Transient Solution of a Two-Processor Heterogeneous System

S. Dharmaraja
Center for Advanced Computing and Communication
Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708-0921, U.S.A.

(Received July 1999; revised and accepted May 2000)

Abstract—A two-processor system heterogeneous with Poisson arrival of jobs having exponentially distributed execution times is considered. Here, the service rates of these processors are not identical. Each job requires exactly one processor for its execution and the scheduling policy is FCFS. When both the processors are idle, the faster processor is scheduled for service before the slower one. For this system, exact time-dependent system size probabilities are obtained using a suitable probability generating function. Finally, some important performance measures are also obtained. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords—Markovian queues, Multiprocessor systems, Generating function.

1. INTRODUCTION

Multiprocessor systems are fast emerging as a powerful computing tool for many applications. Advances in computing technology have made it possible to build multiprocessor systems. Performance is one of the key factors that influences the design, development, and tuning of multiprocessor systems. The most important issue in a multiprocessor system is scheduling of tasks onto a given set of processors [1]. The power of multiprocessor systems cannot be fully exploited if poor scheduling strategy is employed. In this context, the application of queueing theory to study the performance of such systems assumes great significance [2,3].

Of specific interest is the stationary system size probabilities. This may be due to the fact that the balance equations involved are simple and relatively straightforward techniques can be employed. However, steady state measures do not reveal the complete picture of the system behaviour, because they ignore the transient and start-up effects. In many potential applications, steady-state measures of system performance simply do not make sense because the system may never attain equilibrium [4].

Transient analysis is useful in obtaining optimal solutions leading to the control of the system. Also, the transient solutions exist for a wider class of problems in comparison with the steady-state

This work was done when the author was a Ph.D. student at IIT Madras, India.
The author would like to thank Prof. P. R. Parthasarathy, Indian Institute of Technology Madras, India, for his valuable suggestions and the referees for their constructive criticism that led to considerable improvement in the presentation of this paper.

0895-7177/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved. Typeset by A\LaTeX
Pll: 0895-7177(00)00194-1
solutions which exist only under convergence conditions. Hence, the time-dependent analysis is important if one has to obtain meaningful results in real world problems [4].

In particular, transient behaviour of Markovian queues has received considerable attention because of their nature of admitting closed form solutions [5,6]. Transient solutions can be obtained numerically for complex systems [7,8]. As closed-form solutions provide deep insight into the system behaviour, we study the time-dependent performance of multiprocessor systems with state-dependent arrivals which admits closed form solution [9].

In multiprocessors systems, various job assignment schemes have been proposed and implemented to provide higher performance than single-processor systems. Many researchers have used an approach involving a queueing system with multiservers, primarily homogeneous servers, in order to deal with the problem of evaluating the performance of multiprocessor systems. Most steady state analyzes dealing with systems with heterogeneous multiservers have limited the model because of the mathematical complexity of analysis, e.g., they limited the number of servers and priority classes, while still other multiserver systems offer multiple customer classes [10]. Because we seek only the closed form transient measures, we study the system having two processors with single class job, which is analytically tractable.

The analysis using conventional probability generating functions is known to be helpful in finding the time-dependent probabilities of queues when the arrival and service rates are constant or linear [5]. In this paper, we use this technique by defining a suitable probability generating function to study the transient behaviour of a two-processor system where one processor is faster than the other and no queue is allowed in front of the slower processor. Trivedi [11] analyzes this model and restricts his analysis to steady-state measures. In this paper, our objective is to present the closed form solution for the time-dependent performance measures.

The remainder of the paper is organised as follows. In Section 2, we describe the model of a two-processor heterogeneous system and give the system of differential-difference equations of probabilities governing this model. In Section 3, we completely solve this system of equations. In Section 4, we present some important performance measures that are derived from the system size probabilities.

2. MODEL DESCRIPTION

Consider a two-processor system with Poisson arrival of jobs and the service rates of these processors are $\mu_1$ and $\mu_2$. Mean arrival rates of each jobs are $\lambda$ and assume without loss of generality that $\mu_1 > \mu_2$. Each job requires exactly one processor for its execution and the scheduling policy is FCFS. When both processors are idle, the faster processor is scheduled for service before the slower one [11]. The state of the system is defined to be a pair $(n_1, n_2)$ where $n_1 \geq 0$ denotes the number of jobs in the queue, including any at the faster processor, and $n_2 \in \{0, 1\}$ denotes the number of jobs at the slower processor. The queueing structure is shown in Figure 1. The state-transition diagram of the system is given in Figure 2.

We assume that there are zero jobs in the system at time $t = 0$. The Kolmogorov forward equations for the system are

\begin{align}
P'_{00}(t) &= -\lambda P_{00}(t) + \mu_1 P_{10}(t) + \mu_2 P_{01}(t), \\
P'_{10}(t) &= \lambda P_{00}(t) - (\lambda + \mu_1) P_{10}(t) + \mu_2 P_{11}(t), \\
P'_{01}(t) &= - (\lambda + \mu_2) P_{01}(t) + \mu_1 P_{11}(t), \\
P'_{11}(t) &= \lambda P_{10}(t) + \lambda P_{01}(t) - (\lambda + \mu_1 + \mu_2) P_{11}(t) + (\mu_1 + \mu_2) P_{21}(t), \\
P'_{n1}(t) &= \lambda P_{n-1,1}(t) - (\lambda + \mu_1 + \mu_2) P_{n1}(t) + (\mu_1 + \mu_2) P_{n+1,1}(t), \quad n \geq 2.
\end{align}

In the following section, we solve the above system of equations by using a suitable probability generating function.
3. TRANSIENT SOLUTION

Define

\[ P(z, t) = q_0(t) + \sum_{n=0}^{\infty} Q_{2+n}(t)z^{n+1}, \]
\[ P(z, 0) = 1, \]

with

\[ q_0(t) = P_{00}(t) + P_{01}(t) + P_{10}(t) + P_{11}(t), \]
\[ Q_n(t) = P_{n1}(t), \quad n \geq 1, \]

and \( \mu_1 + \mu_2 = \mu. \)

The system of equations (2.1)–(2.5) then yields

\[ \frac{\partial}{\partial t} P(z, t) = \lambda(z - 1)Q_1 + \left[ \frac{\mu}{z} + \lambda z - (\lambda + \mu) \right] P(z, t) - q_0(t). \]

The solution of this differential equation is obtained as

\[ P(z, t) = P(z, 0) \exp \left\{ - \left[ \lambda + \mu - \frac{\mu}{z} - \lambda z \right] t \right\} \]
\[ + \int_0^t \left[ \lambda(z - 1)Q_1 + \left\{ \lambda + \mu - \frac{\mu}{z} - \lambda z \right\} q_0(u) \right] \exp \left\{ - \left[ \lambda + \mu - \frac{\mu}{z} - \lambda z \right] (t - u) \right\} du. \]  

(3.6)

It is known that if \( \alpha = 2\sqrt{\lambda\mu} \) and \( \beta = \sqrt{\lambda/\mu} \), then

\[ \exp \left\{ \left( \lambda z + \frac{\mu}{z} \right) t \right\} = \sum_{n=-\infty}^{\infty} (\beta z)^n I_n(\alpha t), \]

where \( I_n(.) \) is the modified Bessel function of first kind of order \( n. \) Using this in (3.6) and comparing the coefficients of \( z^n \) on either side, we get, for \( n = 1, 2, \ldots, \)

\[ \beta^{-n}Q_{1+n}(t) = \exp (-\lambda u + \mu t) \frac{\beta I_n(\alpha t)}{\alpha} + \int_0^t \exp (-\lambda u + \mu (t - u)) \left\{ \lambda I_{n-1}(\alpha (t - u)) \right\} du, \]

(3.7)
and for \( n = 0 \),

\[
\beta q_0(t) = \exp(-\lambda \mu t) \beta I_0(\alpha t) + \int_0^t \exp(-(\lambda + \mu)(t-u)) [\lambda \{I_1(\alpha (t-u))
-\beta I_0(\alpha (t-u))\}] Q_1(u) + \{\beta(\lambda + \mu) I_0(\alpha (t-u)) - 2\lambda I_1(\alpha (t-u))\} q_0(u) \, du.
\] (3.8)

As \( P(z,t) \) does not contain terms with negative powers of \( z \), the right-hand side of (3.7) with \( n \) replaced by \(-n\), must be zero. Thus,

\[
\exp(-(\lambda + \mu) t) \beta I_n(\alpha t) + \int_0^t \exp(-(\lambda + \mu)(t-u)) [\lambda \{I_{n+1}(\alpha (t-u)) - \beta I_n(\alpha (t-u))\} \\
\times Q_1(u) + \{\beta(\lambda + \mu) I_n(\alpha (t-u)) - \lambda [I_{n-1}(\alpha (t-u)) + I_{n+1}(\alpha (t-u))]\} q_0(u) \, du = 0,
\] (3.9)

where we have used \( I_{-n}(.) = I_n(.) \). Usage of (3.9) in (3.9) considerably simplifies the work and results in simple expressions for \( Q_n(t) \). This yields, for \( n = 1, 2, \ldots \),

\[
Q_{n+1}(t) = \beta^n \int_0^t \frac{I_n(\alpha (t-u))}{t-u} \exp\{-(\lambda + \mu)(t-u)\} Q_1(u) \, du.
\] (3.10)

Now, the probabilities \( P_{00}(t) \), \( P_{01}(t) \), \( P_{10}(t) \), and \( P_{11}(t) \) remain to be solved. For this we consider the system of equations (2.1)–(2.3) subject to condition (3.8). Equations (2.1)–(2.3) can be expressed in the form

\[
\frac{d}{dt} Q(t) = AQ(t) + \mu_2 Q_1(t) e_1 + \mu_1 Q_1(t) e_2,
\] (3.11)

where

\[
Q(t) = (P_{00}(t), P_{01}(t), P_{10}(t), P_{11}(t))^T,
\]

\[
A = \begin{pmatrix}
-\lambda & \mu_1 & \mu_2 \\
\lambda & -(\lambda + \mu_1) & 0 \\
0 & 0 & -(\lambda + \mu_2)
\end{pmatrix},
\]

\( e_1 = (0,1,0)^T \), and \( e_2 = (0,0,1)^T \).

In the sequel, for any function \( f(.) \), let \( \hat{f}(s) \) denote its Laplace transform. Now, by taking Laplace transform, the solution of (3.11) is obtained as

\[
\hat{Q}(s) = [sI - A]^{-1}\left[\mu_2 \hat{Q}_1(s) e_1 + \mu_1 \hat{Q}_1(s) e_2 + Q(0)\right],
\] (3.12)

with

\[
Q(0) = (1,0,0)^T.
\] (3.13)

Thus, only \( \hat{Q}_1(s) \) is to be found. We observe that if

\[
e = (1,1,1)^T, \quad e^T \hat{Q}(s) + \hat{Q}_1(s) = \hat{q}_0(s).
\]

Using (3.8) in the above equation and simplifying, we get, with \( p = s + \lambda + \mu \),

\[
\hat{Q}_1(s) = \frac{-se^T(sI - A)^{-1}Q(0)}{s + \lambda - \left(\frac{(p - \sqrt{p^2 - \alpha^2}) / 2}{s + \lambda}\right) + se^T(sI - A)^{-1}[\mu_2 e_1 + \mu_1 e_2].
\] (3.14)

Let

\[
(sI - A)^{-1} = (\hat{a}_{kj}(s))_{3x3}.
\]
We can use the usual method to find \((sI - A)^{-1}\) and is given by

\[
\frac{1}{|D|} \begin{pmatrix}
(s + \lambda + \mu_1)(s + \lambda + \mu_2) & \mu_1(s + \lambda + \mu_2) & \mu_2(s + \lambda + \mu_1) \\
\lambda(s + \lambda + \mu_2) & (s + \lambda)(s + \lambda + \mu_2) & \lambda \mu_2 \\
0 & 0 & (s + \lambda)(s + \lambda + \mu_1) - \lambda \mu_1
\end{pmatrix},
\]

where \(|D| = (s + \lambda + \mu_2)(s^2 + s(2\lambda + \mu_1) + \lambda^2)\).

Let \(s_k, (k = 1, 2, 3)\) be the characteristic roots of the matrix \(A\). Then

\[
s_1 = -\lambda + \mu_2, \\
s_2, s_3 = -\frac{(2\lambda + \mu_1) \pm \sqrt{4\lambda \mu_1 + \mu_1^2}}{2}.
\]

We observe that \(\hat{a}_{kj}(s)\) are rational algebraic functions in \(s\). The cofactor of the \((i, j)^{th}\) element of \((sI - A)\) is a polynomial of degree \(2 - |i - j|\). Since the characteristic roots of \(A\) are distinct, the inverse transforms \(a_{kj}(t)\) of \(\hat{a}_{kj}(s)\) can be obtained by partial fraction decomposition and are given below.

\[
a_{11}(t) = \frac{\mu_1}{2\sqrt{4\lambda \mu_1 + \mu_1^2}} (e^{s_2 t} - e^{s_3 t}),
\]

\[
a_{12}(t) = \frac{\mu_2 (\mu_1 - \mu_2)}{2\sqrt{4\lambda \mu_1 + \mu_1^2}} (e^{s_2 t} - e^{s_3 t}),
\]

\[
a_{13}(t) = \frac{\mu_2 (\mu_1 - \mu_2)}{2\lambda \mu_1 - \mu_1^2} e^{s_1 t} - \frac{\mu_2}{2\sqrt{4\lambda \mu_1 + \mu_1^2}} \left( \frac{\mu_1 + \sqrt{4\lambda \mu_1 + \mu_1^2}}{\mu_1 - 2\mu_2 - \sqrt{4\lambda \mu_1 + \mu_1^2}} e^{s_2 t} - \frac{\mu_1 - \sqrt{4\lambda \mu_1 + \mu_1^2}}{\mu_1 - 2\mu_2 + \sqrt{4\lambda \mu_1 + \mu_1^2}} e^{s_3 t} \right),
\]

\[
a_{21}(t) = \frac{\lambda}{2\sqrt{4\lambda \mu_1 + \mu_1^2}} (e^{s_2 t} - e^{s_3 t}),
\]

\[
a_{22}(t) = \frac{\sqrt{4\lambda \mu_1 + \mu_1^2} - \mu_1}{2\sqrt{4\lambda \mu_1 + \mu_1^2}} e^{s_2 t} + \frac{\sqrt{4\lambda \mu_1 + \mu_1^2} + \mu_1}{2\sqrt{4\lambda \mu_1 + \mu_1^2}} e^{s_3 t},
\]

\[
a_{23}(t) = \frac{\lambda \mu_2}{2\lambda \mu_1 - \mu_1^2} e^{s_1 t} - \frac{2\lambda \mu_2}{\sqrt{4\lambda \mu_1 + \mu_1^2}} \left( \frac{1}{\mu_1 - 2\mu_2 - \sqrt{4\lambda \mu_1 + \mu_1^2}} e^{s_2 t} - \frac{1}{\mu_1 - 2\mu_2 + \sqrt{4\lambda \mu_1 + \mu_1^2}} e^{s_3 t} \right),
\]

\[
a_{33}(t) = e^{s_1 t},
\]

and \(a_{31}(t) = a_{32}(t) = 0\).

After considerable simplification, (3.14) reduces to

\[
\frac{d^2}{dt^2} = \frac{2}{p + \sqrt{p^2 - \alpha^2}} \left\{ -\hat{b}_1(s) \right\} \left[ 1 - \left( \frac{\mu}{\lambda} \right)^{1/2} \frac{p - \sqrt{p^2 - \alpha^2}}{\alpha} \left( 1 - \frac{\hat{b}_2(s)}{\mu} \right) \right]^{-1} - \frac{p - \sqrt{p^2 - \alpha^2}}{\alpha^2} \left\{ -2\hat{b}_2(s) \right\} \left[ 1 - \left( \frac{\mu}{\lambda} \right)^{1/2} \frac{p - \sqrt{p^2 - \alpha^2}}{\alpha} \left( 1 - \frac{\hat{b}_2(s)}{\mu} \right) \right]^{-1},
\]

where

\[
\hat{b}_1(s) = s [\hat{a}_{11}(s) + \hat{a}_{21}(s)],
\]

\[
\hat{b}_2(s) = s [\mu_2 (\hat{a}_{12}(s) + \hat{a}_{22}(s)) + \mu_1 (\hat{a}_{13}(s) + \hat{a}_{23}(s) + \hat{a}_{33}(s))].
\]
Inversion yields
\[
P_{11}(t) = \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda} \right)^{m/2} \sum_{k=0}^{m} (-1)^k \binom{m}{k} \int_{0}^{t} \frac{1}{\mu} b_2^k(t - u) \left[ \exp(-\lambda u + \mu u) \{I_m(\alpha u) - I_{m+2}(\alpha u)\} \right] \, du,
\]
where \( b_2^k \) is the \( k \)-fold convolution of \( b_2(t) \) with itself. We note that \( b_2^0(t) = \delta(t) \) (delta function). Inverse Laplace transforms of \( b_i(s) \) \( i = 1, 2 \) reduce to
\[
b_1(t) = \frac{\lambda^2}{\sqrt{4\lambda \mu_1 + \mu_1^2}} \left( e^{\lambda t} - e^{\mu_1 t} \right),
b_2(t) = \frac{\lambda \mu_1 (\mu_2 - \mu_1) (\lambda + \mu_2)}{\lambda \mu_1 - \mu_2 (\mu_2 - \mu_1)} e^{\lambda t}
\]
\[+ \frac{\left( (\mu_1 + \mu_2) \left( \mu_1 + \sqrt{4\lambda \mu_1 + \mu_1^2} \right) \right)}{4\sqrt{4\lambda \mu_1 + \mu_1^2}} \left( \sqrt{4\lambda \mu_1 + \mu_1^2} - 2\lambda - \mu_1 \right) e^{\mu_2 t} \]
\[+ \frac{\lambda \mu_1 (\mu_2 - \mu_1) \left( 2\lambda + \mu_1 - \sqrt{4\lambda \mu_1 + \mu_1^2} \right) \right)}{4\lambda \mu_1 + \mu_2 - \mu_1} e^{\mu_2 t} \]
\[+ \frac{\left( (\mu_1 + \mu_2) \left( \mu_1 - \sqrt{4\lambda \mu_1 + \mu_1^2} \right) \right)}{4\sqrt{4\lambda \mu_1 + \mu_1^2}} \left( \sqrt{4\lambda \mu_1 + \mu_1^2} + 2\lambda + \mu_1 \right) e^{\mu_1 t} \]
\[+ \frac{\lambda \mu_1 (\mu_2 - \mu_1) \left( 2\lambda + \mu_1 + \sqrt{4\lambda \mu_1 + \mu_1^2} \right) \right)}{4\lambda \mu_1 + \mu_2 + \mu_1} e^{\mu_1 t}. \]

Using (3.13) in (3.13) and inverting, we obtain
\[
P_{00}(t) = a_{11}(t) + \int_{0}^{t} (\mu_2 a_{12}(u) + \mu_1 a_{13}(u)) P_{11}(t - u) \, du, \quad (3.16)
P_{10}(t) = a_{21}(t) + \int_{0}^{t} (\mu_2 a_{22}(u) + \mu_1 a_{23}(u)) P_{11}(t - u) \, du, \quad (3.17)
P_{01}(t) = a_{31}(t) + \int_{0}^{t} (\mu_2 a_{32}(u) + \mu_1 a_{33}(u)) P_{11}(t - u) \, du. \quad (3.18)
\]

Thus, (3.10), (3.15)–(3.18) completely determine all the state probabilities.

4. PERFORMANCE MEASURES

Let \( N(t) \) denote the number of jobs in the system at time \( t \). The average number of jobs in the system at time \( t \) is computed by observing the number of jobs in state \( (n_1, n_2) \) which is \( n_1 + n_2 \). Therefore, the average number of jobs is given by
\[
E[N(t)] = P_{10}(t) + P_{01}(t) + \sum_{k \geq 1} (k + 1) P_{k1}(t).
\]

The probability that an arriving job is required to join the queue at time \( t \) is given as
\[
P[\text{queueing}](t) = \sum_{k \geq 1} P_{k1}(t).
\]
Let the random variable $M(t)$ denote the number of busy processors at time $t$; the probability that the system has $n$ busy processors is given by

$$P(M(t) = n) = \begin{cases} P[N(t) = 1] = P_{10}(t) + P_{01}(t), & n = 1, \\ P[N(t) > 1] = \sum_{k \geq 1} P_{k1}(t), & n = 2. \end{cases}$$

The average number of busy processors at time $t$ is given by

$$E[M(t)] = P_{10}(t) + P_{01}(t) + \sum_{k \geq 1} 2P_{k1}(t),$$

which can be simplified as

$$E[M(t)] = 2 \left[ 1 - P_{00}(T) \right] - \left[ P_{10}(t) + P_{01}(t) \right].$$

The utilizations of the faster and slower processor at time $t$ are given by $1 - P_{00}(t)$ and $1 - P_{00}(t) - P_{10}(t)$, respectively. We can compute the above time-dependent measures by using (3.10) and (3.15)–(3.18).

REFERENCES