Queueing Systems 1(1987)289 – 309

AN APPROXIMATE ANALYSIS FOR A CLASS OF ASSEMBLY-LIKE QUEUES

F. BONOMI

AT&T Bell Laboratories, Crawfords Corner Road, HO 3M-306, Holmdel, NJ 07733, USA

Received 2 June 1986
(Revised 20 August 1986)

Abstract

Assembly-like queues model assembly operations where separate input processes deliver different types of component (customer) and the service station assembles (serves) these input requests only when the correct mix of components (customers) is present at the input. In this work, we develop an effective approximate analytical solution for an assembly-like queueing system with N ($N \geqslant 2$) classes of customers forming N independent Poisson arrival streams with rates $\{\lambda_i\}_{i=1,\ldots,N}$. The arrival of a class of customers is "turned off" whenever the number of customers of that class in the system exceeds the number for any of the other classes by a certain amount. The approximation is based on the decomposition of the original N input stream stage into a cascade of $N-1$ two-input stream stages. This allows one to refer to the theory of paired customer systems as a foundation of the analysis, and makes the problem computationally tractable. Performance measures such as server utilization, throughput, average delays, etc., can then be easily computed. For illustrative purposes, the theory and techniques presented are applied to the approximate analysis of a system with $N = 3$. Numerical examples show that the approximation is very accurate over a wide range of parameters of interest.

Keywords

Queueing theory, synchronization, assembly-like queues, decomposition, approximation.

1. Introduction

Assembly-like queues model assembly operations often found in manufacturing systems, and can be pictorially represented as shown in fig. 1. Such operations consist of N different input processes and a service station. Each input process delivers a different type of component (customer) and the service station assembles (serves) these input requests only when the correct mix of components (customers) is present at the input.

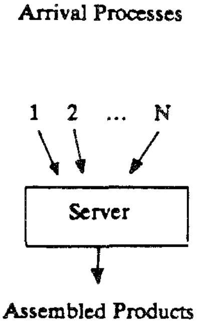


Fig. 1. Representation of an assembly operation.

A similar type of mechanism, involving synchronization between different arrival processes, can be encountered in a variety of contexts, not exclusively relating to manufacturing systems. Relevant examples can be found among the numerous synchronization mechanisms typical of a computing, communication, or manufacturing system environment.

Despite the importance of such situations, there is a limited literature on the subject of mathematical modeling of assembly-like operations, because of the difficulty in analyzing such systems. Harrison [1] studies assembly-like queues with the assumption that N ($N \geqslant 2$) classes of customers arrive according to N mutually independent renewal processes. He proves that such queueing systems, where no arrival control strategy is present, are inherently unstable. Latouche [2] analyzes particular assembly-like queues with $N = 2$ and shows that imposing a control mechanism, which is a function of the excess (i.e. the difference) between the queue lengths for the two classes of customers, can make the system stable. The analytical solution for the equilibrium probability distribution of the system is shown to be matrix-geometric. Neuts [3] and Ramaswami [4] treat some generalizations of Latouche's work to the case of input streams that are phase-type processes and to the general service time case, respectively.

It is worth noting that a special case of the problem solved by Latouche, obtained by assuming deterministically zero service times, is known as the "taxi problem", and is treated by Srivastava and Kashyap [6]. The case of finite buffers was first considered by Bhat [7], who considers the $N = 2$ case with a small buffer size, and by Lipper and Sengupta [8], who give approximations for systems with N classes of customers and Poisson arrivals with the same rate, and an arrival is blocked and lost if the buffer is full. Baccelli and Massey [9] treat networks of queues obtained by combining series and fork-join networks, obtaining bounds on the moments of the total delay to traverse such networks.

In this work, we develop an effective approximate analytical solution for an assembly-like queueing system with N ($N \geqslant 2$) classes of customers forming N independent Poisson arrival streams with rates $\{\lambda_i\}_{i=1,\ldots,N}$. The arrival control mechanism adopted is a generalization of the one introduced by Latouche to the case of N classes. The approximation is based on the decomposition of the original N input stream stage into a cascade of $N-1$ two-input stream stages. This allows reference to the theory of paired customer systems as a foundation of the analysis, and makes the problem computationally tractable.

Sections 2 and 3 are devoted to the problem statement and to the general description of the approximate approach. In sect. 4, we briefly review Latouche's treatment of the $N = 2$ case. An extension of this theory to the case of a system with a Poisson and a Markov Modulated Poisson input stream is presented in sect. 5. The output processes for the systems considered in sects. 4 and 5 are shown in sect. 6 to be generalized N-processes [5]. Section 7 discusses the approximation of a generalized N-process by a Markov Modulated Process. Finally, for illustrative purposes, the theory and techniques presented in sects. 3–7 are applied to the approximate analysis of a system with $N = 3$. The description and results of such analysis are considered in sect. 8, while the last section discusses feasible extensions and conclusions of this work.

2. Problem statement

Let $n_k(t)$, $k = 1, 2, \ldots, N$, be the number of customers of class k in the system at time t. The state of the system can be represented by the vector $(n_K(t), e_j(t), j = 1, 2, \ldots, N-1)$, with

$$n_K(t) = \min_{j=1,\ldots,N} n_j(t)$$

$$e_j(t) = n_j(t) - n_N(t) \qquad j = 1, \ldots, N.$$

Also, let

$$e_{jk} = e_{jk}(t) = n_j(t) - n_k(t) = e_j(t) - e_k(t) \qquad j, k = 1, 2, \ldots, N.$$

The function $e_{jk}(t)$ is called the excess between class j and class k customers.

For each class j, the arrival stream is assumed to form an independent Poisson process with parameters depending on the excess between class j and each of the remaining classes. Such a dependency is formulated as follows:

$$\lambda_j = \lambda_j(e_{j1}, \ldots, e_{jN}) \qquad j = 1, \ldots, N$$

$$\lambda_j = 0 \text{ if } \max_i e_{ji} = J_j, \quad J_j > 0, \ j = 1, \ldots, N.$$

In our treatment we will consider the case $\lambda_j(e_{j1}, \ldots, e_{jN}) = \text{const}(j) = \lambda_j$ if

$$\max_i e_{ji} < J_j, \quad j = 1, 2, \ldots, N.$$

This is the simplest possible control law of this kind that can be considered. Although more sophisticated strategies could be analyzed, we will restrict ourselves to this case for the sake of clarity of exposition.

The time taken to serve an N-plet of customers, one for each class, has exponential distribution with parameter μ. Service can begin only when at least one customer per class is requesting service. The cases where different numbers of customers per class are required may be handled by appropriate scaling.

3. An approximation

The assembly system described above can be conceptually viewed as shown in fig. 2. Here we explicitly decompose the assembly process into a synchronization

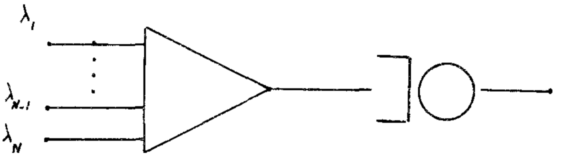


Fig. 2. A conceptual view of an assembly-like system.

phase and a service phase. An output is offered to the server as soon as a customer in each class is present at the input, and the server operates on such a group of customers as in a regular single-class queue. The synchronization phase can, in turn, be considered as the result of a cascade of $N-1$ stages, each of which operates an elementary synchronization, as shown in fig. 3. An elementary synchronizer produces an output as soon as the second component of a pair arrives at the input.

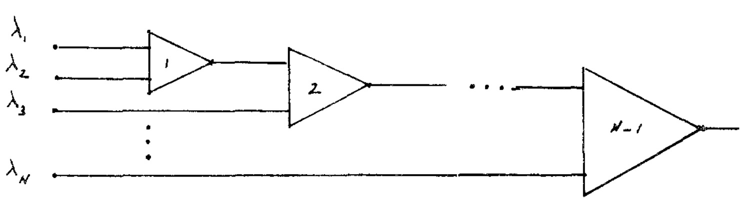
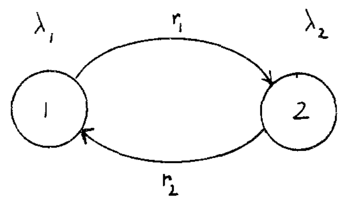


Fig. 3. Decomposition of the synchronization phase.

We will approach the problem using the decomposition suggested in fig. 3. The output for each stage, discussed in sect. 6, is naturally approximated by a two-state Markov Modulated Process (MMP). This process is characterized by two Poisson arrival rates λ_1 and λ_2 corresponding to the two possible system states. The system moves between the states as a Markov process with rates r_1 and r_2, as shown in fig. 4.



λ_j : arrival rate when in state j; r_j : rate out of state j.

Fig. 4. The two-state Markov Modulated Process.

The MMP has already been successfully used as an approximation [10] because of its simplicity and sufficient flexibility. As we will discuss in sect. 4, the matrix-geometric solution for the paired customer system with Poisson arrival streams [2] can be readily extended to provide a solution for the case when one of the arrival streams is an MMP. Thus, after the output streams of every paired synchronizer is approximated by an MMP, we can treat analytically every stage in our decomposition. Whenever only global parameters such as throughput and utilization are of interest, we just need to compute the equilibrium distribution for the excess at each stage, determining the complete matrix-geometric solution only at the final stage. The former computation can be performed very quickly by using a recursive procedure proposed by Chandy et al. [11].

The major challenge to be faced taking this approach relates to the fact that the control mechanism we are considering implies the presence of feedback loops from successive stages. Taking this feedback into account explicitly and exactly is equivalent to solving the problem in its full dimensionality, which is a prohibitive task. To avoid such a "curse of dimensionality", we first open the feedback loop by neglecting the influence of the later stages on the previous ones. We approximately compensate by (a) relabeling, without loss of generality, the input streams so that $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$, and (b) using an iterative procedure of the fixed point type.

Since the feedback loop is active only when the excess for a certain stage is positive in the direction of the arrival stream from previous stages and large, with respect to the imposed bounds, we expect (a) to drastically decrease the probability of active feedback whenever the arrival rates are not approximately equal. The residual feedback influence is then compensated by using an iterative procedure which suitably adjusts the bounds to a point of equilibrium. A more detailed treatment of this technique is given in sect. 8, where we illustrate the case $N = 3$.

4. A queue with paired customers and Poisson arrivals

In this section, we briefly review the analysis presented by Latouche [2], which plays a fundamental role in our discussion. We consider here the assembly-like system introduced in sect. 2 in the case $N = 2$.

The state of the system is represented by $(n(t), e(t))$, where $n(t) = \min(n_1(t), n_2(t))$ is the number of pairs and $e(t) = n_1(t) - n_2(t)$ is the excess. The model can be described as a continuous-time Markov chain on the state space (i, j), $i \geqslant 0$, $-J_2 \leqslant j \leqslant J_1$ with infinitesimal generator Q given by

$$Q = \begin{bmatrix} A_{10} & A_0 & & \\ A_2 & A_1 & A_0 & 0 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}, \tag{1}$$

where all entries are square matrices of order $(J_1 + J_2 + 1)$. Each row in the block-partitioned matrix Q corresponds to a different value of $i \geqslant 0$. The matrices A_{10}, A_0, A_1, A_2 are given by:

$$A_{10} = \begin{array}{c} \qquad\qquad\qquad \text{column } j = 0 \\ \qquad\qquad\qquad \downarrow \\ \begin{bmatrix} -\lambda_1 & 0 & & & & & \\ \lambda_2 & (-\lambda_1-\lambda_2) & 0 & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \lambda_2 & (-\lambda_1-\lambda_2) & \lambda_1 & & \\ & & & 0 & (-\lambda_1-\lambda_2) & \lambda_1 & \\ & & & & \ddots & \ddots & \ddots \\ & & & & & (-\lambda_1-\lambda_2) & \lambda_1 \\ & & & & & 0 & -\lambda_2 \end{bmatrix} \end{array} \tag{2}$$

$$A_1 = A_{10} - \mu I, \quad A_2 = \mu I \tag{3}$$

$$A_0 = \begin{array}{c} \begin{array}{ccccccc} & & & \text{column } j = 0 & & & \\ & & & \downarrow & & & \end{array} \\ \left[\begin{array}{ccccccc} 0 & \lambda_1 & & & & & \\ & \ddots & \ddots & & & & \\ 0 & & \ddots & \ddots & & & \\ & \ddots & & \ddots & \lambda_1 & & \\ & & \ddots & & \ddots & \ddots & \\ & & 0 & 0 & 0 & & \\ & & & \lambda_2 & \ddots & \ddots & \\ & & & & \ddots & \ddots & 0 \\ & & & & & \ddots & \ddots \\ & & & & & \lambda_2 & 0 \end{array}\right] \end{array} \tag{4}$$

I is the identity matrix of order $(J_1 + J_2 + 1)$. The structure of Q suggests that the steady-state probability vector $\boldsymbol{x}$ associated with Q has a matrix-geometric structure [3], i.e. $\boldsymbol{x} \doteq (x_0, x_1, \ldots)$, with x_i a $(J_1 + J_2 + 1)$-dimensional row vector and

$$x_i = x_0 R^i, \quad i = 1, 2, \ldots, \tag{5}$$

where R is the maximal non-negative solution of the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0 \tag{6}$$

such that its spectral radius is smaller than one ($\mathrm{sp}(R) < 1$). R can be computed by the recursion

$$R_{n+1} = C_0 + R_n^2 C_2 \quad \text{for } n \geqslant 0, \tag{7}$$

where $C_0 = -A_0 A_1^{-1}$, $C_2 = -A_2 A_1^{-1}$. Note that A_1 is proved to be nonsingular [2]. An interesting and powerful accuracy check on the computation of the matrix R is given by the relation

$$R\boldsymbol{e} = \mu^{-1} A_0 \boldsymbol{e}, \tag{8}$$

where $\boldsymbol{e}$ is the unit column vector.

5. A more general queue with paired customers

In this section, we consider a queue with paired customers characterized by a Poisson arrival stream and an MMP arrival stream. This extension is important since we will later see that the output processes of the cascaded synchronizers can be effectively approximated by MMP processes. This case is slightly more general than the one studied by Latouche and has not been treated in the literature. We show that, even in this case, the system still has a matrix-geometric equilibrium distribution.

The presence of two states in the MMP doubles the number of states necessary to describe the system we are considering. For each value of the excess and of $\min(n_1(t),\ n_2(t))$ we have two possible values for the state of the MMP. We can thus consider the state of our system to be the triplet $(e(t),\ n(t),\ s(t))$, with $e(t) = n_1(t) - n_2(t)$, $n(t) = \min(n_1(t),\ n_2(t))$, and $s(t)$ = state of the MMP at time t. We decide to order the states as follows:

$$(-J_2, 0, 0) \ldots (J_1, 0, 0), (-J_2, 0, 1) \ldots (J_1, 0, 1),$$

$$(-J_2, 1, 0) \ldots (J_1, 1, 0), (-J_2, 1, 1) \ldots (J_1, 1, 1),$$

$$\ldots\ldots\ldots\ldots\ .$$

With this particular ordering, the infinitesimal generator $Q^\star$ for the continuous-parameter Markov chain describing the system is, again, given as

$$Q^\star = \begin{bmatrix} A_{10}^\star & A_0^\star & & 0 \\ A_2^\star & A_1^\star & A_0^\star & \\ & \ddots & \ddots & \ddots \\ 0 & & \ddots & \ddots \end{bmatrix} \tag{9}$$

with all the entries being matrices of order $2(J_1 + J_2 + 1)$. Again, each row in the block-partitioned matrix $Q^\star$ corresponds to a different value of $n(t)$.

The matrices $A_{10}^\star$, $A_0^\star$, $A_1^\star$, $A_2^\star$ are given by eq. (10).

$$
A_{10}^{\star}=\left[\begin{array}{ccccc|ccccc}
(-r_1-\lambda_1) & 0 & & & & r_1 & & & & \\
\lambda_3 & (-r_1-\lambda_1-\lambda_3) & 0 & & & & \ddots & 0 & & \\
 & \ddots & \ddots & \ddots & & & & \ddots & & \\
 & & \lambda_3 & \ddots & 0 & & 0 & & \ddots & \\
 & & & \lambda_3 & (-r_1-\lambda_3) & & & & & r_1 \\
\hline
r_2 & & & & & (-r_2-\lambda_2) & 0 & & & \\
 & \ddots & 0 & & & \lambda_3 & (-r_2-\lambda_2-\lambda_3) & 0 & & 0 \\
0 & & \ddots & & & & \ddots & \ddots & \ddots & \\
 & & & \ddots & & & & \ddots & \ddots & 0 \\
 & & & & r_2 & & & \lambda_3 & & (-r_2-\lambda_3)
\end{array}\right] \tag{10}
$$

$$A_1^\star = A_{10}^\star - \mu I^\star,\ A_2^\star = \mu I^\star, \tag{11}$$

where $I^\star$ is now the identity matrix of order $2(J_1 + J_2 + 1)$.

$$A_0^\star = \left[\begin{array}{cccccc:cccccc}
0 & \lambda_1 & & & & & & & & & & \\
 & \ddots & \ddots & & & & & & & & & \\
 & & \cdot & \lambda_1 & & & & & & & & \\
 & & 0 & 0 & 0 & & & & 0 & & & \\
 & & & \lambda_3 & \ddots & & & & & & & \\
 & & & & \ddots & \ddots & & & & & & \\
 & & & & \lambda_3 & 0 & & & & & & \\
\hdashline
 & & & & & & 0 & \lambda_2 & & & & \\
 & & & & & & & \ddots & \ddots & & & \\
 & & & & & & & & \cdot & \lambda_2 & & \\
 & & 0 & & & & & 0 & 0 & 0 & & \\
 & & & & & & & \lambda_3 & \ddots & & & \\
 & & & & & & & & \ddots & \ddots & & \\
 & & & & & & & & & \lambda_3 & 0 &
\end{array}\right] \tag{12}$$

The solution for the equilibrium distribution for the system under consideration is thus matrix-geometric and the treatment presented in sect. 3 easily extends to this case with respect to the matrices $A_0^\star$, $A_{10}^\star$, $A_1^\star$, $A_2^\star$.

6. The synchronizer output process

In this section, we investigate the output process for the class of paired customer systems discussed above, when the service time is deterministically equal to zero. We call a system of this kind a synchronizer. First, we introduce the basic ideas by discussing the case of Poisson arrival streams with rates λ_1, λ_2. The case of Poisson and MMP input streams will follow as an immediate consequence.

For the Poisson paired customer system, the output will be a stream of pairs. In the case of a synchronizer, at each instant at least one of the queues must be empty. A pair is produced whenever there is an arrival in the class whose queue is empty, except for the case when both queues are empty. We can describe this phenomenon by

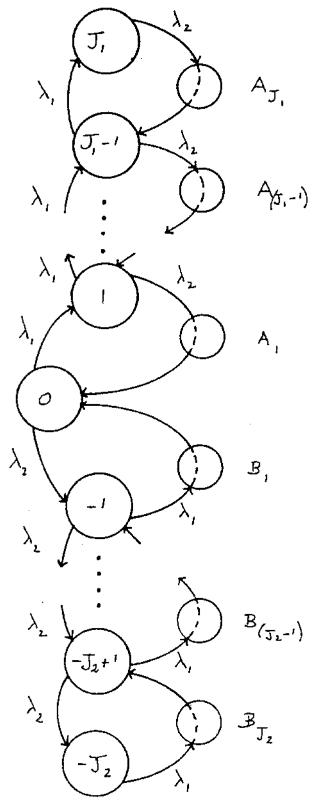


Fig. 5. State transition diagram for P-P synchronizer output process.

using a state transition diagram as shown in fig. 5. We distinguish a set of regular states $R = \{-J_2, \ldots, 0, \ldots J_1\}$, and a set of instantaneous states $I = \{B_{J_2}, \ldots B_1, A_1 \ldots A_{J_1}\}$. The regular states describe the value of the excess at the input of the system, while the instantaneous states are states that are visited with zero sojourn time upon arrival of customers that can form a pair. This implies that for every transition through an instantaneous state, a pair is produced at the output.

Let $\hat{Q}$ be the infinitesimal generator for the continuous-time Markov chain describing the transitions with respect to the regular states (i.e. the presence of the instantaneous states is disregarded). The matrix $\hat{Q}$ is given by

$$\hat{Q} = \begin{bmatrix} -\lambda_1 & \lambda_1 & & & \\ \lambda_2 & (-\lambda_1 - \lambda_2) & \lambda_1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \lambda_1 \\ & & & \lambda_2 & -\lambda_2 \end{bmatrix} \tag{13}$$

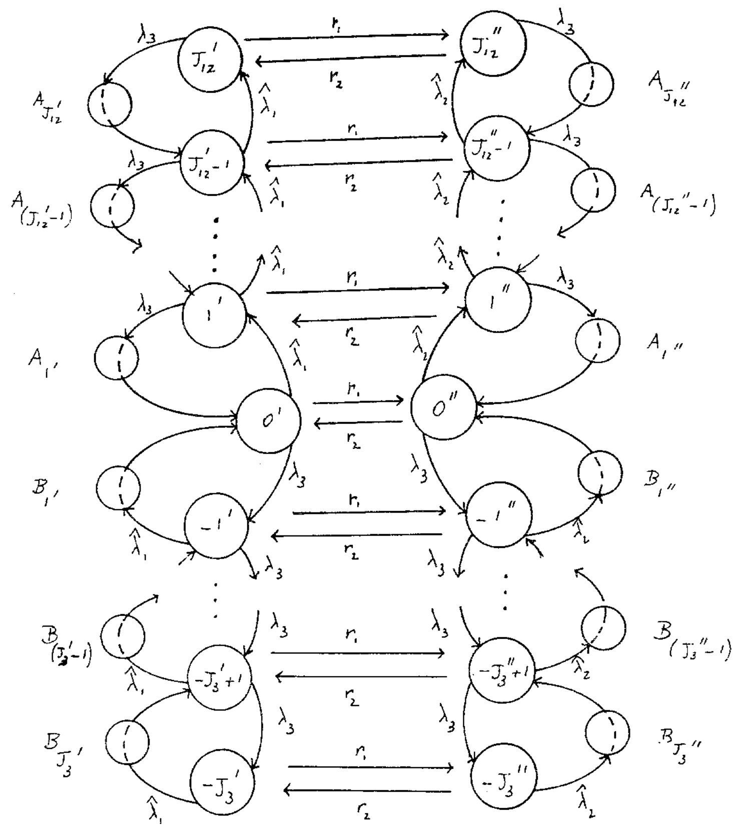


Fig. 6. State transition diagram for P-MMP synchronizer output process.

where we have ordered the states from $-J_2$ to J_1.

At every (i, j) transition involving the visit to an instantaneous state there is an arrival (i.e. an output from the synchronizer) with probability one. The point process obtained in this way is a particular case of a Generalized N-process (GNP). A thorough treatment for this versatile class of processes can be found in [4] and [5].

Let us now consider the case of a synchronizer with a Poisson process of rate λ_3 and an independent MMP with parameters $\hat{\lambda}_1$, $\hat{\lambda}_2$, r_1, r_2 as input processes. The control bounds are J_3 and J_{12}, respectively. It is immediately seen from the state transition description in fig. 6 that, again, the output process from the synchronizer is a GNP described by the matrix $\hat{Q}^\star$ given by eq. (14), where the ordering for the states is obvious and $\hat{Q}^\star$ is a square matrix with dimension $2(J_{12} + J_3 + 1)$.

$$
\hat{Q}^{\star} = \left[\begin{array}{cccc|cccc}
(-r_1-\hat{\lambda}_1) & \hat{\lambda}_1 & & & r_1 & & & \\
\lambda_3 & (-r_1-\hat{\lambda}_1-\lambda_3) & \hat{\lambda}_1 & & & \ddots & & 0 \\
 & \ddots & \ddots & \ddots & & 0 & \ddots & \\
 & & \lambda_3 & \hat{\lambda}_1 \\ & & & (-r_1-\lambda_3) & & & & r_1 \\
\hline
r_2 & & & & (-r_2-\hat{\lambda}_2) & \hat{\lambda}_2 & & \\
 & \ddots & 0 & & \lambda_3 & (-r_2-\hat{\lambda}_2-\lambda_3) & \hat{\lambda}_2 & 0 \\
0 & & \ddots & & & \ddots & \ddots & \ddots \\
 & & & & & & & \hat{\lambda}_2 \\
 & & & r_2 & & & \lambda_3 & (-r_2-\lambda_3)
\end{array}\right] \qquad (14)
$$

7. The MMP approximation

The output process for the synchronizer with Poisson arrivals displays basically two phases, or stages, corresponding to positive and negative values for the excess, respectively, and a special state in between, corresponding to zero excess. It seems thus natural to approximate such a generalized N-process (GNP) by the MMP described in fig. 4. The generalization to the case where the arrival streams are Poisson and MMP is straightforward. As a consequence, every paired synchronizer of the decomposition we consider will have either Poisson input streams or a Poisson and an MMP input stream.

The MMP model has four parameters that will be chosen to match the expected value m, the variance ν, the third moment μ_3, and the time constant τ_c of the GNP. From [10] we can express the above quantities for the MMP in terms of the four parameters $\hat{\lambda}_1$, $\hat{\lambda}_2$, r_1 and r_2 as follows:

$$m^{\mathrm{MMP}} = \frac{\hat{\lambda}_1 r_2 + \hat{\lambda}_2 r_1}{r_1 + r_2}$$

$$\nu^{\mathrm{MMP}} = \frac{r_1 r_2 (\hat{\lambda}_1 - \hat{\lambda}_2)^2}{(r_1 + r_2)^2}$$

$$\mu_3^{\mathrm{MMP}} = \frac{\hat{\lambda}_1^3 r_2 + \hat{\lambda}_2^3 r_1}{r_1 + r_2}$$

$$\tau_c^{\mathrm{MMP}} = \frac{1}{\nu^{\mathrm{MMP}}} \int_0^\infty r^{\mathrm{MMP}}(t)\,\mathrm{d}t = \frac{1}{r_1 + r_2}\,, \tag{15}$$

where $r^{\mathrm{MMP}}(t)$ is the covariance function of the arrival rate for the MMP. We use superscripts to distinguish between quantities relating to the different processes.

We need to express similar quantities for the GNP, in terms of the parameters λ_1, λ_2, J_1, J_2. Let $\pi = [\pi(-J_2), \ldots, \pi(o), \ldots, \pi(J_1)]$ be the row vector of equilibrium probabilities for the state of the GNP. Let $\hat{Q}$ be its infinitesimal generator.

Let Λ be the $(J_1 + J_2 + 1)$-dimensional diagonal matrix of Poisson intensities

$$\Lambda = \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_1 & & & & & \\ & & \ddots & & & & 0 \\ & & & \lambda_1 & & & \\ & & & & 0 & & \\ & 0 & & & & \lambda_2 & \\ & & & & & & \ddots \\ & & & & & & & \lambda_2 \end{bmatrix} \tag{16}$$

Then the kth moment ($k = 1, 2, 3$) for the arrival rate is given by

$$\mu_k^{\mathrm{GNP}} = \lambda_1^k \sum_{j=-J_2}^{-1} \pi(j) + \lambda_2^k \sum_{j=1}^{J_1} \pi(j) \tag{17}$$

which allows us to determine m^{GNP}, ν^{GNP} and μ_3^{GNP}. The covariance function can be expressed as

$$r^{\mathrm{GNP}}(t) = \pi\Lambda\,[\mathrm{e}^{\hat{Q}t} - 1\pi]\,\Lambda 1, \tag{18}$$

with $1 = [1, \ldots, 1]^{\mathrm{T}}$, the unit vector of dimension $(J_1 + J_2 + 1)$.

To determine $\tau_{\mathrm{c}}^{\mathrm{GNP}}$ we need to integrate $r^{\mathrm{GNP}}(t)$. This problem is non-trivial since $\hat{Q}$ is singular as well as 1π. We follow here a procedure suggested in [12] and [13]. Let

$$A = \int_0^\infty (\mathrm{e}^{\hat{Q}t} - 1\pi)\,\mathrm{d}t, \tag{19}$$

then

$$\hat{Q}\int_0^\infty (\mathrm{e}^{\hat{Q}t} - 1\pi)\,\mathrm{d}t = \int_0^\infty \hat{Q}\,\mathrm{e}^{\hat{Q}t}\,\mathrm{d}t = \hat{Q}A \tag{20}$$

since $\hat{Q}1\pi = 0$. Integrating, we have

$$\mathrm{e}^{\hat{Q}t}\big]_0^\infty = \hat{Q}A,\ (1\pi - I) = \hat{Q}A. \tag{21}$$

Also,

$$1\pi\int_0^\infty (\mathrm{e}^{\hat{Q}t} - 1\pi)\,\mathrm{d}t = 1\pi A$$

$$\int_0^\infty (1\pi\,\mathrm{e}^{\hat{Q}t} - 1\pi)\,\mathrm{d}t = 1\pi A$$

$$\int_0^\infty (1\pi - 1\pi)\,\mathrm{d}t = 1\pi A,$$

since by definition of π, $1\pi\,\mathrm{e}^{\hat{Q}t} = 1\pi$. Thus,

$$1\pi A = 0. \tag{22}$$

Now adding (21) and (22), we obtain

$$(1\pi - I) = \hat{Q}A + 1\pi A, \quad \text{and} \quad A = (\hat{Q} + 1\pi)^{-1} (1\pi - I),$$

since $(\hat{Q} + 1\pi)$ can be shown to be non-singular [13]. The matrix $(\hat{Q} + 1\pi)$ plays the role of a pseudo-inverse for the singular matrix $\hat{Q}$. We can now determine τ_c^{GNP} as

$$\tau_c^{\mathrm{GNP}} = \left(\frac{1}{\nu^{\mathrm{GNP}}}\right) \pi\Lambda (\hat{Q} + 1\pi)^{-1} (1\pi - I)\Lambda 1. \tag{23}$$

We can obtain the MMP parameters as follows, where are right-hand side quantities refer to the GNP:

$$r_1 = \frac{1}{\tau_c(1+\eta)}, \quad r_2 = \frac{\eta}{\tau_c(1+\eta)}$$

$$\hat{\lambda}_1 = m + \sqrt{\nu}/\eta, \quad \hat{\lambda}_2 = m - \sqrt{\nu}/\eta, \tag{24}$$

where

$$\eta = 1 + \frac{\delta}{2}[\delta - \sqrt{\eta + \delta^2}], \quad \text{and} \quad \delta = \frac{\mu_3 - 3m\nu - m^3}{\nu^3/2}.$$

8. Numerical results for $N = 3$

In this section, we discuss in detail the case $N = 3$. Following the procedure outlined in sect. 3, we look at the assembly-like system with $N = 3$ as a cascade of two synchronizers followed by a simple exponential server station (fig. 7). The customers queueing for such a station are triplets of customers, one for each of the three input classes.

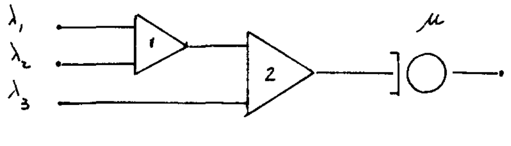


Fig. 7. Decomposition for $N = 3$.

The original Poisson arrival streams are labeled, without loss of generality, such that $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3$. At the elementary synchronizers inputs, one or the other queue will always be empty. We are interested in determining the following parameters:

(a) Expected excess at the second synchronizer, i.e. $E(e_{II}) = E(n_3 - \min(n_1, n_2))$.
(b) System utilization ρ_3.

According to the procedure outlined in sect. 3, we proceed as follows:

Step 0: Set $\hat{J}_i = J_i$, $i = 1, 2$.

Step 1: Determine the invariant distribution for the excess at the first synchronizer. This computation is equivalent to the computation of the invariant distribution for a birth and death process between the states $-\hat{J}_2$ and $\hat{J}_1$.

Step 2: Determine $\hat{\lambda}_1$, $\hat{\lambda}_2$, r_1, r_2, the parameters of the MMP approximation for the output of the first synchronizer. (See sect. 7.)

Step 3: Determine the limiting distribution of the excess at the second synchronizer $\pi(e_{II})$. The control parameters are, in this case, $J = \min(\hat{J}_1, \hat{J}_2)$ and J_3. This computation can be efficiently performed by a recursive technique described by Chandy [11].

Step 4: If $\hat{J}_i \leqslant J_i - \ulcorner E(e_{II}(\hat{J}_1, \hat{J}_2)^+)\urcorner$, $i = 1, 2$, proceed with step 5. Otherwise let $\hat{J}_i = \hat{J}_i - 1$, $i = 1, 2$, and go to step 1. Here, the notation $e_{II}(\hat{J}_1, \hat{J}_2)$ is used to stress the fact that e_{II} is determined when $\hat{J}_1$ and $\hat{J}_2$ are the bounds for the first synchronizer. Also, the procedure will always terminate since $\ulcorner E(e_{II}(\hat{J}_1, \hat{J}_2)^+)\urcorner$ is a non-increasing function of both $\hat{J}_1$ and $\hat{J}_2$.

Step 5: Determine the matrix-geometric solution for the paired customer system with MMP and Poisson arrival streams corresponding to the second synchronizer and the service process. Compute interesting performance measures.

Note that the iteration through steps 1–4 implements the basic compensation for the decoupling between the two synchronizers.

To validate the performance of the approximation procedure we used a SIMSCRIPT II simulation of the original model. In figs. 8–9 and table 1 we show some of the results in a graphic and tabulated form.

We expect that whenever the original system is weakly coupled, the approximation should work well, even without the compensation in step 4. This is the case when the arrival rates are not balanced. When the system is strongly coupled, the approximation relies on the compensation mechanism.

In fact, from our experience, we feel that compensation is only necessary in a narrow range of values for the arrival rates. In particular, such a range R is approximately characterized by

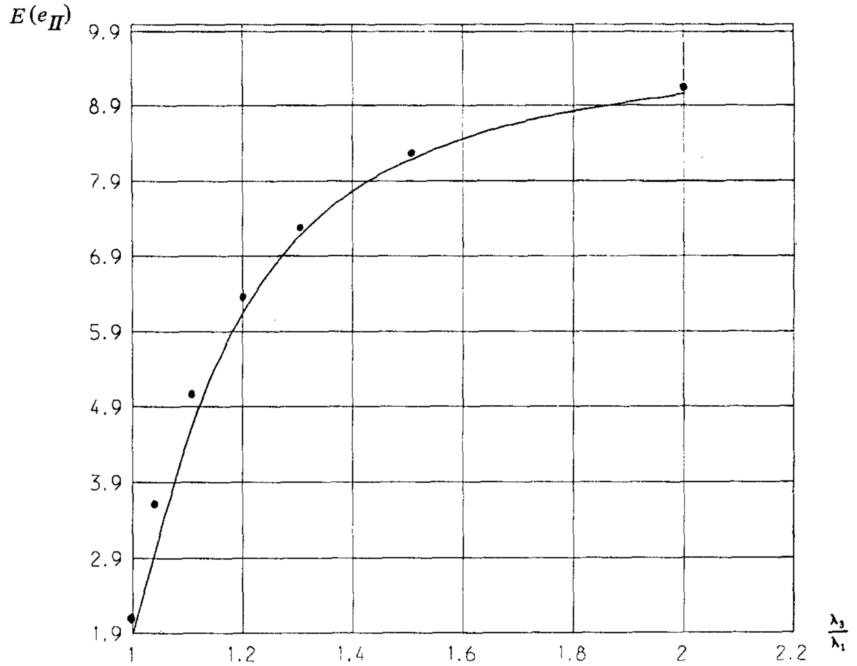


Fig. 8. Analytical approximation and simulation results versus λ_3/λ_1 for expected excess as the second synchronizer, i.e. $E(e_{II}) = E(n_3 - \min(n_1, n_2))$. ($J_1 = J_2 = J_3 = 10$, $\lambda_2 = \lambda_1$, $\mu = \lambda_3$).

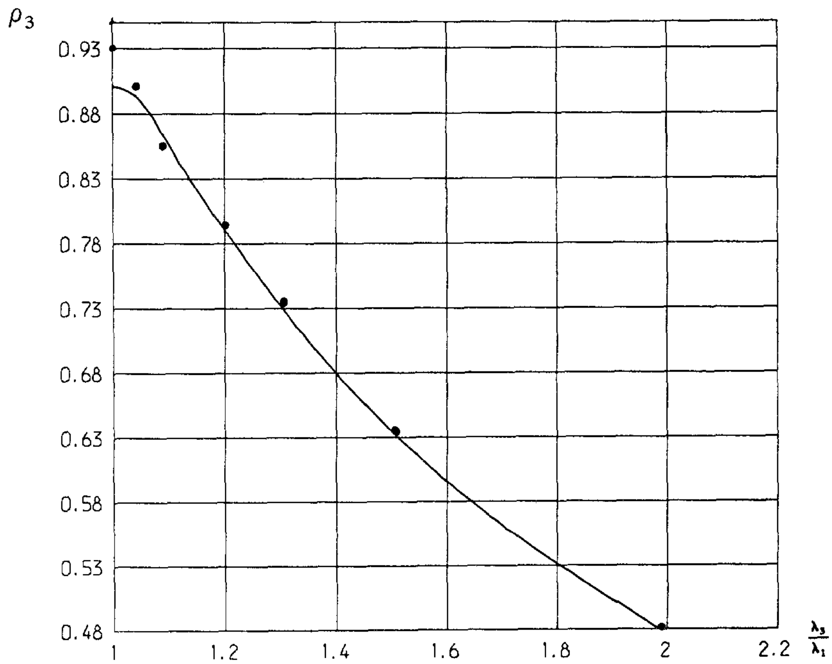


Fig. 9. Analytical approximation and simulation results versus λ_3/λ_1 for system utilization, ρ_3. ($J_1 = J_2 = J_3 = 10$, $\lambda_2 = \lambda_1$, $\mu = \lambda_3$).

Table 1

Numerical results. ($J_1 = J_2 = J_3 = 10$, $\lambda_2 = \lambda_1$, $\mu = \lambda_3$)

λ_3/λ_1	Simul. $E(e_{II})$	Approx. $E(e_{II})$	Error % $E(e_{II})$	Simul. ρ_3	Approx. ρ_3	Error % ρ_3
1.0	– 2.1750	– 2.0759	4.8	0.9333	0.91233	2.3
1.05	– 3.8560	– 3.5328	9.0	0.8963	0.88983	0.7
1.1	– 5.0416	– 4.5538	10.6	0.8581	0.85636	0.2
1.2	– 6.4796	– 6.13	5.7	0.7957	0.79077	0.6
1.3	– 7.4624	– 7.1364	4.5	0.7370	0.73186	0.8
1.5	– 8.3201	– 8.1696	1.9	0.6390	0.63481	0.6
2.0	– 9.1221	– 9.0560	0.7	0.4776	0.47613	0.3

Table 2

Numerical results for arrival rates in region R. ($J_1 = J_2 = J_3 = 10$, $\lambda_2 = \lambda_1 = \mu = 0.1$)

λ_3/λ_1	Simul. $E(e_{II})$	Approx. $E(e_{II})$	Error % $E(e_{II})$	Simul. ρ_3	Approx. ρ_3	Error % ρ_3
1.0	– 2.1750	– 2.07	4.8	0.9333	0.91233	2.3
1.01	– 2.6244	– 2.3899	10.0	0.9325	0.91490	1.9
1.02	– 3.0524	– 2.6929	13.3	0.9324	0.91509	1.8
1.03	– 3.2355	– 2.9845	8.3	0.9384	0.91743	2.2
1.05	– 3.8560	– 3.5328	9.0	0.9459	0.92389	2.3
1.07	– 4.3612	– 3.8672	12.9	0.9473	0.93224	1.6
1.1	– 5.0416	– 4.5538	10.6	0.9508	0.93624	1.5

Table 3

Numerical results for different values of the control bounds. ($\lambda_1 = \lambda_2 = \lambda_3 = \mu = 0.1$)

$J_1 = J_2 = J_3$	Simul. $E(e_{II})$	Approx. $E(e_{II})$	Error % $E(e_{II})$	Simul. ρ_3	Approx. ρ_3	Error % ρ_3
5	– 1.1425	– 1.1170	2.7	0.8664	0.83852	3.3
10	– 2.1750	– 2.0759	4.8	0.9333	0.91233	2.3
15	– 3.4459	– 3.0373	13.5	0.9503	0.93915	1.1
20	– 4.4109	– 3.9955	10.5	0.9655	0.95313	1.2
25	– 5.2759	– 4.9629	6.2	0.9730	0.96532	0.8

$$R = \left\{ \lambda_1, \lambda_2, \lambda_3 : \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \lambda_1 + 0.1\, \lambda_1 \right\}.$$

The results show how, for arrival rates outside the range R, the approximation performs better and better as λ_3/λ_1 increases.

Table 2 presents a comparison between approximation and simulation for values of the arrival rates inside the set R. There is a clear trend toward better approximations as λ_3/λ_1 increases, i.e. as the system becomes less strongly coupled. The worst approximation for the expected excess at the second synchronizer gives an error of about 13%. The system utilization is always well approximated and, again, the error decreases as λ_3/λ_1 increases. The approximation error seems also to be decreasing with increasing J_1, J_2, and J_3, as described in table 3. In this case, there should be a decreasing influence of the control mechanism with increasing values for the bounds.

The approximation is expected to become worse as N becomes larger. However, since the main cause of error is due to the decoupling of successive stages in the decomposition, while the MMP approximation is quite precise, we still expect a good performance of the procedure (10% error or less) whenever the arrival rates are not approximately equal.

9. Conclusion

In this work we presented an approximation for assembly-like systems with N ($N > 2$) Poisson arrival streams based on the idea of decomposition. Also, as by-products of this investigation, we obtained an analytical solution for the case $N = 2$, and one of the arrival streams is a two-state Markov Modulated process. The study presented in this document lends itself to numerous extensions. In particular, more general control mechanisms than the one considered can be analyzed, as discussed in [2] for the case of the paired customer system. Also, more general service time distributions can be treated. All the generalizations for the case of the paired customer system considered in [3] can in fact be adopted here.

Acknowledgements

I would like to express my gratitude to all the members of Department 45312 at AT&T Bell Laboratories, who very patiently helped to dispel a little bit of my ignorance. Thanks are also due to David Lucantoni for his useful suggestions on the procedure discussed in section 7, and to Y.T. Wang for his stimulating support.

References

[1] J.D. Harrison, Assembly-like queues, J. Appl. Prob. 10(1973)354.
[2] G. Latouche, Queues with paired customers, J. Appl. Prob. 18(1981)684.
[3] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models* (Johns Hopkins, 1981).
[4] V. Ramaswami, The N/G/1 queue and its detailed analysis, Adv. Appl. Prob. 12(1980)222.
[5] H.M. Srivastava and B.R.K. Kashyap, *Special Functions in Queueing Theory* (Academic Press, 1982) pp. 69 – 75.
[6] V. Ramaswami, A note on Neuts' versatile Markovian point process, Tech. Rep. No. 7-79, Dept. of Mathematics, Drexel University, Philadelphia, PA (1979).
[7] U.N. Bhat, Finite capacity assembly-like queues, Queueing Systems: Theory and Applications 1(1986)85.
[8] E.H. Lipper and B. Sengupta, Assembly-like queues with finite buffer, Queueing Systems: Theory and Applications 1(1986)67.
[9] F. Baccelli and W.A. Massey, Series-parallel, fork-join queueing networks and their stochastic ordering, to appear.
[10] H. Heffes, A class of data traffic processes – covariance function characterization and related queueing results, BSTJ 59(1980)897.
[11] U. Herzog, L. Woo and K.M. Chandy, Solution of queueing problems by a recursive technique, IBM J. Res. Develop. (May, 1975) 295.
[12] D. Lucantoni, private communication.
[13] J.J. Hunter, *Mathematical Techniques of Applied Mathematics*, Vol. 1 – 2 (Academic Press, 1983).