the one in which the average data traffic, given by the data utilization ρ_d, is restricted to its reserved capacity only.

$$\text{Underload region:} \quad 0 \le \rho_d < N_d. \tag{2.1}$$

The overload region is the one in which the data load is large enough to require, on the average, voice-dedicated channels. For large voice holding times compared to data packet lengths this can result in extraordinarily long data queues and hence large data queuing delays if not limited by buffer size [3], [7].

$$\begin{aligned} \text{Overload region: for infinite} M \quad & N_d \le \rho_d < \rho_d^{\max} \\ & = N - \rho_v(1 - P_{BV}) \\ \text{for finite } M \quad & N_d \le \rho_d < \infty. \end{aligned} \tag{2.2}$$

The maximum data utilization $\rho_d^{\max}$ corresponds to the data load beyond which infinite data buffer systems become unstable. At this point the data load is equal to the average number of channels not occupied by the voice. P_{BV} is the voice blocking probability. In much of the analysis we take M to be finite. In some examples we assume infinite buffers.

Under the assumptions made above of Poisson arrivals and exponential service times, and if we also assume that the voice and data holding times are large compared to the frame period T, then the state of the system is given by a continuous-time two-dimensional Markov Process $(V(t), D(t))$ where $V(t) =$ number of voice calls in the system, and $D(t) =$ number of data packets in the system.

[1] We shall henceforth use the shorthand notation $N(N_v, N_d), M$ to denote the movable boundary scheme. Where M is not specified it is taken to be infinite.

We denote the equilibrium state probabilities by the two dimensional vector $\boldsymbol{p} = [p_{i,j}]$, i.e.,

$$p_{i,j} = \Pr(V(t) = i,\, D(t) = j),$$
$$i = 0, 1, \cdots, N_v, \qquad j = 0, 1, \cdots, M. \quad (2.3)$$

Define $q(i, j; k, l)$ as the transition rate from state (i, j) to (k, l). Using the arrival rate and service time parameters defined above, the nonzero transition rates are given by

$$\begin{array}{ll} q(i, j; i+1, j) = \lambda_v & i = 0, \cdots, N_v - 1 \\ q(i, j; i-1, j) = i\mu_v & i = 1, 2, \cdots, N_v \\ q(i, j; i, j+1) = \lambda_d & j = 0, \cdots, M-1 \end{array} \quad (2.4)$$

$$q(i, j; i, j-1) = \begin{cases} j\mu_d & j = 0, \cdots, N-i; \\ (N-i)\mu_d & j \geq N-i, \cdots, M-1. \end{cases}$$

The resulting state transition diagram for this system is shown in Fig. 2. The equilibrium probabilities $p_{i,j}$ must satisfy the global balance conditions. These can be written in matrix form by first mapping the two dimensional vector $\boldsymbol{p}$ into a one dimensional vector $\boldsymbol{v} = [v_l]$, by choosing some 1 : 1 mapping function

$$f : (i, j) \rightarrow l. \quad (2.5)$$

(An example appears in the next section). Thus, $v_l = p_{ij}$, for $l = f(i, j)$. The generator matrix $\boldsymbol{Q} = [Q_{k,l}]$ for the Markov process is then formed. The scalar elements $Q_{k,l}$ of $\boldsymbol{Q}$ are given by

$$Q_{k,l} = \begin{cases} q(i, j; m, n) \text{ if } k = f(i, j), l = f(m, n) & l \neq k; \\ -\sum_{i \neq k}^{i} Q_{k,i} & l = k \end{cases}. \quad (2.6)$$

The global balance conditions are then written as

$$\boldsymbol{vQ} = \boldsymbol{0}, \qquad \boldsymbol{v1} = 1 \quad (2.7)$$

where $\boldsymbol{0}$ and $\boldsymbol{1}$ are the appropriate size vectors of zeros and ones. Our goal is now to solve (2.7) for $\boldsymbol{v}$. Since $\boldsymbol{v}$ is a vector whose size n is given by the number of states in the two dimensional Markov chain (Fig. 2), $n = (N_v + 1)(M + 1)$, and $\boldsymbol{Q}$ is a square matrix of size $n \times n$. Hence, we have n linear simultaneous equations to solve. Solving these exactly by say Gaussian elimination would take $O(n^3)$ arithmetic operations. For narrowband T1 links N_v may be about 20, whereas M may be several hundred. For broad-band switching where we may have a movable boundary scheme involving say voice, data and video the number of states in the state space grows as a product of the number of states of each class of traffic. In addition if we have multiple nodes, the total state space again grows as a product of each individual node state space.

Hence clearly any sort of exact technique would quickly become computationally too expensive. Even numerical iterative techniques would have the same problem, besides giving no insight into the solution. Thus instead we exploit the structure of the generator matrix $\boldsymbol{Q}$. We assume that the ratio of the holding times, given by the symbol α, is large

$$\alpha = \frac{1/\mu_v}{1/\mu_d} \gg 1. \quad (2.8)$$

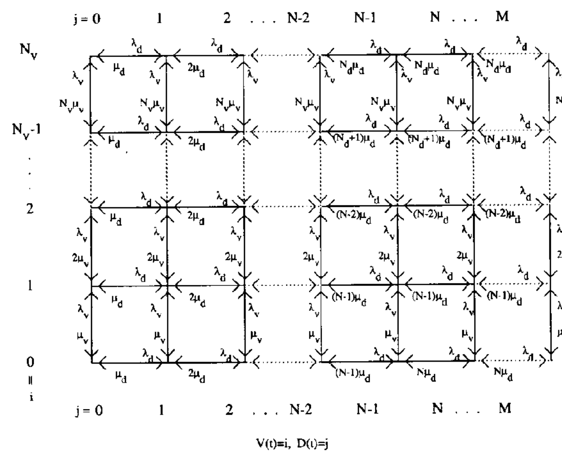

Fig. 2. Senet state transition diagram.

This immediately indicates that for each voice state i, the data states $\{(i, 0), (i, 1), \cdots, (i, M)\}$, for $i = 0, \cdots, N_v$, may achieve equilibrium. In this case the probability of being in state (i, j) is approximately given by

$$p_{i,j} \approx p(j|i)p_v(i) \quad (2.9)$$

where the following shorthand notation is used: $p(j|i) = \Pr(D(t) = j | V(t) = i)$, and $p_v(i) = \Pr(V(t) = i)$.

The conditional probability $p(j|i)$ is easily evaluated in closed form and is given by the probability of state j of a finite $M/M/N - i/M$ queue (where the last M refers to the maximum number of data packets allowed in the system). Similarly the exact probabilities $p_v(i)$ can also be written down exactly in closed form, and are given by

$$p_v(i) = \frac{\rho_v^i}{i!} / \sum_{k=0}^{N_v} \rho_v^k / k! \qquad i = 0, \cdots, N_v.$$

Having written down the decomposed solution following intuitive reasons we now ask the question, how good an approximation is (2.9)? Can we find error estimates, error bounds, and can we improve this first order approximation? As we shall show in this paper the answer to all these questions is in the affirmative and lies in showing that the solution (2.9) can also be obtained by using Courtois' decomposition/aggregation technique.

We will first give the Courtois Decomposition approximation in its general form in the next section, after which we will apply it to Senet.

III. Coutois' Decomposition Approximation

In this section we briefly summarize some of the results of the Courtois Decomposition approximation as described in [4]. The Courtois approximation is applicable to the following problem: Find the solution row vector $\boldsymbol{v}$ for a system of linear equations of the form

$$\boldsymbol{vP} = \boldsymbol{v}, \quad \boldsymbol{v1} = 1. \quad (3.1)$$

$\boldsymbol{P}$ is a $n \times n$ stochastic matrix. Let us define matrices $\boldsymbol{P}^*$ and $\boldsymbol{C}$ satisfying the following equations.[2]

$$\boldsymbol{P} = \boldsymbol{P}^* + \epsilon \boldsymbol{C} \tag{3.2}$$

$$\boldsymbol{P}^* = \mathrm{diag}\,(\boldsymbol{P}_I^*), \qquad I = 1, \cdots, N \tag{3.3}$$

with $\boldsymbol{P}_I^* = [P_{I_i J_j}^*]$ row stochastic square matrices defined to be of size $n(I) \times n(I)$. The elements $C_{I_i J_j}$ of $\boldsymbol{C}$ must each satisfy

$$|C_{I_i J_j}| \le 1, \qquad \forall I,\, J \text{ and } \forall i,\, j. \tag{3.4}$$

If an ϵ can be found such that

$$0 < \epsilon \ll 1 \tag{3.5}$$

then $\boldsymbol{P}$ is called near completely decomposable (NCD) [4]. The smallest ϵ is achieved by setting $\max |C_{I_i J_j}| = 1$. Hence ϵ then equals the largest element of the matrix $|\boldsymbol{P} - \boldsymbol{P}^*|$, i.e.

$$\epsilon = \max_{I_i J_j} |P_{I_i J_j} - P_{I_i J_j}^*|. \tag{3.6}$$

At this point we also describe a similar notation used for any eigenvector $\boldsymbol{v}$ of $\boldsymbol{P}$. Let $\boldsymbol{v}$ be a vector of subvectors $\boldsymbol{v}_I$

$$\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N] \tag{3.7}$$

where each $\boldsymbol{v}_I$ is a subvector of scalars v_{I_i}:

$$\boldsymbol{v}_I = [v_{I_1}, \cdots, v_{I_{n(I)}}] \qquad I = 1, \cdots, N. \tag{3.8}$$

Each $\boldsymbol{P}_I^*$ corresponds to a set of $n(I)$ states $\mathcal{A}_I = \{I_i,\, i = 1, \cdots, n(I)\}$. We call $\mathcal{A}_I$ an *aggregate state.* Thus the states $i = 1, \cdots, n$ have been partitioned into N aggregate states $\mathcal{A}_I$, $I = 1, \cdots, N$ each of size $n(I)$, such that transition probabilities between the aggregate states are weak compared to those within the aggregate states. Here ϵ represents the degree of coupling between the aggregate states. We note here that (3.2)–(3.6) imply that $\boldsymbol{P}^*$ can be constructed in an unlimited number of ways. We are merely required to distribute the off-diagonal block elements of $\boldsymbol{P}$ over the nonzero diagonal block elements of $\boldsymbol{P}$. One of these methods which we term "diagonal folding" will be discussed later.

If $\boldsymbol{P}$ is NCD then the Decomposition method approximates $\boldsymbol{v}$ by another vector $\boldsymbol{x}^*$, using the following decomposition/aggregation steps.[3]

Let $\boldsymbol{v}^*$ be a vector of the form (3.7, 3.8). First the solutions $\boldsymbol{v}_I^*$ of the N equations

$$\boldsymbol{v}_I^* \boldsymbol{P}_I^* = \boldsymbol{v}_I^*, \qquad \boldsymbol{v}_I^* \mathbf{1} = 1, \quad I = 1, \cdots, N \tag{3.9}$$

are found, either numerically or in closed form. Next the aggregate matrix $\boldsymbol{G}$ is formed. This is an $N \times N$ transition matrix for the aggregate states. $\boldsymbol{P}$ is called *lumpable* if there exists a set of constants $\hat{k}_{IJ}$, such that

$$\boldsymbol{P}_{IJ} \mathbf{1} = \hat{k}_{IJ} \mathbf{1}, \qquad \forall I,\, J = 1, \cdots, N \tag{3.10}$$

[2] The following N is not to be confused with the one used in Section II.

[3] Henceforth, we use the notation * to denote the decomposition approximation.

As we shall see later, the $\boldsymbol{P}$ matrix corresponding to Senet systems has this property. For lumpable systems, $\boldsymbol{G} = [G_{IJ}]$ can be found exactly by setting

$$G_{IJ} = \hat{k}_{IJ} \qquad I,\, J = 1, \cdots, N. \tag{3.11}$$

Using the matrix $\boldsymbol{G}$, the equilibrium state probability vector $\boldsymbol{X} = [X_I]$ for the aggregate states is given by the solution of the linear equations

$$\boldsymbol{X}\boldsymbol{G} = \boldsymbol{X}, \qquad \boldsymbol{X}\mathbf{1} = 1. \tag{3.12}$$

X_I is the probability of being in aggregate state I. For lumpable systems, $\boldsymbol{X}$ represents an exact solution.

The approximate solution $\boldsymbol{x}^*$ to $\boldsymbol{v}$, $\boldsymbol{x}^*$ having the form (3.7, 3.8), is given by

$$\boldsymbol{x}_I^* = \boldsymbol{v}_I^* \boldsymbol{X}_I. \tag{3.13}$$

Thus we have weighted the conditional distribution $\boldsymbol{v}^*$ by the probability X_I of being in the aggregate state I. Courtois has shown that (3.13) is an $O(\epsilon)$ approximation to $\boldsymbol{v}$ [4].

A. Reduction in Computational Complexity

In calculating the approximate solution (3.13) we have reduced the original problem (3.1) into the solution of $N+1$ smaller problems. N of these given by (3.9) are of size $n(I)$ each, and one of size N is given by (3.12). The importance of this decomposition lies in that often the solution of these smaller problems can be obtained in closed form, hence giving a closed form approximation to (3.1). This is the case in the voice/data integration considered next.

B. Courtois' Decomposition Applied to Senet

We stated that decomposition is applicable to systems whose solution is of the form (3.1) whereas our problem is in the form (2.7). Hence, we must first convert the generator matrix $\boldsymbol{Q}$ with solution of the form (2.7) into a transition probability matrix $\boldsymbol{P}$ with solution of the form (3.1) such that the state probability vector $\boldsymbol{v}$ remains unchanged. This can be done as follows. Define

$$\boldsymbol{P} = k\boldsymbol{Q} + \boldsymbol{I} \tag{3.14}$$

where k is a scalar constant satisfying $0 < k \le 1/\max_i |Q_{ii}|$, and hence $\boldsymbol{P} \ge \boldsymbol{0}$. Then $\boldsymbol{v}\boldsymbol{P} = k\boldsymbol{v}\boldsymbol{Q} + \boldsymbol{v}\boldsymbol{I} = \boldsymbol{v}$ from (2.7). Thus, $\boldsymbol{P}$ is the desired stochastic transition probability matrix.

We show later in (3.23) that $\boldsymbol{P}$ is NCD with

$$\epsilon \le \frac{1}{\alpha} \frac{N_v - 1 + \max(\rho_v, 1)}{\rho_d + N + \rho_v/\alpha} \equiv \epsilon_0. \tag{3.15}$$

The exact condition for $\boldsymbol{P}$ to be NCD, i.e., satisfying (3.5), would be for ϵ_0, given by (3.15), to satisfy

$$\epsilon_0 \ll 1. \tag{3.16}$$

We can simplify condition (3.16) for typical voice/data systems which have

$$\begin{aligned} 0 &< \rho_v < N_v, \\ 0 &< \rho_d < N. \end{aligned} \tag{3.17}$$

Fig. 3. Discrete time probability matrix P for $N(N_v, N_d), M$.

If

$$\alpha = \frac{1/\mu_v}{1/\mu_d} \gg \frac{N_v}{N} \tag{3.18}$$

then

$$\epsilon_0 \approx O\left(\frac{1}{\alpha}\frac{N_v}{N}\right) \ll 1. \tag{3.19}$$

We note that (3.18) clearly follows from (2.8), and hence (3.19) is a direct consequence of (2.8). Thus, (3.18) gives a simpler condition than (3.16) for $\boldsymbol{P}$ to be NCD.

Now let us look at the transition diagram of $\boldsymbol{Q}$, Fig. 2. (3.18) implies $\mu_d N \gg \mu_v N_v$. Under average loading conditions, i.e., $\rho_d \approx O(N_d)$, $\rho_v \approx O(N_v)$, (3.18) also implies $\lambda_d \gg \lambda_v$. This suggests that the state transition diagram can be naturally divided into groups $\mathcal{A}_i$ with $\mathcal{A}_i = \{(i, 0), (i, 1), \cdots, (i, M)\}$, for $i = 0, \cdots, N_v$ such that the interaction *between* groups (order of λ_v, μ_v) is weak compared to interaction *within* the groups (order of λ_d, μ_d).

Thus, in order to obtain an NCD matrix $\boldsymbol{P}$ having the form (3.2)–(3.4) we choose a mapping function f such that the elements within a group are mapped such that they are lumped together. Thus the following lexicographic mapping is chosen: $f(i, j) = Mi + j$. The resulting $\boldsymbol{P}$ matrix is shown in Fig. 3. From our definition of an NCD matrix (3.2), together with the conditions (3.17) and (3.18), it can now be seen that $\boldsymbol{P}$ is NCD with the number of subblocks N being $N_v + 1$. The size of each subblock $n(I) = M + 1$.

Having obtained $\boldsymbol{P}$ in NCD form the Decomposition method approximates $\boldsymbol{v}$ in the following Decomposition/Aggregation stages described in Section II.

C. Decomposition

$\boldsymbol{P}$ is decomposed into groups of irreducible Markov chains whose solution is obtained independently of the others. As mentioned earlier, there is no unique way of forming $\boldsymbol{P}^*$ from $\boldsymbol{P}$. Alternative methods of "folding" $\boldsymbol{P}$ into $\boldsymbol{P}^*$ have been investigated. At present we choose "diagonal folding" which, as we shall see, results in a closed form solution.

In diagonal folding, we form $\boldsymbol{P}^*$ by the following "perturbation" of $\boldsymbol{P}$. For every row of $\boldsymbol{P}$, collect all the nonzero terms outside the diagonal block and add them to the diagonal term. Thus,

$$P^*_{I_i I_i} = P_{I_i I_i} + \sum_{\substack{J=1\\ J\neq I}}^{N} \sum_{J=1}^{n(J)} P_{I_i J_j}, \qquad i = 1, \cdots, n(I), \quad I = 1, \cdots, N \tag{3.20}$$

$$P^*_{I_i I_j} = P_{I_i I_j}, \quad i, j = 1, \cdots, n(I), j \neq i, \quad I = 1, \cdots, N. \tag{3.21}$$

The matrix $\boldsymbol{P}^*$ is shown in Fig. 4.

$\boldsymbol{P} - \boldsymbol{P}^*$ is easily evaluated to be

$$P_{I_i J_j} - P^*_{I_i J_j} = \begin{cases} -(\lambda_v 1(N_v - I) + I\mu_v)k & I = J, \quad i = j \\ \lambda_v k & I = J - 1, \quad i = j \\ I\mu_v k & I = J + 1, \quad i = j \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

for $I, J = 0, \cdots, N_v$, $i, j = 0, \cdots, M$. Here

$$1(n) = \begin{cases} 1 & n \geq 1 \\ 0 & n < 1 \end{cases}.$$

$$P^* = \begin{pmatrix} P_0^* & & & \\ & P_1^* & & \\ & & \ddots & \\ & & & P_{N_v}^* \end{pmatrix}$$

$$\begin{array}{cc}
 & \begin{array}{cccccc} 0 & 1 & 2 & N & N+1 & M \end{array} \\
\begin{array}{c} 0\\ 1\\ 2\\ \vdots\\ N\\ N+1\\ \vdots\\ M-1\\ M \end{array} &
\begin{pmatrix}
1-\lambda_d k & \lambda_d k & 0 & & & \\
\mu_d k & 1-(\lambda_d+\mu_d)k & \lambda_d k & & & \\
0 & 2\mu_d k & 1-(\lambda_d+2\mu_d)k & & & \\
 & & \ddots & & & \\
 & & & 1-(\lambda_d+N\mu_d)k & \lambda_d k & \\
 & & & N\mu_d k & 1-(\lambda_d+N\mu_d)k & \\
 & & & & \ddots & \\
 & & & & & \lambda_d k \\
 & & & & & 1-N\mu_d k
\end{pmatrix}
\end{array}$$

$$P_0^*$$

$$\begin{array}{cc}
 & \begin{array}{ccccc} 0 & 1 & N-1 & N & M \end{array} \\
\begin{array}{c} 0\\ 1\\ 2\\ \vdots\\ N-1\\ N\\ \vdots\\ M-1\\ M \end{array} &
\begin{pmatrix}
1-\lambda_d k & \lambda_d k & & & \\
\mu_d k & 1-(\lambda_d+\mu_d)k & & & \\
0 & 2\mu_d k & & & \\
 & \ddots & & & \\
 & & 1-(\lambda_d+(N-1)\mu_d)k & \lambda_d k & \\
 & & (N-1)\mu_d k & 1-(\lambda_d+(N-1)\mu_d)k & \\
 & & & \ddots & \\
 & & & & \lambda_d k \\
 & & & & 1-(N-1)\mu_d k
\end{pmatrix}
\end{array}$$

$$P_1^*$$

$$\vdots$$

$$\begin{array}{cc}
 & \begin{array}{ccccc} 0 & 1 & N_d & N_d+1 & M \end{array} \\
\begin{array}{c} 0\\ 1\\ 2\\ \vdots\\ N_d\\ N_d+1\\ \vdots\\ M-1\\ M \end{array} &
\begin{pmatrix}
1-\lambda_d k & \lambda_d k & & & \\
\mu_d k & 1-(\lambda_d+\mu_d)k & & & \\
0 & 2\mu_d k & & & \\
 & \ddots & & & \\
 & & 1-(\lambda_d+N_d\mu_d)k & \lambda_d k & \\
 & & N_d\mu_d k & 1-(\lambda_d+N_d\mu_d)k & \\
 & & & \ddots & \\
 & & & & \lambda_d k \\
 & & & & 1-N_d\mu_d k
\end{pmatrix}
\end{array}$$

$$P_{N_v}^*$$

Fig. 4. P^* matrix for $N(N_\nu, N_d)$, M.

Once $\boldsymbol{P}^*$ is determined, ϵ and $\boldsymbol{C}$ follow from (3.6) and (3.2). Using (3.6) and (3.22) it is easy to show that

$$\begin{aligned}\epsilon &= \max_{I_iJ_j}|P_{I_iJ_j} - P^*_{I_iJ_j}| \\ &= \max\{(\lambda_v + (N_v-1)\mu_v)k,\ N_v\mu_v k\} \\ &= k\mu_v[N_v - 1 + \max(\rho_v, 1)].\end{aligned}$$

Now $0 < k \le 1/\max_i|Q_{ii}|$. $\max_i|Q_{ii}|$ is the maximum rate of leaving a state i, and by inspection of the $\boldsymbol{Q}$ matrix: $\max_i|Q_{ii}| \le \lambda_d + \lambda_v + N\mu_d$. Hence,

$$\begin{aligned}k &\le 1/(\lambda_d + \lambda_v + N\mu_d) \\ \epsilon &\le \frac{\mu_v[N_v - 1 + \max(\rho_v, 1)]}{(\lambda_d + \lambda_v + N\mu_d)} \\ &\le \frac{1}{\alpha}\frac{N_v - 1 + \max(\rho_v, 1)}{\rho_d + N + \rho_v/\alpha}.\end{aligned} \tag{3.23}$$

Since the matrices $\boldsymbol{P}_I^*$ represent a one dimensional birth–death process, the normalized equilibrium probability vectors of (3.9) are easily found to be given by

$$v^*_{I_j} = \begin{cases} P_{I_0}\rho_d^j/j! & 0 \le j < N-I \\ P_{I_0}\frac{\rho_d^j}{(N-I)!\,(N-I)^{j-(N-I)}} & N-I \le j \le M \end{cases} \tag{3.24}$$

where

$$P_{I_0} = \left(\frac{\rho_d^{(N-I)}(1-(\rho_d/(N-I))^{(M-N+I+1)})}{\left(1-\frac{\rho_d}{N-I}\right)(N-I)!} + \sum_{k=0}^{N-I-1}\rho_d^k/k!\right)^{-1} \qquad I = 0, \cdots, N_v.$$

These are the conditional solutions $p(j|I)$ to a finite $M/M/N-I/M$ queue, precisely as mentioned in Section II.

D. Aggregation

The solution of the decomposed Markov chains is combined to approximate the exact solution. Since we have assumed that voice calls have preemptive priority over data packets, the voice process is independent of the data process. Hence, it is easy to show that $\boldsymbol{P}$ is lumpable (see 3.11). Thus, the $(N_v+1)\times(N_v+1)$ aggregate matrix $\boldsymbol{G}$ is formed using (3.11). The matrix $\boldsymbol{G}$ is shown in Fig. 5. $\boldsymbol{X}$, the solution to (3.12), which represents the marginal distribution of the number of voice calls, is again easily obtained in closed form

$$X_I = \frac{\rho_v^I}{I!}\Big/\sum_{k=0}^{N_v}\rho_v^k/k! \qquad I = 0, \cdots, N_v. \tag{3.25}$$

This is precisely the form of $p_v(I)$ noted in Section II. Hence, the approximate solution $\boldsymbol{x}^*$ to $\boldsymbol{v}$ (3.13) is given in closed form as

$$\begin{aligned}x^*_{I_j} &= v^*_{I_j}X_I \qquad j = 0, \cdots, M, \quad I = 0, \cdots, N_v \\ &= p(j|I)p_v(I).\end{aligned} \tag{3.26}$$

Thus, we have shown that by choosing diagonal folding we get the decomposed solution (2.9). We can now find the relevant performance criteria. The waiting time, for example, is found, using Little's formula, to be given by

$$E^*(W) = \frac{1}{\mu_d}\left(\frac{E^*(D)}{\rho_d(1-P^*_{BD})} - 1\right). \tag{3.27}$$

Here P^*_{BD} is the decomposition approximation to the data blocking probability P_{BD}. P^*_{BD} is found by summing the probability the data buffer is full over all the voice states, i.e., $P^*_{BD} = \sum_{I=0}^{N_v} x^*_{I_M}$, and $\rho_d(1-P^*_{BD})$ is the data throughput. Of course for infinite M, P^*_{BD} is zero. The term $E^*(D)$ represents the approximate mean number of packets in the system, and is given by

$$E^*(D) = \sum_{I=0}^{N_v}\sum_{i=0}^{M} i x^*_{I_i} \tag{3.28a}$$

$$= \sum_{I=0}^{N_v} X_I \sum_{i=0}^{M} i v^*_{I_i} \tag{3.28b}$$

$$= \sum_{I=0}^{N_v} X_I E^*(D_I). \tag{3.28c}$$

$$
\begin{array}{c}
\\ 0 \\ 1 \\ 2 \\ \vdots \\ N_v-1 \\ N_v
\end{array}
\begin{array}{c}
\begin{array}{ccccccc}
0 & 1 & 2 & \cdots & N_v-2 & N_v-1 & N_v
\end{array}
\\
\begin{pmatrix}
1-\lambda_v k & \lambda_v k & & & & & \\
\mu_v k & 1-(\lambda_v+\mu_v)k & \lambda_v k & & & & \\
 & 2\mu_v k & 1-(\lambda_v+2\mu_v)k & \cdots & & & \\
 & & & \ddots & & & \\
 & & & & (N_v-1)\mu_v k & 1-(\lambda_v+(N_v-1)\mu_v)k & \lambda_v k \\
 & & & & & N_v\mu_v k & 1-N_v\mu_v k
\end{pmatrix}
\end{array}
$$

Fig. 5. G matrix for $N(N_\nu, N_d)$, M.

Here we have used (3.26), and defined $E^*(D_I) = E^*(D|V(t) = I)$, i.e.,

$$E^*(D_I) \equiv \sum_{i=0}^{M} i v^*_{I_i}. \tag{3.29}$$

$E^*(D_I)$ can be found by substituting (3.24) in (3.29). After some simplification this gives (3.30) shown below. For infinite buffer size, i.e., $M \to \infty$, with $\rho_d < N - N_v$, $E^*(D_I)$ reduces to (3.31) below. Results using these approximations appear in Section IV.

IV. Results

We now compare the decomposition results with exact and simulation results for various systems. The exact results were obtained, unless otherwise specified, by matrix iteration methods until sufficient accuracy was achieved in evaluating $\boldsymbol{v}$.

We first consider an infinite data buffer system 2(1, 1) using the notation $N(N_v, N_d)$ introduced in section 2. Fig. 6 shows the decomposition results obtained with (3.27), (3.28c) and (3.31) for the underload region, along with exact results for $\alpha = 1.1$, 10, 10^2, 10^3. $\rho_v = 0.5$ and $1/\mu_d = 10$ ms are kept constant. Hence, $1/\mu_v = 11$, 10^2, 10^3, 10^4 ms, respectively, for the four curves. We note that the data waiting times increase as α increases. This is as expected since voice calls preempt data packets on the N_v shared channels and the longer the holding time of voice calls the longer these data packets are delayed.

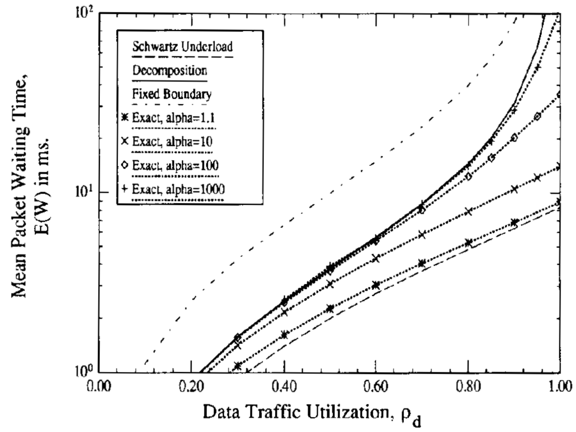

Fig. 6. Underload packet waiting time: 2(1, 1).

$$
\begin{aligned}
E^*(D_I) = P_{I_0}\Bigg\{ & \frac{\rho_d^{N-I+1}}{(N-I)!} \\
& \cdot \left[\frac{1 - \left(\frac{\rho_d}{N-I}\right)^{M-N+I+1} - (M-N+I+1)\left(\frac{\rho_d}{N-I}\right)^{M-N+I}\left(1 - \frac{\rho_d}{N-I}\right)}{(N-I)\left(1 - \frac{\rho_d}{N-I}\right)^2} \right] \\
& + \frac{\rho_d^{N-I+1}}{(N-I)!} \frac{\left[1 - \left(\frac{\rho_d}{N-I}\right)^{M-N+I}\right]}{1 - \frac{\rho_d}{N-I}} + \sum_{k=1}^{N-I} \frac{\rho_d^k}{(k-1)!} \Bigg\} \qquad I = 0, \cdots, N_v.
\end{aligned} \tag{3.30}
$$

$$
\begin{aligned}
E^*(D_I) = & \left(\frac{\rho_d^{N-I+1}}{(N-I)!} \left[\frac{1}{(N-I)} \frac{1}{\left(1 - \frac{\rho_d}{N-I}\right)^2} + \frac{1}{1 - \frac{\rho_d}{N-I}} \right] + \sum_{k=1}^{N-I} \frac{\rho_d^k}{(k-1)!} \right) \\
& \cdot \left(\frac{\rho_d^{N-I}}{\left(1 - \frac{\rho_d}{N-I}\right)(N-I)!} + \sum_{k=0}^{N-I-1} \rho_d^k / k! \right)^{-1} \qquad I = 0, \cdots, N_v.
\end{aligned} \tag{3.31}
$$

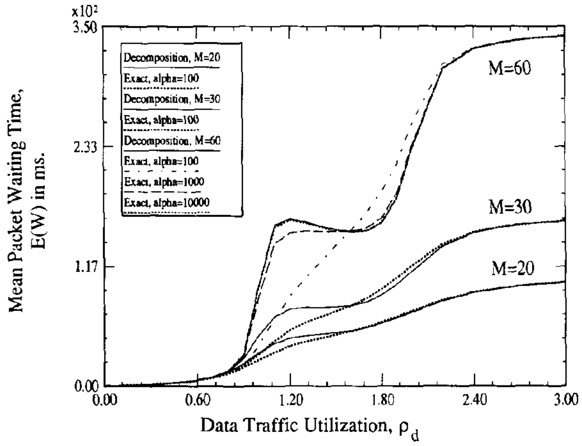

Fig. 7. Packet waiting time: 2(1, 1), $M = 20, 30, 60$.

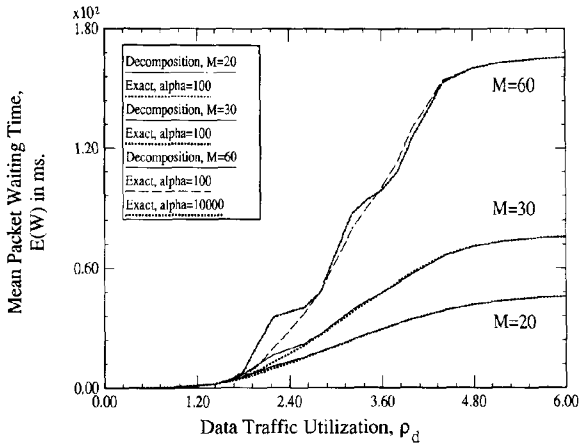

Fig. 8. Packet waiting time: 4(,2 2), $M = 20, 30, 60$.

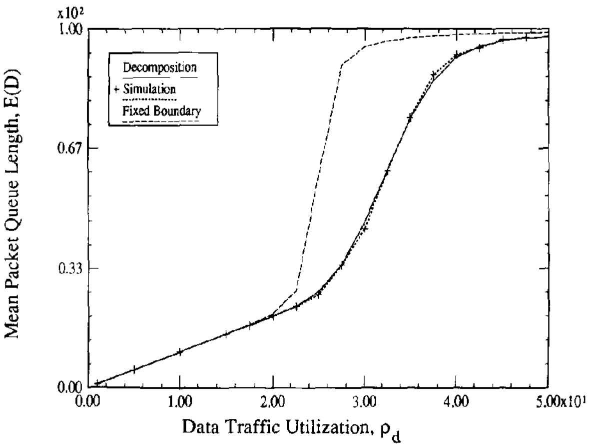

Fig. 9. Packet queue length: 50(25, 25), $M = 100$, $\alpha = 100$.

Since the decomposition approximation assumes the voice holding times are long enough for the data process to achieve complete equilibrium in each voice state, it effectively assumes $\alpha \rightarrow \infty$. We thus find that for infinite buffers the decomposition approximation serves as an upper bound. This is apparent in Fig. 6. As $\alpha \rightarrow \infty$ the exact curves approach the decomposition curves from below.

For comparison purposes we have plotted the fixed boundary curve which does not allow data packets to use any of the N_v voice channels. Clearly the movable boundary results are superior to those of fixed boundary results. The decomposition results are also clearly far tighter an upperbound than the fixed boundary result.

In the overload region for the infinite data buffer case the assumption of packet queue equilibrium in any given voice state, required to obtain approximation (2.9) or (3.26), is no longer valid. Thus in Fig. 6 as $\rho_d \rightarrow N - N_v = N_d$, the boundary between the underload and overload regions, the decomposition approximation is shown going to infinity. The exact curve is still valid, however, going out of bound only at the point $\rho_d = \rho_d^{\max} = 1.6666$ (this point is not shown in Fig. 6).

We also plot for comparison purposes in Fig. 6 an underload approximation proposed in [3]. We note that this serves as a lower bound, i.e., is approached as $\alpha \rightarrow 1$ from above.

We now consider finite data buffer systems. Fig. 7 shows three sets of curves for $M = 20, 30, 60$. For each set we plot the decomposition results along with exact curves for $\alpha = 10^2, 10^3, 10^4$. Again $\rho_v = 0.5$, and $1/\mu_d = 10$ ms. As before, as $\alpha \rightarrow \infty$, the exact curves approach the decomposition curves.

Figs. 8 and 9 show that the results grow more accurate for larger finite data buffer systems, 4(2, 2), 60 and 50(25, 25), 100. $1/\mu_d = 10$ ms for all the figures. $\rho_v = 0.7808$ for Fig. 8, which makes the normalized maximum number of channels available for data $\rho_d^{\max}/N_d$ the same (1.6666) as in Figs. 6 and 7. Fig. 8 shows that as we go to larger systems, from 2(1, 1) to 4(2, 2) for M and α fixed, the error in the approximation decreases. But for the same number of channels, as M increases, the error in the approximation increases. Fig. 9 shows results for a realistic sized system with 50 channels. $\rho_v = 17.25$ has been chosen as the voice utilization to make the voice blocking probability small: $P_{BV} = 1.78\%$. This gives $\rho_d^{\max} = 33.06$ channels for data. The decomposition results are compared to simulation results with 95% confidence interval. Clearly, the approximation is very accurate, as expected. For comparison purposes the mean delay for a fixed boundary scheme is also plotted.

V. Error Bounds

We had mentioned in Section III that the decomposition approximation is an $O(\epsilon)$ approximation, with a bound on ϵ given by (3.15). The constant associated with the error could be large, hence this result does not guarantee a small error even if ϵ is small. Hence what we need in order to guarantee a small error is a tight error bound. This is what we shall deal with here.

In this section we will use a technique different from Courtois' Decomposition approximation described in Section III, called *bounded aggregation*, to obtain two new approximations, a lower bound and an upper bound to the exact mean queue length. It will be shown that these two approximations provide a lower and upper bound to the decomposition approximation. Hence they can be used to obtain error bounds for the decomposition approximation. The bounded aggregation method was used by Courtois in [8] to obtain bounds to the equilibrium probability distribution vector $\boldsymbol{v}$. We will first obtain bounds to the mean queue length using $\boldsymbol{v}$ as obtained in [8], and then secondly using an improved method also based on bounded aggregation. It will be shown that the second method provides tighter bounds than the first. We first briefly summarize the relevant results as obtained in [8].

Define $\mathcal{B}(1, \boldsymbol{L})$ to be the set of all $n \times n$ nonnegative irreducible matrices $\boldsymbol{B}$ having spectral radius $\rho(\boldsymbol{B})$ less than or equal to 1, and a lower bound $n \times n$ matrix $\boldsymbol{L}$:

$$\mathcal{B}(1, \boldsymbol{L}) = \{\boldsymbol{B} \in \Re^{+}_{n\times n}: \rho(\boldsymbol{B}) \leq 1, \boldsymbol{0} \leq \boldsymbol{L} \leq \boldsymbol{B}, \boldsymbol{B} \text{ irreducible}\}. \quad (5.1)$$

We also require that $\rho(\boldsymbol{L}) < 1$.

It is shown in [8] that it is possible to easily calculate a polyhedron $\mathcal{P}$ that contains the left-positive eigenvector of any matrix $\boldsymbol{B} \in \mathcal{B}(1, \boldsymbol{L})$. Having obtained such a polyhedron, it is then possible to bound the left-positive eigenvector of all matrices $\boldsymbol{B} \in \mathcal{B}(1, \boldsymbol{L})$. We summarize these results briefly as follows. Define the normalized inverse, $\boldsymbol{Z} = \Sigma^{-1}(\boldsymbol{I} - \boldsymbol{L})^{-1}$, where $\boldsymbol{I}$ is the identity matrix and Σ is a normalization diagonal matrix diag $(\Sigma) = (\boldsymbol{I}-\boldsymbol{L})^{-1}\boldsymbol{1}$. Define the polyhedron $\mathcal{P}$ as the convex combination of the rows of the matrix $\boldsymbol{Z}$: $\mathcal{P} = \{\boldsymbol{\beta}^T\boldsymbol{Z}, \boldsymbol{\beta} \in \Re^{+}_{n}, \boldsymbol{\beta}^T\boldsymbol{1} = 1\}$. We now state without proof the following theorem from [8].

Theorem 1: The positive-left eigenvector of a matrix $\boldsymbol{B} \in \mathcal{B}(\lambda, \boldsymbol{L})$, belongs to the polyhedron the vertex of which are the rows of the matrix $\boldsymbol{Z}$, which have indices in the set

$$\mathcal{J} = \{j = 1, 2, \cdots, n; \exists i \quad \text{s.t.} B_{ij} > L_{ij}\}, \quad \text{if } \rho(\boldsymbol{B}) = \lambda. \quad (5.2)$$

The next theorem in [8] gives the two approximations $v_i^{\inf}$, $v_i^{\sup}$ to eigenvector $\boldsymbol{v} = [v_i]$; here $v_i^{\inf} \leq v_i \leq v_i^{\sup}$, for $i = 1, \cdots, n$. It states that

$$v_i^{\inf} = \max\left\{\min_{k\in\mathcal{J}}(Z_{ki}); 1 - \sum_{j\neq i}\max_{k\in\mathcal{J}}(Z_{kj})\right\} \quad (5.3a)$$

$$v_i^{\sup} = \min\left\{\max_{k\in\mathcal{J}}(Z_{ki}); 1 - \sum_{j\neq i}\min_{k\in\mathcal{J}}(Z_{kj})\right\}. \quad (5.3b)$$

Z_{ki} are the elements of $\boldsymbol{Z}$. In this paper we simplify (5.3a) and (5.3b), by using

$$v_i^{\inf} = \min_{k\in\mathcal{J}}(Z_{ki}) \quad (5.4a)$$

$$v_i^{\sup} = \max_{k\in\mathcal{J}}(Z_{ki}). \quad (5.4b)$$

We have found that this provides an occasionally less tight but much more simply calculated bound.[4]

We now apply the bounded aggregation method to nearly completely decomposable (NCD) systems, discussed previously in Section III. We first describe a method of obtaining exact results called *exact aggregation* as given in [8].

Consider a stochastic $\boldsymbol{P}$ matrix of the form given in (3.2) with the left eigenvector $\boldsymbol{v}$ of the form (3.7), (3.8). Define a stochastic block diagonal matrix $\overline{\boldsymbol{P}}$ of the form (3.3), i.e.,

$$\overline{\boldsymbol{v}}_I\overline{\boldsymbol{P}}_{II} = \overline{\boldsymbol{v}}_I, \qquad \overline{\boldsymbol{v}}_I\boldsymbol{1} = 1 \qquad I = 1, \cdots, N \quad (5.5)$$

such that $\overline{\boldsymbol{v}}_I = \boldsymbol{v}_I/\boldsymbol{v}_I\boldsymbol{1}$. For example, $\overline{\boldsymbol{P}}_{II}$ is a stochastic matrix whose left eigenvector $\overline{\boldsymbol{v}}_I$ is parallel to the exact subvector $\boldsymbol{v}_I$. Using the method of exact aggregation described briefly in [8], we can decompose the transition probability matrix $\boldsymbol{P}$ as follows:

$$\begin{pmatrix} \boldsymbol{P}_{II} & \boldsymbol{E}_{II} \\ \boldsymbol{F}_{II} & \boldsymbol{G}_{II} \end{pmatrix}$$

here $\boldsymbol{E}_{II}$ and $\boldsymbol{F}_{II}$ are the matrices of transition probabilities between $\boldsymbol{P}_{II}$ and the remaining part of the system $\boldsymbol{G}_{II}$. Thus we have $\overline{\boldsymbol{P}}_{II} = \boldsymbol{P}_{II} + \boldsymbol{E}_{II}(\boldsymbol{I}-\boldsymbol{G}_{II})^{-1}\boldsymbol{F}_{II}$. We define $\tilde{\boldsymbol{P}}_{II}$ as $\tilde{\boldsymbol{P}}_{II} = \overline{\boldsymbol{P}}_{II} - \boldsymbol{P}_{II} = \boldsymbol{E}_{II}(\boldsymbol{I}-\boldsymbol{G}_{II})^{-1}\boldsymbol{F}_{II}$. From (3.3) we can easily see that no row of $\boldsymbol{E}_{II}$ or column of $\boldsymbol{F}_{II}$ is completely null. Since $\boldsymbol{P}$ is irreducible $(\boldsymbol{I}-\boldsymbol{G}_{II})^{-1}$ is nonnegative. Hence $\tilde{\boldsymbol{P}}_{II}$ is a nonnegative matrix with no zero columns. Thus $\overline{\boldsymbol{P}}_{II} \geq \boldsymbol{P}_{II}$. Since the $\overline{\boldsymbol{P}}_{II}$ are stochastic, $\rho(\overline{\boldsymbol{P}}_{II}) = 1$. Hence, $\overline{\boldsymbol{P}}_{II} \in \mathcal{B}(1, \boldsymbol{P}_{II})$. Thus the exact eigenvector $\overline{\boldsymbol{v}}_I$, obtained by using $\overline{\boldsymbol{P}}_{II}$, can be bounded by using $\mathcal{B}(1, \boldsymbol{L})$ to obtain the bounds.

We now show that these bounds also bound the decomposition approximation. The decomposition method uses the submatrices $\boldsymbol{P}_I^*$ to solve for $\boldsymbol{v}_I^*$, as given in (3.9). The $\boldsymbol{P}_I^*$ are exactly equal to $\boldsymbol{P}_{II}$ except for the diagonal entries $\boldsymbol{P}_{I_{ii}}^*$, which are *increased* to make $\boldsymbol{P}_I^*$ stochastic. This observation can be verified by referring to (3.20) and (3.21). This clearly indicates that $\boldsymbol{P}_I^* \geq \boldsymbol{P}_{II}$. Since the $\boldsymbol{P}_I^*$ are stochastic $\boldsymbol{P}_I^* \in \mathcal{B}(1, \boldsymbol{P}_{II})$. Then since both $\boldsymbol{P}_I^*$ and $\overline{\boldsymbol{P}}_{II} \in \mathcal{B}(1, \boldsymbol{P}_{II})$, the bounds obtained by using $\mathcal{B}(1, \boldsymbol{P}_{II})$ bound both the exact and decomposition results, i.e., $\boldsymbol{v}^{\inf} \leq \boldsymbol{v}^*, \overline{\boldsymbol{v}} \leq \boldsymbol{v}^{\sup}$. Note that the aggregation step involves no approximation for lumpable systems.

Thus, we can find bounds to the subvectors $\boldsymbol{v}_I$ using (5.4a), (5.4b). We proceed as follows, by first calculating the index set $\mathcal{J}(I)$ given by (5.2) which for NCD systems becomes

$$\mathcal{J}(I) = \{j = 1, 2, \ldots n(I); \exists i \quad \text{s.t.} \overline{P}_{I_iI_j} > P_{I_iI_j}\}. \quad (5.6)$$

As shown earlier, $\tilde{\boldsymbol{P}}_{II} = \overline{\boldsymbol{P}}_{II} - \boldsymbol{P}_{II}$ has no zero columns, hence (5.6) simplifies to $\mathcal{J}(I) = \{j = 1, 2, \ldots n(I)\}$. Thus, knowing $\mathcal{J}(I)$ we can write down the solution as follows:

$$v_{I_i}^{\inf} = \min_{k\in\mathcal{J}(I)}(Z(I)_{ki}) \quad (5.7a)$$

$$\sup_{v_{I_i}} = \max_{k\in\mathcal{J}(I)}(Z(I)_{ki}) \quad (5.7b)$$

[4] Typically for systems considered in this paper (5.3a), (5.3b) and (5.4a), (5.4b) give the same bounds.

with $Z(I)_{ki}$ the elements of $\boldsymbol{Z}(I)$

$$\boldsymbol{Z}(I) = \sum_I {}^{-1}(\boldsymbol{I} - \boldsymbol{P}_{II})^{-1} \tag{5.8}$$

$$\mathrm{diag}\left(\sum_I\right) = (\boldsymbol{I} - \boldsymbol{P}_{II})^{-1}\mathbf{1}. \tag{5.9}$$

Once we have the bounds $\boldsymbol{v}_I^{\inf}$, $\boldsymbol{v}_I^{\sup}$, then $\boldsymbol{x}_I^{\inf/\sup} = \boldsymbol{v}_I^{\inf/\sup}X_I$ from (3.13). Now one way to obtain bounds on the mean queue length is to simply use $\boldsymbol{x}^{\inf/\sup}$ instead of $\boldsymbol{x}^*$ in (3.28a). This gives our first set of bounds:

$$E_1(D)^{\inf/\sup} = \sum_{I=0}^{N_v} X_I \sum_{i=0}^{M} i v_{I_i}^{\inf/\sup} \tag{5.10a}$$

thus

$$E_1(D)^{\inf} = \sum_{I=0}^{N_v} X_I \sum_{i=0}^{M} i \min_{k\in\mathcal{J}(I)} (Z(I)_{ki}) \tag{5.10b}$$

$$E_1(D)^{\sup} = \sum_{I=0}^{N_v} X_I \sum_{i=0}^{M} i \max_{k\in\mathcal{J}(I)} (Z(I)_{ki}). \tag{5.10c}$$

From a closer look at (5.10a) and by using Theorem 1 we can show that the bounds (5.10b), (5.10c) can be improved as follows. We know from Theorem 1 that if $\boldsymbol{v}$ is the left eigenvector of $\boldsymbol{B} \in \mathcal{B}(1, \boldsymbol{L})$ then $\boldsymbol{v}^T$ can be written as the convex combination of the rows of $\boldsymbol{Z}$, which have indices in $\mathcal{J}$, i.e.,

$$v_i = \sum_{k\in\mathcal{J}} \beta_k Z_{ki} \text{ for some } \boldsymbol{\beta} = [\beta_k] \in \Re_n^+,$$

$$\textstyle\sum_{k\in\mathcal{J}} \beta_k = 1.$$

For NCD systems this becomes

$$v_{I_i} = \sum_{k\in\mathcal{J}(I)} \beta_{I_k} Z(I)_{ki}, \tag{5.11}$$

with $\boldsymbol{\beta}_I = [\beta_{I_k}]$ satisfying

$$\boldsymbol{\beta}_I \in \Re_{n(I)}^+, \qquad \sum_{k\in\mathcal{J}(I)} \beta_{I_k} = 1.$$

The mean queue length is given by (3.28b), $E(D) = \sum_I X_I \sum_i i v_{I_i}$. Substituting (5.11) we get

$$E(D) = \sum_I X_I \sum_i i \sum_{k\in\mathcal{J}(I)} \beta_{I_k} Z(I)_{ki}$$

$$= \sum_I X_I \sum_{k\in\mathcal{J}(I)} \beta_{I_k} \sum_i i Z(I)_{ki}. \tag{5.12}$$

Define $E(D)_{I_k} \triangleq \sum_i i Z(I)_{ki}$. Thus, $\sum_i i v_{I_i}$ is a convex combination of $E(D)_{I_k}$ for $k \in \mathcal{J}(I)$. Hence the bounds of $\sum_i i v_{I_i}$ are simply

$$\min_{k\in\mathcal{J}(I)} E(D)_{I_k} \quad \text{and} \quad \max_{k\in\mathcal{J}(I)} E(D)_{I_k}.$$

We can thus take our new bounds as

$$E_2(D)^{\inf} = \sum_{I=0}^{N_v} X_I \min_{k\in\mathcal{J}(I)} \left(\sum_{i=0}^{M} i Z(I)_{ki}\right) \tag{5.13a}$$

$$E_2(D)^{\sup} = \sum_{I=0}^{N_v} X_I \max_{k\in\mathcal{J}(I)} \left(\sum_{i=0}^{M} i Z(I)_{ki}\right) \tag{5.13b}$$

Proposition 1: Mean queue length bounds (5.13a) and (5.13b) are tighter than those obtained using (5.10b) and (5.10c), i.e.,

$$E_1(D)^{\inf} \le E_2(D)^{\inf} \tag{5.14a}$$

$$E_1(D)^{\sup} \ge E_2(D)^{\sup}. \tag{5.14b}$$

Proof: We will prove only (5.14a) since (5.14b) has the same proof with min replaced by max. By definition

$$\min_{k\in\mathcal{J}(I)} Z(I)_{ki} \le Z(I)_{ji} \qquad \forall j \in \mathcal{J}(I).$$

Multiplying by i and summing on each side

$$\sum_i i \min_{k\in 8F J(I)} Z(I)_{ki} \le \sum_i i Z(I)_{ji} \quad \forall j \in \mathcal{J}(I).$$

Thus, clearly,

$$\sum_i i \min_{k\in\mathcal{J}(I)} Z(I)_{ki} \le \min_{j\in\mathcal{J}(I)} \left(\sum_i i Z(I)_{ji}\right).$$

Multiplying by X_I and summing over I

$$\sum_{I=0}^{N_v} X_I \sum_{i=0}^{M} i \min_{k\in\mathcal{J}(I)} (Z(I)_{ki}) \le \sum_{I=0}^{N_v} X_I \min_{k\in\mathcal{J}(I)} \left(\sum_{i=0}^{M} i Z(I)_{ki}\right).$$

This completes the proof.

Examples of plots of the mean queue length are shown in Figs. 10 and 11. Fig. 10 uses the same parameters as Fig. 7, with $M = 60$, and $\alpha = 100$ except here the mean queue length is plotted instead of the mean waiting times. Both sets of upper and lower bounds $E_1(D)^{\inf/\sup}$ and $E_2(D)^{\inf/\sup}$ are plotted. As expected, $E_2(D)^{\inf/\sup}$ are clearly tighter bounds. Fig. 11 shows results for a larger system with 8 channels where the bounds used are those of (5.13a), (5.13b).

A. Computational Requirements

The error bounding scheme requires the computation of n matrix inverses (5.8), each of size $n(I) \times n(I)$. This computation is of order $n \times n^2(I)$. Hence, for Senet, the original problem that was of size $(MN_v)^2$ has been reduced to $N_v \times M^2$. Thus a computational reduction by a factor of N_v for the calculation of error bounds for the closed form solution given by (3.30) has been achieved.

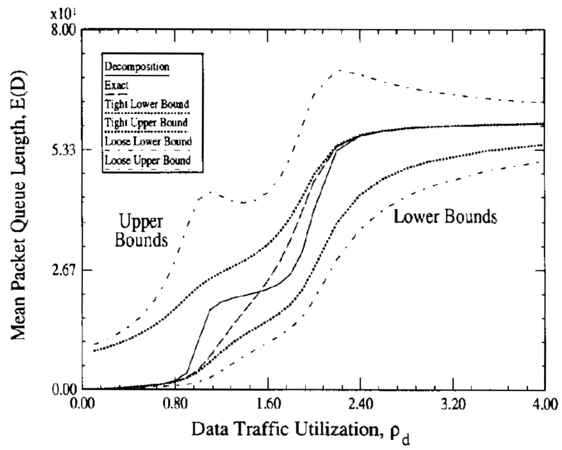

Fig. 10. Queue length with bounds: 2(1, 1), $M = 60$.

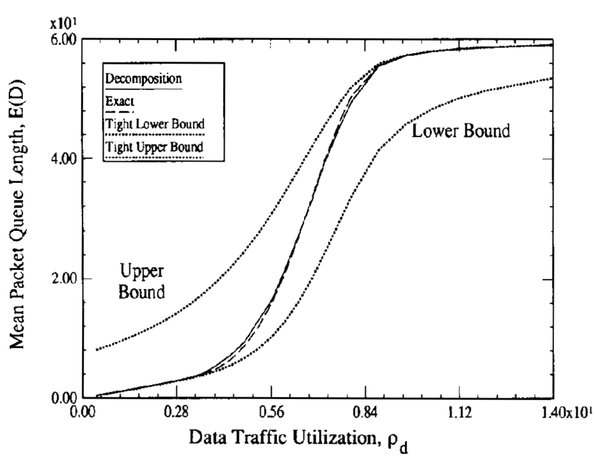

Fig. 11. Queue length with bounds: 8(4, 4), $M = 60$, $\alpha = 100$.

VI. Summary and Conclusions

We have presented an approximation technique with error bounds for the data performance analysis in a movable boundary integration scheme. The technique gives a closed form solution to the distribution of the Markov chain for the number of packets in the data queue. From this all the necessary performance criteria such as the mean data queue length and the data blocking probability can be derived. It is found that the approximation becomes more accurate as the ratio of voice holding to data holding time increases. It also improves as the number of channels increases, and hence is particularly useful for large systems.

Apart from giving insight into the system's behavior, the main importance of the closed form solution lies in the fact that it can be used for optimization purposes. In addition, since the decomposition mean queue length and data blocking probability approximations both appear to be upper bounds, i.e., worst case behavior, in the low to medium traffic region ($\rho_d < \rho_d^{\max}$), they can be useful in system design.

The decomposition technique can also be easily extended to include more classes of traffic. These could be circuit switched users with differing holding times and bandwidth requirements, such as that of video. A study of such systems with various access policies for the circuit switched traffic is underway and we intend to report the results in the future, when available.

Acknowledgment

We would like to acknowledge helpful discussions and suggestions made by P. O'Reilly of GTE Labs., Waltham, MA, and his direction of part of the research while the principal author was at GTE. We also thank the reviewers for their valuable suggestions.

References

[1] D. P. Gaver and J. P. Lehoczky, "Channels that cooperatively service a data stream and voice messages," *IEEE Trans. Commun.*, vol. COM-30, pp. 1153–1162, May 1982.

[2] A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis.* New York: McGraw-Hill, 1978.

[3] M. Schwartz, *Telecommunication Networks: Protocols, Modeling and Analysis.* Reading, MA: Addison-Wesley, 1987.

[4] P. J. Courtois, *Decomposability, Queueing and Computer System Applications.* New York: Academic, 1977.

[5] G. Coviello and P. A. Vena, "Integration of circuit/packet switching in a SENET (slotted envelope network) concept," *Nat. Telecommun. Conf. Rec.*, New Orleans, LA, Dec. 1975, pp. 42-12–42-17.

[6] K. Kummerle, "Multiplexor performance for integrated line- and Packet-switched traffic," *ICCC*, Stockholm, Sweden, 1974, pp. 508–515.

[7] C. J. Weinstein, M. L. Malpass, and M. J. Fischer, "Data traffic performance of an integrated circuit- and packet-switched multiplex structure," *IEEE Trans. Commun.*, vol. COM-28, pp. 873–878, June 1980.

[8] P.-J. Courtois and P. Semal, "Bounds for the positive eigenvectors of nonnegative matrices and for their approximations by decomposition," *J. Assoc. Comput. Machinery*, vol. 31, no. 4, pp. 804–825, Oct. 1984.

[9] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences.* New York: Academic, 1979.

[10] B. Noble and J. W. Daniel, *Applied Linear Algebra.* Englewood Cliffs, NJ: Prentice Hall, 1969.

[11] H. A. Simon, and A. Ando, "Aggregation of variables in dynamic systems," *Econometrica*, vol. 29, 1961, pp. 111–138.

Sayeed Ghani received the S.B. degree in electrical engineering from Massachusetts Institute of Technology (M.I.T.) in 1984. From 1984 to 1989 he was a Graduate Research Assistant at Columbia University, from where he received the M.S. and Ph.D. degrees in electrical engineering in 1986 and 1990, respectively.

He has been a Member of Technical Staff at NYNEX Science & Technology, Inc. since 1989, where he has been working on the switched multi-megabit data service (SMDS) project. He has been an active member of the IEEE 802.6 committee, and the technical editor of the IEEE 802.6 standard since 1989. His research interests include Metropolitan Area Networks, Wide Area Networks, and performance analysis of Broadband Networks.

Mischa Schwartz (S'46–A'49–M'54–SM'54–F'66–LF'92) received the B.E.E degree from Cooper Union, New York, NY, in 1947, the M.E.E. degree from the Polytechnic Institute of Brooklyn, NY, in 1949, and the Ph.D. degree in applied physics from Harvard University, Cambridge, MA, in 1951.

From 1947 to 1952 he was a Project Engineer with the Sperry Gyroscope Company, working in the fields of statistical communication theory, radar detection, and radar design. From 1952 to 1974 he was Professor of Electrical Engineering at the Polytechnic Institute of Brooklyn, serving as Head of the Electrical Engineering Department from 1961 to 1965. During the year 1965–1966 he was an NSF Science Faculty Fellow at the Laboratoire de Physique, Ecole Normale Superieure, Paris, France. During the academic year 1973–1974 he was a Visiting Professor at Columbia University, New York, NY. He joined that institution in September 1974 as Professor of Electrical Engineering. For the 1980 calendar year, he was on leave as a Visiting Scientist with IBM Research. During 1986, he was on leave as a Resident Consultant with NYNEX Science and Technology.

Dr. Schwartz is a member of the National Academy of Engineering. He is a former Director of IEEE and formerly Chairman of the Information Theory Group, and past President of the Communications Society. He is author or co-author of eight books on communications, signal processing, and computer communications networks, and has published extensively in the technical literature. He is on the editorial board of *Networks* and *Computer Networks*. He has lectured extensively both in this country and abroad on various aspects of communication theory, communication systems, digital communications, and computer communications. He was a recipient of a Distinguished Visitor Award in 1975 from the Australian-American Education Foundation. He is a Fellow of the AAAS, and past Chairman, Commission C., U.S. National Committee/URSI. He has been consultant on communications, computer communications, signal processing and radar to many companies. In 1983, Dr. Schwartz was the recipient of the IEEE's annual Education Medal. Columbia University awarded him its Great Teacher Medal in 1983. In 1984, the IEEE Centennial year, Dr. Schwartz was cited as one of the 10 all-time outstanding electrical engineering educators. He served from 1985 to 1988 as Director of the Columbia University Center for Telecommunications Research, one of six national Engineering Research Centers established in 1985 under major grants from the National Science Foundation. In 1986, Dr. Schwartz was the recipient of the Cooper Union Gano Dunn Award, given annually for outstanding achievement in Science Technology. In 1989 he received the IEEE Region I Award for outstanding engineering management and leadership. In December 1992, he was Awarded Finalist, Mayor's Awards for Excellence in Science and Technology, New York City, NY.