Limiting Average Availability of a System
Supported by Several Spares and Several Repair Facilities

by

Jyotirmoy Sarkar

PR # 01-10

This manuscript and other can be obtained via the
World Wide Web from www.math.iupui.edu

December 2001

Department of Mathematical Sciences
Indiana University Purdue University Indianapolis
LIMITING AVERAGE AVAILABILITY OF A SYSTEM
SUPPORTED BY SEVERAL SPARES AND SEVERAL REPAIR
FACILITIES

Jyotirmoy Sarkar
Department of Mathematical Sciences
Indiana University Purdue University Indianapolis
402 N. Blackford Street
Indianapolis, IN 46202-3216 USA
jsarkar@math.iupui.edu

December 19, 2001

Abstract: A one-unit system, supported by $s$ spare units and $r$ repair facilities, fails when all units are either undergoing or awaiting repair. Under the perfect repair policy, the limiting average availability is obtained when both life and repair times are exponential.

Keywords: Continuous monitoring, Perfect repair, Stand-by, Mean system up time, Mean system down time, Transition probability, Sojourn time.

1991 AMS Subject Classification: Primary 90B25; Secondary 62N05.
1 Introduction

We consider a one-unit system under continuous monitoring. Initially, one unit is put on operation and \(s\) identical spares remain on cold stand-by. As soon as the operating unit fails one of the spare units, if available, is placed on service while the failed unit joins the repair queue. There are \(r\) repair facilities. Repair starts on a failed unit as soon as one of the repair facilities becomes free. We assume that repair takes a random amount of time, after which the unit is restored back to a level equivalent to a new unit (this is called the perfect repair policy) and becomes a viable spare. A spare unit cannot fail, hence is said to remain on cold stand-by until placed on service. Since there are only a finite number of spares, a replacement of the failed operating unit may not be always possible. The system fails when the operating unit fails, and no spare unit is available to replace it, because all units are either undergoing or awaiting repair. Specifically, this implies that \(r \leq s + 1\), since otherwise some repair facilities will always remain idle.

A more general model requiring \(q\) operating units to run the system has been described in Barlow and Proschan (1975, page 204). Of interest is the limiting average availability (the limiting expected proportion of time the system is up), which is defined as

\[
A_{av} := \frac{\text{MSUT}}{\text{MSUT} + \text{MSDT}}
\]

where MSdT stands for mean system down time and is the mean duration from the moment the system fails until it is revived through repair, and MSUT stands for mean system up time and is the mean duration from the epoch when the failed system is revived to the next system failure. Barlow and Proschan (1975, page 206) provide the limiting average availability only for the case of one repair facility and either no or one spare unit, assuming exponential lifetime and exponential repair time distributions. However, their expression (3.6) for the limiting average availability of a one-unit system supported by one spare and one repair facility, when lifetime and repair time distributions are arbitrary, is incorrect as pointed out by Sen and Bhattacharya (1986, page 283) who gave the correct expression.
A more ambitious objective is to obtain an expression for the instantaneous availability function $A(t)$, the probability that the system is up at a specified time $t > 0$. $A(t)$ measures the performance of a maintained system and is an important aspect of reliability theory. For an excellent account on the subject see, for example, Høyland and Rausand (1994). Explicit evaluation of exact availability is often very difficult. For some examples see Sarkar and Chaudhuri (1999). No attempt is made in this article to evaluate $A(t)$ in our model.

However, in many practical situations, the limiting average availability is sufficient for decision making. In this paper, therefore, we focus on the limiting average availability of a one-unit system supported by $s$ spares and $r$ repair facilities, when the lifetime of each unit is exponential($\alpha$) and the repair time is exponential ($\beta$). For this model, the MSDT is simply $1/(r\beta)$. Obtaining the expression for the MSUT is a lot more complicated even for the case of exponential lifetime distribution of the operating unit. Specifically, it requires solving a system of $s + 1$ linear equations. This precisely is the main contribution of this article. Some preliminaries are given in Section 2. Section 3 gives the main result, while Section 4 presents some applications.

## 2 Preliminaries

Assume that $X_0, X_1, X_2, \ldots$ are independent and identically distributed (IID) random variables representing the times to failure of operating units with exponential($\alpha$) distribution; and $Y_1, Y_2, \ldots$ are IID random variables representing the repair times with exponential($\beta$) distribution. Assume also that the times to failure are independent of the repair times.

We say the system is in state $i$ ($i = 0, 1, \ldots, s, s + 1$) if there are $i$ failed units undergoing or awaiting repair. At time $t = 0$, the system is in state 0. When the system enters state $s + 1$, it fails. From any state $i$ the system transitions to an adjacent state after a random amount of time. Using the lack of memory property of the exponential distribution, it is straightforward to see that the sojourn time in state $i$ is exponentially distributed with scale parameter $a_i \alpha + b_i \beta$ where $a_i$ is
the number of operating unit (that is, \( a_i = 1 \) if \( 0 \leq i \leq s \) and \( a_{s+1} = 0 \)) and \( b_i \) is the number of unit(s) undergoing repair (that is, \( b_i = \min\{i, r\} \)). Furthermore, the transition probabilities \( P_{ij} \) from state \( i \) to state \( j \) \((i, j = 0, 1, \ldots, s + 1)\) are given by

\[
\begin{align*}
P_{i,i+1} &= \frac{a_i \alpha}{a_i \alpha + b_i \beta} \\
P_{i,i-1} &= \frac{b_i \beta}{a_i \alpha + b_i \beta} \\
P_{i,j} &= 0 \text{ for } |i - j| \neq 1.
\end{align*}
\]

(1)

Note that states 0 and \( s + 1 \) are reflecting states (that is, they transition to states 1 and \( s \) respectively, with probability one).

Let \( E_i \) denote the mean time to transition from state \( i \) to state \( i + 1 \) for \( i = 0, 1, \ldots, s \); and let \( E_{s+1} = 0 \). Then for \( i = 0, 1, \ldots, s \) we have

\[
E_i = \frac{1}{a_i \alpha + b_i \beta} + \frac{b_i \beta}{a_i \alpha + b_i \beta} (E_{i-1} + E_i).
\]

(2)

The first term on the right hand side of (2.2) is the expected duration of time the system remains in state \( i \), and the second term is the product of \( P_{i,i-1} \) and the mean time to transition from state \( i - 1 \) to state \( i + 1 \) (of course, via state \( i \)). After simplification (2.2) yields

\[
E_i = \frac{1}{a_i \alpha} + \frac{b_i \beta}{a_i \alpha} E_{i-1}.
\]

(3)

Note that the failed system (in state \( s+1 \)), when revived, enters state \( s \) in which only one unit is operating and all other units are undergoing or awaiting repair (hence, there is no viable spare). The revived system remains operational until it enters state \( s + 1 \), that is, until the next system failure. Therefore, MSUT\( = E_s \).

Theorem 3.1 below gives the expression for \( E_s \) by solving the system of equations (2.3). The expression for MSDT is rather straight-forward. Note that when the system is down (in state \( s+1 \)) all \( r \leq s+1 \) repair facilities are busy, each having an exponential(\( \beta \)) repair time distribution. Therefore, by the independence of the repair times, the system down time has an exponential(\( r \beta \)) distribution. Hence,

\[
\text{MSDT} = (r \beta)^{-1}.
\]

(4)
3 Evaluation of MSUT

Theorem 3.1: For a one-unit system supported by \( r \geq 1 \) repair facilities and \( s \geq r - 1 \) spare units, with lifetime distribution exponential(\( \alpha \)) and repair time distribution exponential(\( \beta \)), the (long-run) mean system up time is given by

\[
MSUT = E_s = \alpha^{-1} \sum_{i=0}^{s} \gamma_i \rho^{s-i}
\]

(1)

where \( \rho = \beta/\alpha \) and

\[
\gamma_i = \begin{cases} 
  r^{s-r} r!/i! & \text{if } i = 0, 1, \ldots, r - 1 \\
  r^{s-i} & \text{if } i = r, \ldots, s.
\end{cases}
\]

(2)

Hence, the limiting average availability is given by

\[
A = \frac{r \rho \sum_{i=0}^{s} \gamma_i \rho^{s-i}}{1 + r \rho \sum_{i=0}^{s} \gamma_i \rho^{s-i}}.
\]

(3)

Proof: Writing (2.3) as \( a_i E_i - b_i E_{i-1} = 1 \) \((i = 0, 1, \ldots, s)\) and substituting the values of \( a_i \) and \( b_i \), we have

\[
\begin{pmatrix}
  E_0 \\
  E_1 \\
  E_2 \\
  \vdots \\
  E_r \\
  E_{r+1} \\
  E_s
\end{pmatrix} = \begin{pmatrix}
  \alpha & 0 & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\
  -\beta & \alpha & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\
  0 & -2\beta & \alpha & 0 & \ldots & \ldots & \ldots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
  0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
  0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
  0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \alpha \\
  \end{pmatrix}^{-1} \begin{pmatrix}
  1 \\
  1 \\
  1 \\
  \vdots \\
  1 \\
  \vdots \\
  1 \\
  \end{pmatrix}
\]

(4)

Denote the inverse matrix on the right hand side of (3.4) by \( D^{-1} = ((d^{si})) \). Then \( E_s = (0, 0, \ldots, 0, 1) D^{-1} (1, 1, \ldots, 1)^T = \sum_{i=0}^{s} d^{si} \). Note that

\[
d^{si} = (-1)^{s+i} \frac{\det(D_{is})}{\det(D)},
\]

(5)
with

$$\det(D) = \alpha^{s+1}$$

(6)

and the \((i, s)\)-th cofactor

$$\det(D_{is}) = \begin{cases} \alpha^i \left[-(i+1)\beta \right] \cdots \left[-r\beta \right] \times \left[-r\beta \right]^{s-r} & \text{if } i = 0, 1, \ldots, r - 1 \\ \alpha^i \left[-r\beta \right]^{s-i} & \text{if } i = r, \ldots, s \end{cases}$$

$$= \gamma_i \alpha^i (-\beta)^{s-i}. \quad (7)$$

Using (3.5)-(3.7), we have

$$E_s = \frac{1}{\det(D)} \sum_{i=0}^{s} (-1)^{i+s} \det(D_{is}) = \alpha^{-(s+1)} \sum_{i=0}^{s} \gamma_i \alpha^i \beta^{s-i} = \alpha^{-1} \sum_{i=0}^{s} \gamma_i \rho^{s-i}$$

proving (3.1). Finally, (3.3) follows from (1.1), (2.4) and (3.1). \(\square\)

In particular, when there is only one repair facility (that is, \(r = 1\)), we have \(\gamma_i = 1\) for \(i = 0, \ldots, s\). Hence, \(E_s = \alpha^{-1}(\rho^{s+1} - 1)/(\rho - 1)\) and

$$A_{av} = \frac{\rho(\rho^{s+1} - 1)}{\rho^{s+2} - 1} = \frac{\beta(\beta^{s+1} - \alpha^{s+1})}{\beta^{s+2} - \alpha^{s+2}}. \quad (8)$$

Likewise, when there are two repair facilities (that is, \(r = 2\)), we have \(\gamma_0 = 2^{s-1}\) and \(\gamma_i = 2^{s-i}\) for \(i = 1, \ldots, s\). Hence, \(E_s = (2\alpha)^{-1}(2\rho^s - 2)(2\rho + 1)/(2\rho - 1)\) and

$$A_{av} = \frac{\rho(2\rho^s - 2)(2\rho + 1)}{\rho(2\rho^s - 2)(2\rho + 1) + (2\rho - 1)}. \quad (9)$$

### 4 Applications

Consider a slightly more general model which requires \(q \geq 1\) operating units which have independent and identical exponential(\(\alpha'\)) life distributions. Suppose the system operation is suspended whenever there are fewer than \(q\) viable units, so that none of these units can fail during suspension of operation. Then the expressions for MSUT and the limiting average availability are obtained by simply putting \(\alpha = q\alpha'\) in Theorems 3.1 and 4.1. This is because the system lifetime distribution is exponential(\(q\alpha\)) in this model.
Example 4.1. A manned spacecraft requires four generators in order to maintain course; otherwise the spacecraft must return to earth. Assume each generator has exponential lifetime with mean 1000 hours and repair takes exponential time with mean 25 hours. It is desired to ensure that at least four generators are functioning with a (long-run) probability of $1 - 10^{-6}$. How many spare generators must be carried on the spacecraft if (a) there is only one repair person, (b) there are two repair persons who work independently but at a time only one can work on a failed unit.

Here, $\alpha = .004, \beta = .04$. Hence, $E_s(r = 1) = 250(10^{s+1} - 1)/9$, MSDT($r = 1$) = 25 and $E_s(r = 2) = 125(21 \times 20^s - 2)/19$, MSDT($r = 2$) = 12.5. Table 1 below lists the limiting average availability for various values of $s$. Thus, if there is only one repair person, the spacecraft should carry five spare generators. If there are two repair persons, four spare generators will suffice.

Table 1 about here

Example 4.2. Metro Railway operates underground trains which run on an 80 minute cycle. Metro Railway strives to make trains available at ten minute interval on an average. So they keep eight trains on operation and two on reserve. It is estimated that each train runs on an average 400 days before breakdown. There is only one repair facility which takes an average of 25 days to complete repair. In order to improve the availability of trains at ten minute intervals, should Metro Railway (a) buy one more train to remain on reserve or (b) set up another repair facility? Assume the cost is about the same for each alternative.

Assume the lifetime is exponential(1/400) and the repair time is exponential(1/25). Then $\alpha = .02, \beta = .04$, $E_s(r = 1) = 25(2^{s+1} - 1)$, MSDT($r = 1$) = 25, MSDT($r = 2$) = 12.5 and $E_s(r = 2) = 25(5 \times 4^s - 2)/3$. Since $A_{av}(s = 3, r = 1) = 0.9677 < A_{av}(s = 2, r = 2) = 0.9811$, option (b) is preferable to option (a).

Example 4.3. At a certain customer service agency phone calls arrive following a Poisson Process with rate 1/2 per minute. That is, the inter-arrival time distr-
bution is exponential($\alpha = 1/2$). The agency has $1+s$ telephone attendants. Unless all attendants are busy, an incoming call is immediately answered by one of the free attendants. Each phone call can last for a random amount of time described by exponential($\beta = 1/5$) distribution. We assume that customers hang up if their calls are not answered immediately.

In this example, the attendants are the operating units. An attendant scheduled to pick up the next incoming call is in the operating state. Other free attendants are on stand by. When an attendant is talking to a customer, she is not available to pick up an incoming call. Hence, conversation is identified with the unit being down and undergoing repair. The system is up when at least one attendant is free and it is down when all attendants are busy. Since each busy attendant can become free after talking to her customer, we have $r = 1+s$.

Of interest to the agency is the determination of the number of telephone attendants so that an acceptable proportion of customers’ calls are answered immediately. This proportion is given by the limiting average availability of the system. Another important quantity of interest to the agency (say, for determining the attendant’s wage) is the proportion $\theta$ of time an attendant is busy answering calls. To obtain this proportion of busy time we reason as follows: During a long period of $N$ units of time, roughly there were $\alpha N$ calls of which $A_{av} \alpha N$ were answered, requiring a total of roughly $A_{av} \alpha N/\beta$ units of busy time, which was shared by $1+s$ attendants. Hence, each attendant has been busy for roughly $A_{av} \alpha N/[\beta(1+s)]$ units of time out of the $N$ units she was on duty. Hence, the proportion of busy time is $\theta = A_{av} \alpha/[\beta(1+s)]$. Table 2 gives the limiting average availability and proportion of busy time for any attendant various values of $s = 0, 1, 2, \ldots$.

![Table 2 about here](image)

**Acknowledgment.** This research was conducted, in part, while the author was visiting the Indian Statistical Institute Calcutta. The author wishes to thank the Institute authorities for their kind hospitality and Dr. S. Bose for some discussions.
REFERENCES


TABLE 1: Limiting average availability in Example 4.1 (a) and (b)

<table>
<thead>
<tr>
<th>$s$</th>
<th>One Repair Facility $E_s$ $A_{av}$</th>
<th>Two Repair Facilities $E_s$ $A_{av}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250 .9090909091</td>
<td>* *</td>
</tr>
<tr>
<td>1</td>
<td>2750 .9909909910</td>
<td>2750 .9954751131</td>
</tr>
<tr>
<td>2</td>
<td>27750 .9990999100</td>
<td>55250 .9997738068</td>
</tr>
<tr>
<td>3</td>
<td>277750 .9999099991</td>
<td>1105250 .9999886905</td>
</tr>
<tr>
<td>4</td>
<td>2777750 .9999910000</td>
<td>22105250 .9999994345</td>
</tr>
<tr>
<td>5</td>
<td>27777750 .99999910000</td>
<td>442105250 .999999717</td>
</tr>
</tbody>
</table>

* If there is no spare unit, the second repair facility is always idle.
TABLE 2: Limiting average availability in Example 4.3 with $\alpha = 1/2, \beta = 1/5$

<table>
<thead>
<tr>
<th>$s$</th>
<th>Formula for $E_s$</th>
<th>MSUT</th>
<th>MSDT</th>
<th>$A_{av}$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\alpha^{-1}$</td>
<td>2.000</td>
<td>5.000</td>
<td>.285714</td>
<td>.714286</td>
</tr>
<tr>
<td>1</td>
<td>$\alpha^{2}(\alpha + \beta)$</td>
<td>2.800</td>
<td>2.500</td>
<td>.528302</td>
<td>.660377</td>
</tr>
<tr>
<td>2</td>
<td>$\alpha^{-3}(\alpha^2 + 2\alpha\beta + 2\beta^2)$</td>
<td>4.240</td>
<td>1.667</td>
<td>.717833</td>
<td>.598194</td>
</tr>
<tr>
<td>3</td>
<td>$\alpha^{-4}(\alpha^3 + 3\alpha^2\beta + 6\alpha\beta^2 + 6\beta^3)$</td>
<td>7.088</td>
<td>1.250</td>
<td>.850084</td>
<td>.531302</td>
</tr>
<tr>
<td>4</td>
<td>$\alpha^{-5}(\alpha^4 + 4\alpha^3\beta + 12\alpha^2\beta^2 + 24\alpha\beta^3 + 24\beta^4)$</td>
<td>8.707</td>
<td>1.000</td>
<td>.896984</td>
<td>.448492</td>
</tr>
</tbody>
</table>