Analytic and Numerical Solutions of An N Duplicate-server System

**K.P. Jun
and
Harry Perros**

**Center for Communications and Signal Processing
Department of Electrical Computer Engineering
North Carolina State University**

March 1987

CCSP-TR-87/3

ANALYTIC AND NUMERICAL SOLUTIONS OF AN N DUPLICATE-SERVER SYSTEM

K. P. Jun

H. G. Perros

Department of Computer Science

Center for Communications and Signal Processing

North Carolina State University

Raleigh, NC 27695

Abstract

We consider an N duplicate-server system, where each server consists of two reconfigurable duplicated units which are subject to breakdowns. This system is analyzed analytically using generating functions and also numerically using the matrix-geometric procedure. Using the generating function approach we obtain a recursive expression of the queue length probability distribution for $N=1$. This expression is difficult to generalize to $N \geq 2$. The numerical method is applicable for any value of N . For any N , we also obtain the condition for stability and the availability of the system.

1. Introduction

In recent years, the interest in fault-tolerant computing systems has been increased because many computing systems require very high reliability and/or availability. There are several techniques to increase reliability and/or availability, such as fault-detection, masking redundancy, and dynamic redundancy. Several dynamic redundancy techniques utilize a combination of fault-detection, fault-masking, and reconfiguration. Among dynamic redundancy techniques, reconfigurable duplication and graceful degradation are of interest.

A reconfigurable duplicate system consists of two duplicated units which function in parallel. Only one of the duplicated units is connected to the system's output. A hardware comparator monitors both duplicate units and a fault is detected by a mismatch. When a fault is detected, the faulty unit is switched out. Normal operation continues on the other duplicate unit. There are several methods for determining the faulty unit. Four methods are discussed in Siewiorek and Swarz[23], these are a) to run a diagnostic program when a mismatch occurs, b) to include self-checking capabilities on each unit, c) to use a watching timer, and d) to use an outside arbiter.

Graceful degradation techniques use the redundant hardware as part of the system's normal resources at all times. For example, consider a gracefully degrading system with N identical processors which are subject to breakdowns. If the system is fault-free, then all processors are active and are able to process

jobs concurrently. Each processor is self-testing, and in the presence of a single faulty processor the system is able to recover (with a specified coverage) to an $(N-1)$ -processor configuration, provided that $N \geq 2$. In this configuration, the system behaves the same as a fault-free version of the system with $N-1$ processors. When only a single processor remains fault-free, fault recovery is no longer possible. Failure to recover from a processor fault results in a total loss of processing capability (system is down).

Since a gracefully degrading system is designed so that continued (though degraded) processing is possible in the presence of failures, performance varies widely over time. So, gracefully degrading systems cannot be modeled by a separation of reliability/availability measures from performance measures. In other words, a combined performance/reliability measure is necessary in order to evaluate gracefully degrading systems.

Such performance/reliability measures have been obtained by many authors using various approaches. Relevant studies include, for example, Beaudry [3] (performance-related reliability measures), Gay and Ketelsen [5] (performance evaluation of degradable systems), Mine and Hitayama [12] (job-related reliability), Castillo and Siewiorek [4] (performance-reliability models for computing systems), Osaki and Nishio [20] (reliability of information), Huslende [6] (combined performance-reliability evaluation for degradable systems), Meyer [10, 11] (performability), and Kulkarni, Nicola, and Trivedi [8, 9] (completion

time of a job).

The performance/reliability of reconfigurable duplication systems has not been adequately studied. In this paper, we present two different solution techniques for analyzing gracefully degrading systems consisting of N servers, where each server is a reconfigurable duplicate system. The system studied in this paper is described in detail in the following section. In section 3, the system is studied analytically using the method of generating functions, and numerically using the matrix-geometric procedure. In particular, in section 3.1 we obtain a recursive expression for the queue-length probability distribution. This procedure is difficult to generalize to $N \geq 2$. In section 3.2, we briefly summarize the matrix-geometric procedure and we obtain the condition for stability and the system's availability.

2. The N duplicate-server System

Let us first consider a gracefully degrading system consisting of N identical servers in parallel. The servers are subject to breakdowns. Failed servers are repaired immediately by one of N repair crews and are brought back to an operative state. A breakdown occurring while a job is being served results in preemption of the job. That is, the interrupted job is not lost but it joins the head of the queue, and it has to be restarted (preemptive-restart) whenever a server becomes available.

This model was analyzed by Mitrani and Avi-Itzhak [13], Meyer [11], Gay and Ketelsen [5], Munarin [15], Nakamura and Osaki [16, 21], Neuts and Lucantoni [17, 18], and Nicola, Kulkarni, and Trivedi [19]. In some of these papers [13, 17, 19] the exact solution with perfect coverage ($c = 1$) was obtained. The remaining papers [5, 11, 15, 16] were concerned with approximate solutions.

In this paper, we analyze this system assuming that each of the N identical servers consists of two reconfigurable duplicated units as shown in Figure 1. We refer to this system as an **N duplicate-server system**. The service mechanism of this system can be described as follows: Any functioning unit of a server may become inoperative due to failure. Once a faulty unit is detected, automatic reconfiguration is executed. In this case, service continues uninterrupted on its duplicate unit. The faulty unit is switched out and it becomes operative again after it has been repaired. During the time that the faulty unit is being repaired, the other duplicate unit may also break down. In this case, the server is completely broken down. The server may also break down if automatic reconfiguration is not successful upon detection of a faulty unit. When a server breaks down, the total number of servers is reduced by one. When all servers are down, the system is down as well. The interrupted job is not lost but it joins the head of the queue and it is restarted (preemptive restart) whenever a server becomes available. The server becomes operative when

one of its two duplicate units is repaired.

For this system, we assume that the automatic switchover time is negligible, the repair of the faulty unit starts as soon as it is detected, and repaired units are perfect. There is an infinite waiting room and the service discipline is first-come-first-served (FCFS). We also assume that all relevant random variables such as the inter-arrival time, the service time, the time to failure, and the repair time of the server have exponential distribution with parameters λ , μ , p , and r , respectively. Finally, let c be the coverage, i.e., the probability that the system reconfigures successfully given that the faulty unit is detected.

Figure 1. N duplicate-server system

3. Analytic and Numerical Solutions

We analyze the N duplicate-server system using a) the generating function approach, and b) the matrix-geometric numerical procedure. The generating function approach is difficult to extend to $N \geq 2$. The numerical method is applicable for any value of N . Below, we obtain the solution to the system using the generating function approach for a single duplicate-server system. In section 3.2, we describe how the matrix-geometric solution can be employed to analyze numerically N duplicate-server systems.

3.1 The Generating Function Approach

Let $P_{s,n}$ be the state probability of the system being in state (s,n) , where $n = 0,1,2,\dots$, is the number of jobs in the system including the one in service, and $s = 0,1,2$, is the number of failed units. The steady-state balance equations of the system are as follows:

$$\begin{aligned}
 (\lambda + 2p)P_{0,0} &= \mu P_{0,1} + rP_{1,0} \\
 (\lambda + \mu + 2p)P_{0,n} &= \lambda P_{0,n-1} + \mu P_{0,n-1} + rP_{1,n}, \quad \text{for } n \geq 1, \\
 (\lambda + p + r)P_{1,0} &= \mu P_{1,1} + 2pcP_{0,0} + 2rP_{2,0} \\
 (\lambda + \mu + p + r)P_{1,n} &= \lambda P_{1,n-1} + \mu P_{1,n-1} + 2pcP_{0,n} + 2rP_{2,n}, \quad \text{for } n \geq 1, \\
 (\lambda + 2r)P_{2,0} &= 2p(1-c)P_{0,0} + pP_{1,0} \\
 (\lambda + 2r)P_{2,n} &= \lambda P_{2,n-1} + 2p(1-c)P_{0,n} + pP_{1,n}, \quad \text{for } n \geq 1,
 \end{aligned} \tag{1}$$

where c is coverage.

Let the generating function $G_s(z)$, $|z| \leq 1$, $s = 0, 1, 2$, be defined by

$$G_s(z) = \sum_{n=0}^{\infty} z^n P_{s,n}. \quad (2)$$

By (2), $G_s(1)$ is the steady-state probability that there are (2-s) operative units(i.e. s failed units). Multiplying both sides of equations in (1) by z^n and summing over all n, we obtain the following equations:

$$\begin{aligned} [\lambda z(1-z) + 2pz - \mu(1-z)] G_0(z) - rz G_1(z) &= -\mu(1-z) P_{0,0} \\ -2pcz G_0(z) + [\lambda z(1-z) + (p+r)z - \mu(1-z)] G_1(z) - 2rz G_2(z) &= -\mu(1-z) P_{1,0} \\ -2p(1-c) G_0(z) - p G_1(z) + [\lambda(1-z) + 2r] G_2(z) &= 0. \end{aligned} \quad (3)$$

The above equations can be written in matrix form,

$$A(z)g(z) = (1-z)b(z), \quad (4)$$

where

$$A(z) = \begin{bmatrix} \lambda z(1-z) + 2pz - \mu(1-z) & -rz & 0 \\ -2pcz & \lambda z(1-z) + (p+r)z - \mu(1-z) & -2rz \\ -2p(1-c)z & -p & \lambda(1-z) + 2r \end{bmatrix},$$

$$g(z) = \begin{bmatrix} G_0(z) \\ G_1(z) \\ G_2(z) \end{bmatrix},$$

and

$$b(z) = \begin{bmatrix} -\mu P_{0,0} \\ -\mu P_{1,0} \\ 0 \end{bmatrix}.$$

It is easy to show that $|A(z)| = 0$ at $z = 1$. We can write

$$|A(z)| = (1-z)Q(z), \quad (5)$$

where $Q(z)$ is a polynomial of degree 4 in z since $|A(z)|$ is a polynomial of degree 5 in z . Equation (4) may now be rewritten in the form

$$Q(z)G_s(z) = |A_s(z)|, \quad s=0,1,2 \quad (6)$$

where $A_s(z)$ is obtained by replacing the $(s+1)$ st column of $A(z)$ with $b(z)$.

Let $G(z)$ denote the generating function of steady-state number of jobs in the system, which is given by

$$G(z) = \sum_{s=0}^2 G_s(z). \quad (7)$$

Now, in order to obtain $G_s(z)$, $s=0,1,2$, and $G(z)$ we need to determine the unknown probabilities $P_{0,0}$ and $P_{1,0}$. From the normalizing equation, we have

$$P_{0,0} + P_{1,0} = \frac{r^2 + 2rp}{r^2 + 2rp + p^2 + rp(1-c)} - \frac{\lambda}{\mu}. \quad (8)$$

To find one more additional equation relating $P_{0,0}$ and $P_{1,0}$ we note from (6) that $|A_s(z_0)|$ must equal to zero whenever $Q(z_0)=0$, $0 < z_0 < 1$. It can be shown that $Q(z)$ has only one real root in the interval $(0,1)$. This statement can be proved using similar arguments as in Mitrani and Avi-Itzhak [13]. For the two-server case, they have obtained an explicit solution for z_0 . Unfortunately, in our case an explicit solution expression for the root z_0 is difficult to obtain. In view of this, z_0 is determined numerically.

Given z_0 , we have from $|A_1(z_0)|=0$ that

$$P_{0,0} = -\frac{(\lambda(1-z_0)+2r) \cdot aa}{2pz_0(c\lambda(1-z_0)+2r)} \cdot P_{1,0}, \quad (9)$$

where $aa = \lambda z_0(1-z_0) + 2pz_0 - \mu(1-z_0)$. Substituting (9) into (8), we have

$$\begin{aligned} P_{1,0} &= \frac{2pz_0(c\lambda(1-z_0)+2r)}{2pz_0(c\lambda(1-z_0)+2r) - (\lambda(1-z_0)+2r) \cdot aa} \cdot bb \\ P_{0,0} &= bb - P_{1,0}, \end{aligned} \quad (10)$$

where

$$bb = \frac{r^2 + 2rp}{r^2 + 2rp + p^2 + rp(1-c)}.$$

We can now easily obtain the expected number of jobs in the system, $E(N)$ as follows:

$$\begin{aligned} E(N) &= \frac{dG(z)}{dz} \Big|_{z=1} \\ &= \frac{d}{dz} (G_0(z) + G_1(z) + G_2(z)) \Big|_{z=1} \\ &= \frac{d}{dz} \left[\frac{|A_0(z)| + |A_1(z)| + |A_2(z)|}{Q(z)} \right] \Big|_{z=1}. \end{aligned} \quad (11)$$

The steady-state probability of the number of jobs in the system can be obtained as follows (cf. Perros[22]): $G(z)$ can be expressed in the following form,

$$G(z) = \frac{S_0 - S_1 z + S_2 z^2 - S_3 z^3}{1 - z(T_1 - T_2 z + T_3 z^2 - T_4 z^3)}, \quad (12)$$

where S_i , $i=0,1,2,3$, and T_j , $j=1,2,3,4$, are represented in terms of parameters λ , μ , p , r , and z_0 . If $|z(T_1 - T_2 z + T_3 z^2 - T_4 z^3)| < 1$, $G(z)$ can be rewritten into

$$G(z) = (S_0 - S_1 z + S_2 z^2 - S_3 z^3) \sum_{k=0}^{\infty} z^k (T_1 - T_2 z + T_3 z^2 - T_4 z^3)^k. \quad (13)$$

Now, using the multinomial expansion formula, we can write:

$$\begin{aligned} & \sum_{k=0}^{\infty} z^k (T_1 - T_2 z + T_3 z^2 - T_4 z^3)^k \\ &= \sum_{k=0}^{\infty} \sum \frac{k!}{n_1! n_2! n_3! n_4!} (-1)^{n_2+n_4} T_1^{n_1} T_2^{n_2} T_3^{n_3} T_4^{n_4} z^{n_1+2n_2+3n_3+4n_4}, \\ &= \sum_{k=0}^{\infty} A_k z^k, \end{aligned} \quad (14)$$

where

$$A_k = \sum \frac{k!}{n_1! n_2! n_3! n_4!} (-1)^{n_2+n_4} T_1^{n_1} T_2^{n_2} T_3^{n_3} T_4^{n_4},$$

and the nonnegative integers n_1, n_2, n_3, n_4 in the above summation take values such that $n_1 + 2n_2 + 3n_3 + 4n_4 = k$. Thus, (13) can be rewritten as follows:

$$\begin{aligned} G(z) &= S_0 A_0 + (S_0 A_1 - S_1 A_0)z + (S_0 A_2 - S_1 A_1 + S_2 A_0)z^2 \\ &+ \cdots + (S_0 A_n - S_1 A_{n-1} + S_2 A_{n-2} - S_3 A_{n-3})z^n + \cdots \end{aligned} \quad (15)$$

Comparing (15) with (7), we have:

$$\begin{aligned}
P_0 &= P_{0,0} + P_{1,0} + P_{2,0} = S_0 A_0 \\
P_1 &= P_{0,1} + P_{1,1} + P_{2,1} = S_0 A_1 - S_1 A_0 \\
P_2 &= P_{0,2} + P_{1,2} + P_{2,2} = S_0 A_2 - S_1 A_1 + S_2 A_0 \\
&\vdots \\
&\vdots \\
&\vdots \\
P_n &= P_{0,n} + P_{1,n} + P_{2,n} \\
&= S_0 A_n - S_1 A_{n-1} + S_2 A_{n-2} - S_3 A_{n-3}, \quad n \geq 3.
\end{aligned} \tag{16}$$

From P_n , we can obtain $P_{s,n}$, $s=0,1,2$, by using $G_0(1)$, $G_1(1)$, and $G_2(1)$:

$$P_{s,n} = G_s(1) \cdot P_n \quad \text{for } s=0,1,2. \tag{17}$$

The quantities A_n , $n=0,1,2,\dots$, can be easily obtained recursively:

$$\begin{aligned}
A_0 &= 1, \\
A_1 &= T_1, \\
A_2 &= A_1 T_1 - A_0 T_2, \\
A_3 &= A_2 T_1 - A_1 T_2 + A_0 T_3, \\
A_n &= A_{n-1} T_1 - A_{n-2} T_2 + A_{n-3} T_3 - A_{n-4} T_4, \quad \text{for } n \geq 4.
\end{aligned} \tag{18}$$

We know from (2) that $G_s(1)$, $s=0,1,2$, is the probability of being (2-s) operative units in the system. From this, we can obtain the availability of this system,

$$\begin{aligned}
\text{Availability} &= G_0(1) + G_1(1) \\
&= \frac{r^2 + 2rp}{r^2 + 2rp + p^2 + rp(1-c)} \\
&= 1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)}.
\end{aligned} \tag{19}$$

The stability condition for one duplicate-server system can now be easily shown

to be

$$\frac{\lambda}{\mu} < \frac{r^2 + 2rp}{r^2 + 2rp + p^2 + rp(1-c)} \quad (20)$$

or

$$\frac{\lambda}{\mu} < 1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)}.$$

3.2 The Matrix-geometric Solution

An N duplicate-server system can be modeled by a discrete-state, continuous time Markov process. We define the states of the system (k_2, k_1, n) , where n denotes the number of jobs in the system, k_2 is the number of servers whose duplicate units are both operative, and k_1 is the number of servers with only one operative unit, for $n=0,1,\dots$, $k_2=0,1,\dots,N$, $k_1=0,1,\dots,N$, and $0 \leq k_2 + k_1 \leq N$. Note that the total number of operative servers in the system is equal to $k_2 + k_1$.

The numerical procedure requires the generation of all possible states subsequently and the generation of the rate matrix Q . Each row of Q corresponds to a specific state and each non-zero element of each row (except a diagonal element) is a transition rate. The diagonal elements of Q are strictly negative. The sum of all the elements of each row must equal to zero. The stationary probability vector $\underline{x} = (x_0, x_1, \dots)$, if it exists, can be found by solving the following system

$$\begin{aligned}\underline{x}Q &= 0, \\ \sum_{i=0}^{\infty} \underline{x}_i \underline{e} &= 1,\end{aligned}\tag{21}$$

where $\underline{x}_i$, $i=0,1,2,\dots$, are $(N+1)(N+2)/2$ vectors and column vector $\underline{e}$ has all elements equal to one. The above system of linear equation $\underline{x}Q = 0$ is solved using the matrix-geometric method(See Neuts [18]). This is because, when the states are ordered lexicographically, the rate matrix Q has a special block tri-diagonal structure as shown in below

$$Q = \begin{bmatrix} A_{00} & A_0 & & & & & \\ A_{10} & A_{11} & A_0 & & & & \\ & A_{21} & A_{22} & A_0 & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & A_{N-1,N-2} & A_{N-1,N-1} & A_0 & \\ & & & & A_2 & A_1 & A_0 \\ & & & & & A_2 & A_1 & A_0 \\ & & & & & & \cdot & \cdot & \cdot \\ & & & & & & & \cdot & \cdot & \cdot \end{bmatrix}.$$

We note that all submatrices A_i , and A_0 , A_1 , A_2 are square of order $(N+1)(N+2)/2$, where N is the number of servers.

In the stable case, there exists the stationary probability vector $\underline{x}$ of the Markov process such that

$$\underline{x}_i = \underline{x}_{N-1} R^{i-N-1}, \text{ for } i \geq N-1, N \geq 2.\tag{22}$$

The stationary probability vector $\underline{x}$ is called the matrix-geometric probability

vector. The vectors $\underline{x}_0, \underline{x}_1, \dots, \underline{x}_{N-1}$ are uniquely determined by

$$\begin{aligned} \underline{x}_0 A_{00} + \underline{x}_1 A_{10} &= 0, \\ \underline{x}_{i-1} A_{00} + \underline{x}_i A_{ii} + \underline{x}_{i-1} A_{i+1,i} &= 0, \quad 1 \leq i \leq N-2, \\ \underline{x}_{N-2} A_{00} + \underline{x}_{N-1} (A_{N-1,N-1} + R A_2) &= 0, \\ \sum_{i=0}^{N-2} \underline{x}_i \underline{e} + \underline{x}_{N-1} (I - R)^{-1} \underline{e} &= 1. \end{aligned} \quad (23)$$

The expected number of jobs in the system, $E(N)$, is given by

$$E(N) = \sum_{i=1}^{N-2} i \underline{x}_i + \underline{x}_{N-1} ((N-1)(I-R)^{-1} + R(I-R)^{-2}).$$

The matrix R is the minimal solution to a matrix nonlinear equation,

$$A_0 + R A_1 + R^2 A_2 = 0 \quad (24)$$

with $R \geq 0$, and it is an irreducible, nonnegative matrix of spectral radius less than one. An iterative approach may be used to compute R as follows:

$$\begin{aligned} R(0) &= 0, \\ R(n+1) &= A_0(-A_1)^{-1} + R^2(n)A_2(-A_1)^{-1}, \quad n \geq 0. \end{aligned} \quad (25)$$

For Markov processes with such generators, Neuts [18] obtained the following stability condition. Consider the infinitesimal generator $A = A_0 + A_1 + A_2$. It can be shown that A is irreducible and that there is a unique row vector $\underline{\pi}$ such that $\underline{\pi} \geq 0$, $\underline{\pi} A = 0$, and $\underline{\pi} \underline{e} = 1$. Then, the stability condition is given by

$$\underline{\pi} A_2 \underline{e} > \underline{\pi} A_0 \underline{e}. \quad (26)$$

From the above inequality, the condition of stability for the N duplicate-server system can be easily obtained. In particular, for the 2 duplicate-server system, the vectors $A_{0\underline{e}}$, $A_{2\underline{e}}$, and $\underline{\pi}$ are given by

$$\begin{aligned} A_{0\underline{e}} &= (\lambda, \lambda, \lambda, \lambda, \lambda, \lambda)^T, \\ A_{2\underline{e}} &= (2\mu, 2\mu, \mu, 2\mu, \mu, 0)^T, \end{aligned}$$

and

$$\begin{aligned} \underline{\pi} &= (\pi_{2,0}, \pi_{1,1}, \pi_{1,0}, \pi_{0,2}, \pi_{0,1}, \pi_{0,0}) \\ &= \frac{1}{D_1} (r^4, 4r^3p, 2r^2(p^2 + rp(1-c)), 4r^2p^2, 4rp(p^2 + rp(1-c)), (p^2 + rp(1-c))^2), \end{aligned}$$

where $D_1 = (r^2 + 2rp + p^2 + rp(1-c))^2$. Then, we have

$$\begin{aligned} \underline{\pi} A_{2\underline{e}} &= \frac{2\mu}{D_1} [r^4 + 4r^3p + r^2(p^2 + rp(1-c)) + 4r^2p^2 + 2rp(p^2 + rp(1-c))] \\ &= 2\mu \left[1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)} \right] \end{aligned}$$

and $\underline{\pi} A_{0\underline{e}} = \lambda$. The condition of stability for the 2 duplicate-server system is

$$\frac{\lambda}{2\mu} < 1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)}. \quad (27)$$

The availability of the 2 duplicate-server system is equal to $1 - \pi_{0,0}$, or

$$Availability = 1 - \left[\frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)} \right]^2.$$

For the 3 duplicate-server system, the vectors $A_{0\underline{e}}$, $A_{2\underline{e}}$, and $\underline{\pi}$ are as follows:

$$A_{0\underline{e}} = (\lambda, \lambda, \lambda, \lambda, \lambda, \lambda, \lambda, \lambda, \lambda, \lambda)^T,$$

$$A_{2\underline{e}} = (3\mu, 3\mu, 2\mu, 3\mu, 2\mu, \mu, 3\mu, 2\mu, \mu, 0)^T,$$

and

$$\begin{aligned}\underline{\pi} &= (\pi_{3,0}, \pi_{2,1}, \pi_{2,0}, \pi_{1,2}, \pi_{1,1}, \pi_{1,0}, \pi_{0,3}, \pi_{0,2}, \pi_{0,1}, \pi_{0,0}) \\ &= \frac{1}{D_2} (r^6, 6r^5p, 3r^4(p^2 + rp(1-c)), 12r^4p^2, 12r^3p(p^2 + rp(1-c)), 3r^2(p^2 + rp(1-c))^2, \\ &\quad 8r^3p^3, 12r^2p^2(p^2 + rp(1-c)), 6rp(p^2 + rp(1-c))^2, (p^2 + rp(1-c))^3),\end{aligned}$$

where $D_2 = (r^2 + 2rp + p^2 + rp(1-c))^3$. Then,

$$\begin{aligned}\underline{\pi} A_{2\underline{e}} &= \frac{3\mu}{D_2} [D - (p^2 + rp(1-c))(r^2 + 2rp + p^2 + rp(1-c))^2] \\ &= 3\mu \left[1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)} \right]\end{aligned}$$

and $\underline{\pi} A_{0\underline{e}} = \lambda$. From inequality (26) we have the condition of stability for the 3 duplicate-server system, i.e.,

$$\frac{\lambda}{3\mu} < 1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)}. \quad (28)$$

The availability of this system is equal to $1 - \pi_{0,0}$, or

$$Availability = 1 - \left[\frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)} \right]^3.$$

Examining the stability conditions (28) (and also (27)), we observe that the right-hand-side of the above inequality is the percentage of time a duplicate server is available. The quantity on the left-hand-side is the percentage of time that the duplicate server is busy. Therefore, in order to have a stable system

the above inequality has to hold. In view of this, the stability condition for arbitrary N can be conjectured to be

$$\frac{\lambda}{N\mu} < 1 - \frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)}. \quad (29)$$

The availability of the N duplicate-server system is equal to $1 - \pi_{0,0}$, or

$$Availability = 1 - \left[\frac{p^2 + rp(1-c)}{r^2 + 2rp + p^2 + rp(1-c)} \right]^N, \quad (30)$$

where $\pi_{0,0}$ denotes the probability that the system is down.

In the case of perfect coverage($c = 1$), the stability condition and the availability of the N duplicate-server system are given by

$$\frac{\lambda}{N\mu} < 1 - \left(\frac{p}{r+p} \right)^2$$

and

$$Availability = 1 - \left(\frac{p}{r+p} \right)^{2N}.$$

5. References

- [1] Baccelli, F. and K. Trivedi, "Analysis of M/G/2 - Standby Redundant System," *Performance '83*, A. K. Agrawala and S. K. Tripathi(editors), 1983, pp. 457-476
- [2] Barlow, R. E. and F. Proschan, *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, Inc., New York, NY, 1975
- [3] Beaudry, M. D., "Performance-Related Reliability Measures for Computing Systems," *IEEE Trans. on Comp.*, Vol. C-27, No. 6, June 1978, pp. 540-547
- [4] Castillo, X. and D. P. Siewiorek, "A Performance-Reliability Model for Computing Systems," *Proc. 1980 Int'l Symp. on FTC*, Kyoto, Japan, pp. 187-192
- [5] Gay, F. A. and M. L. Ketelsen, "Performance Evaluation for Gracefully Degrading Systems," *Proc. 1979 Int'l Symp. on FTC*, Madison, WI, pp. 51-57
- [6] Huslende, R., "A Combined Evaluation of Performance and Reliability for Degradable Systems," *Performance Evaluation Review*, Vol. 10, No. 3, Fall 1981, pp. 157-164
- [7] King, P. J. B. and I. Mitrani, "The Effect of Breakdown on the Performance of Multiprocessor Systems," *Performance '81*, F. J. Kylstra(editor),

1981, pp. 201-211

- [8] Kulkarni, V. G., V. F. Nicola and K. S. Trivedi, "A Unified Model for Performance and Reliability of Fault-Tolerant/Multi-Mode Systems," Technical Report, No. UNC/ORSA/TR-84-12, Univ. of North Carolina at Chapel Hill, Aug. 1984
- [9] Kulkarni, V. G., V. F. Nicola and K. S. Trivedi, "The Completion Time of a Job on Multi-Mode Systems," Technical Report, No. UNC/ORSA/TR-85-2, Univ. of North Carolina at Chapel Hill, Jan. 1985
- [10] Meyer, J. F, "On Evaluating the Performability of Degradable Computing Systems," *IEEE Trans. on Comp.*, Vol. C-29, No. 8, Aug. 1980, pp. 720-731
- [11] Meyer, J. F, "Closed-Form Solutions of Performability," *IEEE Trans. on Comp.*, Vol. C-31, No. 7, July 1982, pp. 648-657
- [12] Mine, H. and K. Hatayama, "Performance Evaluation of a Fault-Tolerant Computing System," *Proc. 1979 Int'l Symp. on FTC*, Madison, WI, pp.59-62
- [13] Mitrani, I. and B. Avi-Itzhak, "A Many-Server Queue with Service Interruptions," *Operations Research*, Vol. 16, 1968, pp. 628-638
- [14] Mitrani, I., "Networks of Unreliable Computers," *Computer Architectures and Networks*, E. Gelenbe and R. Muhl(Eds.), North-Holland, 1974, pp. 359-374

- [15] Munarin, J. A., "Dynamic Workload Model for Performance/Reliability Analysis of Gracefully Degrading Systems," *Proc. 1983 Int'l Symp. on FTC*, Milano, Italy, pp. 290-295
- [16] Nakamura, M. and S. Osaki, "Performance/Reliability Evaluation for Multiprocessor Systems with Computational Demands," *Int. J. System Sci.*, Vol. 15, No. 1, 1984, pp. 95-105
- [17] Neuts, M. F. and D. M. Lucantoni, "A Markovian Queue with N Servers Subject to Breakdowns and Repairs," *Management Science*, Vol.25, 1979, pp. 849-861
- [18] Neuts, M. F., *Matrix-Geometric Solutions in Stochastic Models : An Algorithmic Approach*, The Johns Hopkins University Press, 1981
- [19] Nicola, V. F., V. G. Kulkarni and K.S. Trivedi, "Queueing Analysis of Fault-Tolerant Computer Systems," Technical Report, CS-1985-24, Duke University
- [20] Osaki, S. and T. Nishio, "Reliability Evaluation of Some Fault-Tolerant Computer Architectures," *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 1980
- [21] Osaki, S., "Performance/Reliability Measures for Fault-Tolerant Computing Systems," *IEEE Trans. on Reliability*, Vol. R-33, No. 4, 1984, pp. 268-271

- [22] Perros, H. G., "On the $M/C_k/1$ Queue," *Performance Evaluation* 3, 1983, pp. 83-93
- [23] Siewiorek, D. P. and R. S. Swarz, *The Theory and Practice of Reliable System Design*, Digital Press, Bedford, MA, 1982
- [24] Trivedi. S. K., *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*, Prentice-Hall, Englewoods Cliffs, NJ, 1982