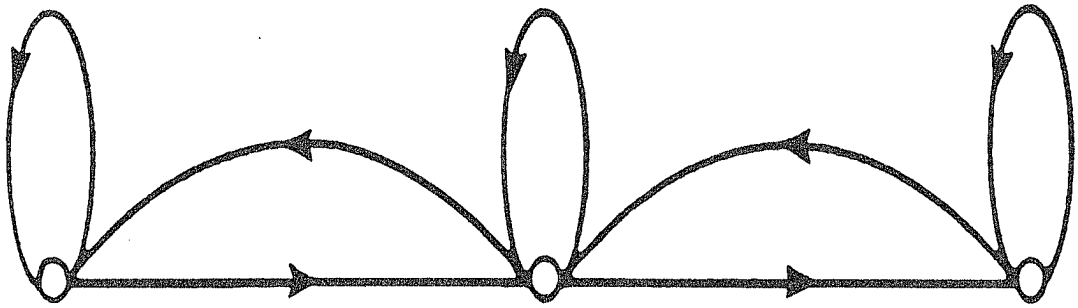


RELIABILITY OF COMPUTER SYSTEMS AND NETWORKS

Fault Tolerance, Analysis, and Design

MARTIN L. SHOOMAN

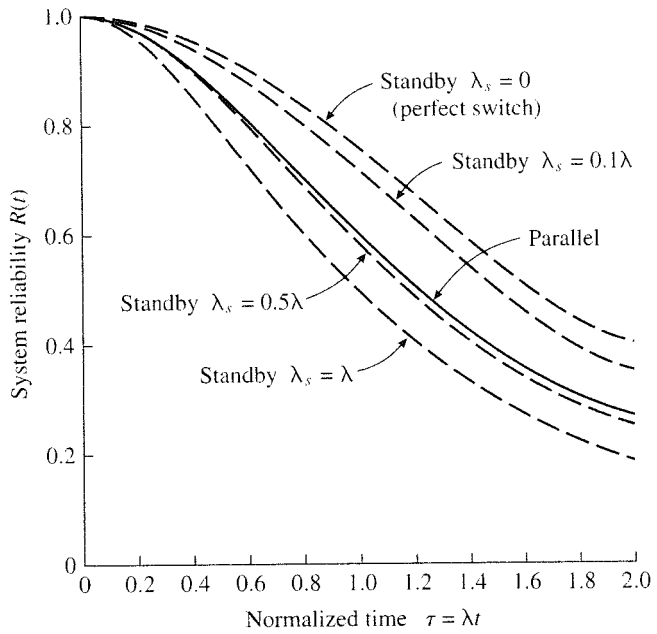


Figure 3.13 A comparison of a two-element ordinary parallel system with a two-element standby system with imperfect switch reliability.

$$R_2(t) = e^{-\lambda t} + \lambda t e^{-\lambda t} e^{-\lambda_s t} \tag{3.60}$$

Clearly, this is less conservative and a more realistic switch model than the previous one.

One can construct even more complex failure models for the switch in a standby system [Shooman, 1990, Section 6.9].

1. Switch failure modes where the switching occurs even when the on-line element is good or where the switch jitters between elements can be included.
2. The failure rate of n nonidentical standby elements was first derived by Bazovsky [1961, p. 117]; this can be shown as related to the gamma distribution and to approach the normal distribution for large n [Shooman, 1990].
3. For n identical standby elements, the system succeeds if there are n–1 or fewer failures, and the probabilities are given by the Poisson distribution that leads to the expression

$$R(t) = e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} \tag{3.61}$$

3.8 REPAIRABLE SYSTEMS

3.8.1 Introduction

Repair or replacement can be viewed as the same process, that is, replacement of a failed component with a spare is just a fast repair. A complete description of the repair process takes into account several steps: (a) detection that a failure has occurred; (b) diagnosis or localization of the cause of the failure; (c) the delay for replacement or repair, which includes the logistic delay in waiting for a replacement component or part to arrive; and (d) test and/or recalibration of the system. In this section, we concentrate on modeling the basics of repair and will not decompose the repair process into a finer model that details all of these substates.

The decomposition of a repair process into substates results in a nonconstant-repair rate (see Shooman [1990, pp. 348–350]). In fact, there is evidence that some repair processes lead to lognormal repair distributions or other nonconstant-repair distributions. One can show that a number of distributions (e.g., lognormal, Weibull, gamma, Erlang) can be used to model a repair process [Muth, 1967, Chapter 3]. Some software for modeling system availability permits nonconstant-failure and -repair rates. Only in special cases is such detailed data available, and constant-repair rates are commonly used. In fact, it is not clear how much difference there is in compiling the steady-state availability for constant- and nonconstant-repair rates [Shooman, 1990, Eq. (6.106) ff.]. For a general discussion of repair modeling, see Ascher [1984].

In general, repair improves two different measures of system performance: the reliability and the availability. We begin our discussion by considering a single computer and the following two different types of computer systems: an air traffic control system and a file server that provides electronic mail and network access to a group of users. Since there is only a single system, a failure of the computer represents a system failure, and repair will not affect the system *reliability* function. The *availability* of the system is a measure of how much of the operating time the system is up. In the case of the air traffic control system, the fact that the system may occasionally be down for short time periods while repair or replacement goes on may not be tolerable, whereas in the case of the file server, a small amount of downtime may be acceptable. Thus a computation of both the reliability and the availability of the system is required; however, for some critical applications, the most important measure is the reliability. If we say the basic system is composed of two computers in parallel or standby, then the problem changes. In either case, the system can tolerate one computer failure and stay up. It then becomes a race to see if the

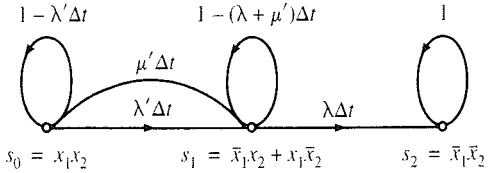
failed element can be repaired and restored before the remaining element fails. The system only goes down in the rare event that the second component fails before the repair or replacement is completed.

In the following sections, we will model a two-element parallel and a two-element standby system with repair and will comment on the improvements in reliability and availability due to repair. To facilitate the solutions of the ensuing Markov models, some simple features of the Laplace transform method will be employed. It is assumed that the reader is familiar with Laplace transforms or will have already read the brief introduction to Laplace transform methods given in Appendix B, Section B8. We begin our discussion by developing a general Markov model for two elements with repair.

3.8.2 Reliability of a Two-Element System with Repair

The benefits of repair in improving system reliability are easy to illustrate in a two-element system, which is the simplest system used in high-reliability fault-tolerant situations. Repair improves both a hot standby and a cold standby system. In fact, we can use the same Markov model to describe both situations if we appropriately modify the transition probabilities. A Markov model for two parallel or standby systems with repair is given in Fig. 3.14. The transition rate from state s_0 to s_1 is given by 2λ in the case of an ordinary parallel system because two elements are operating and either one can fail. In the case of a standby system, the transition is given by λ since only one component is powered and only that one can fail (for this model, we ignore the possibility that the standby system can fail). The transition rate from state s_1 to s_0 represents the repair process. If only one repairman is present (the usual case), then this transition is governed by the constant repair rate μ. In a rare case, more than one repairman will be present, and if all work cooperatively, the repair rate is $>\mu$. In some circumstances, there will be only a shared repairman among a number of equipments, in which case the repair rate is $<\mu$.

In many cases, study of the repair statistics shows a nonexponential distribution (the exponential distribution is the one corresponding to a constant transition rate)—specifically, the lognormal distribution [Ascher, 1984; Shooman, 1990, pp. 348–350]. However, much of the benefits of repair are illustrated by



where $\lambda' = 2\lambda$ for an ordinary system
$\lambda' = \lambda$ for a standby system
$\mu' = \mu$ for one repairman
$\mu' = k\mu$ for more than one repairman ($k > 1$)

Figure 3.14 A Markov reliability model for two identical parallel elements and k repairmen.

the constant transition rate repair model. The Markov equations corresponding to Fig. 3.14 can be written by utilizing a simple algorithm:

1. The terms with 1 and Δt in the Markov graph are deleted.
2. A first-order Markov differential equation is written for each node where the left-hand side of the equation is the first-order time derivative of the probability of being in that state at time t.
3. The right-hand side of each equation is a sum of probability terms for each branch that enters the node in question. The coefficient of each probability term is the transition probability for the entering branch.

We will illustrate the use of these steps in formulating the Markov of Fig. 3.14.

$$\frac{dP_{s_0}(t)}{dt} = -\lambda' P_{s_0}(t) + \mu' P_{s_1}(t) \tag{3.62a}$$

$$\frac{dP_{s_1}(t)}{dt} = \lambda' P_{s_0}(t) - (\lambda + \mu') P_{s_1}(t) \tag{3.62b}$$

$$\frac{dP_{s_2}(t)}{dt} = \lambda' P_{s_1}(t) \tag{3.62c}$$

Assuming that both systems are initially good, the initial conditions are

$$P_{s_0}(0) = 1, \qquad P_{s_1}(0) = P_{s_2}(0) = 0$$

One great advantage of the Laplace transform method is that it deals simply with initial conditions. Another is that it transforms differential equations in the time domain into a set of algebraic equations in the Laplace transform domain (often called the frequency domain), which are written in terms of the Laplace operator s.

To transform the set of equations (3.62a–c) into the Laplace domain, we utilize transform theorem 2 (which incorporates initial conditions) from Table B7 of Appendix B, yielding

$$sP_{s_0}(s) - 1 = -\lambda' P_{s_0}(s) + \mu' P_{s_1}(s) \tag{3.63a}$$

$$sP_{s_1}(s) - 0 = \lambda' P_{s_0}(s) - (\lambda + \mu') P_{s_1}(s) \tag{3.63b}$$

$$sP_{s_2}(s) - 0 = \lambda P_{s_1}(s) \tag{3.63c}$$

Writing these equations in a more symmetric form yields

$$(s+\lambda')P_{s_0}(s) - \mu' P_{s_1}(s) = 1 \tag{3.64a}$$

$$-\lambda' P_{s_0}(s) + (s+\mu'+\lambda)P_{s_1}(s) = 0 \tag{3.64b}$$

$$-\lambda P_{s_1}(s) + sP_{s_2}(s) = 0 \tag{3.64c}$$

Clearly, Eqs. (3.64a–c) lead to a matrix formulation if desired. However, we can simply solve these equations using Cramer's rule since they are now algebraic equations.

$$P_{s_0}(s) = \frac{(s+\lambda+\mu')}{[s^2 + (\lambda+\lambda'+\mu')s + \lambda\lambda']} \tag{3.65a}$$

$$P_{s_1}(s) = \frac{\lambda'}{[s^2 + (\lambda+\lambda'+\mu')s + \lambda\lambda']} \tag{3.65b}$$

$$P_{s_2}(s) = \frac{\lambda\lambda'}{s[s^2 + (\lambda+\lambda'+\mu')s + \lambda\lambda']} \tag{3.65c}$$

We must now invert these equations—transform them from the frequency domain to the time domain—to find the desired time solutions. There are several alternatives at this point. One can apply transform No. 10 from Table B6 of Appendix B to Eqs. (3.65a, b) to obtain the solution as a sum of two exponentials, or one can use a partial fraction expansion as illustrated in Eq. (B104) of the appendix. An algebraic solution of these equations using partial fractions appears in Shooman [1990, pp. 341–342], and further solution and plotting of these equations is covered in the problems at the end of this chapter as well as in Appendix B8. One can, however, make a simple comparison of the effects of repair by computing the MTTF for the various models.

3.8.3 MTTF for Various Systems with Repair

Rather than compute the complete reliabiity function of the several systems we wish to compare, we can simplify the analysis by comparing the MTTF for these systems. Furthermore, the MTTF is given by an integral of the reliability function, and by using Laplace theory we can show [Section B8.2, Eqs. (B105)–(B106)] that the MTTF is just given by the limit of the Laplace transform expression as $s \to 0$.

For the model of Fig. 3.14, the reliability expression is the sum of the first two-state probabilities; thus, the MTTF is the limit of the sum of Eqs. (3.65a, b) as $s \to 0$, which yields

$$\text{MTTF} = \frac{\lambda + \mu' + \lambda'}{(\lambda\lambda')} \tag{3.66}$$

TABLE 3.4 Comparison of MTTF for Several Systems

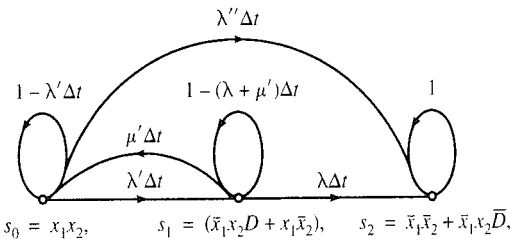
Element	Formula	For $\lambda = 1$, $\mu = 10$
Single element	$1/\lambda$	1.0
Two parallel elements—no repair	$1.5/\lambda$	1.5
Two standby elements—no repair	$2/\lambda$	2.0
Two parallel elements—with repair	$(3\lambda+\mu)/2\lambda^2$	6.5
Two standby elements—with repair	$(2\lambda+\mu)/\lambda^2$	12.0

We substitute the various values of λ' shown in Fig. 3.14 in the expression; since we are assuming a single repairman, $\mu' = \mu$. The MTTF for several systems is compared in Table 3.4. Note how repair strongly increases the MTTF of the last two systems in the table. For large μ/λ ratios, which are common in practice, the MTTF of the last two systems approaches $0.5\mu/\lambda^2$ and μ/λ^2.

3.8.4 The Effect of Coverage on System Reliability

In Fig. 3.12, we portrayed a fairly complex block diagram for a standby system. We have already modeled the possibility of imperfection in the switching mechanism. In this section, we develop a model for imperfections in the decision unit that detects failures and switches from the on-line system to the standby system. In some cases, even in the n-ordinary parallel system (hot standby), it is not possible to have both systems fully connected, and a decision unit and switch are needed. Another way of describing this phenomenon is to say that the decision unit cannot detect 100% of all the on-line unit failures; it only "covers" (detects) the fraction c ($0 < c < 1$) of all the possible failures. (The formulation of this concept is generally attributed to Bouricius, Carter, and Schneider [1969].) The problem is that if the decision unit does not detect a failure of the on-line unit, input and output remain connected to the failed on-line element. The result is a system failure, because although the standby unit is good, there is no indication that it must be switched into use. We can formulate a Markov model in Fig. 3.15, which allows us to evaluate the effect of coverage. (Compare with the model of Fig. 3.14.) In fact, we can use Fig. 3.15 to model the effects of coverage on either a hot or cold standby system. Note that the symbol D stands for the decision unit correctly detecting a failure in the on-line unit, and the symbol $\overline{D}$ means that the decision unit has not been able to (failed to) detect a failure in the on-line unit. Also, a new arc has been added in the figure from the good state s_0 to the failed state s_2 for modeling the failure of the decision unit to "cover" a failure of the on-line element.

The Markov equations for Fig. 3.15 become the following:



where $\lambda' = 2c\lambda$ for an ordinary parallel system
$\lambda'' = 2(1-c)\lambda$ for an ordinary parallel system
$\lambda' = c\lambda$ for a standby system
$\lambda'' = (1-c)\lambda$ for a standby system
$\mu' = \mu$ for one repairman

Figure 3.15 A Markov reliability model for two identical, parallel elements, k repairmen, and coverage effects.

$$sP_{s_0}(s) - 1 = -(\lambda' + \lambda'')P_{s_0}(s) + \mu' P_{s_1}(s) \tag{3.67a}$$
$$sP_{s_1}(s) - 0 = \lambda' P_{s_0}(s) - (\lambda + \mu')P_{s_1}(s) \tag{3.67b}$$
$$sP_{s_2}(s) - 0 = \lambda'' P_{s_0}(s) + \lambda P_{s_1}(s) \tag{3.67c}$$

Compare the preceding equations with Eqs. (3.63a–c) and (3.64a–c). Writing these equations in a more symmetric form yields

$$(s + \lambda' + \lambda'')P_{s_0}(s) - \mu' P_{s_1}(s) = 1 \tag{3.68a}$$
$$-\lambda' P_{s_0}(s) + (s + \mu' + \lambda)P_{s_1}(s) = 0 \tag{3.68b}$$
$$-\lambda'' P_{s_0}(s) - \lambda P_{s_1}(s) + sP_{s_2}(s) = 0 \tag{3.68c}$$

The solution of these equations yields

$$P_{s_0}(s) = \frac{(s + \lambda + \mu')}{s^2 + (\lambda + \lambda' + \lambda'' + \mu')s + (\lambda\lambda' + \lambda''\mu' + \lambda\lambda'')} \tag{3.69a}$$

$$P_{s_1}(s) = \frac{\lambda'}{s^2 + (\lambda + \lambda' + \lambda'' + \mu')s + (\lambda\lambda' + \lambda''\mu' + \lambda\lambda'')} \tag{3.69b}$$

$$P_{s_2}(s) = \frac{\lambda''s + \lambda\lambda' + \mu'\lambda'' + \lambda\lambda''}{s[s^2 + (\lambda + \lambda' + \lambda'' + \mu')s + (\lambda\lambda' + \lambda''\mu' + \lambda\lambda'')]} \tag{3.69c}$$

For the model of Fig. 3.15, the reliability expression is the sum of the first two-state probabilities; thus the MTTF is the limit of the sum of Eqs. (3.69a, b) as $s \to 0$, which yields

TABLE 3.5 Comparison of MTTF for Several Systems

Element	Formula	For $\lambda = 1$, $\mu = 10$, $c = 1$	For $\lambda = 1$ $\mu = 10$, $c = 0.95$	For $\lambda = 1$, $\mu = 10$, $c = 0.90$
Single element	$1/\lambda$	1.0	—	—
Two parallel elements—no repair: $[\mu' = 0, \lambda' = 2c\lambda, \lambda'' = 2(1-c)\lambda]$	$(0.5 + c)/\lambda$	1.5	1.45	1.40
Two standby elements—no repair: $[\mu' = 0, \lambda' = c\lambda, \lambda'' = (1-c)\lambda]$	$(1 + c)/\lambda$	2.0	1.95	1.90
Two parallel elements—with repair: $[\mu' = \mu, \lambda' = 2c\lambda, \lambda'' = 2(1-c)\lambda]$	$\frac{(1+2c)\lambda + \mu}{2\lambda[\lambda + (1-c)\mu]}$	6.5	4.3	3.2
Two standby elements—with repair: $[\mu' = \mu, \lambda' = c\lambda, \lambda'' = (1-c)\lambda]$	$\frac{(1+c)\lambda + \mu}{\lambda[\lambda + (1-c)\mu]}$	12.0	7.97	5.95

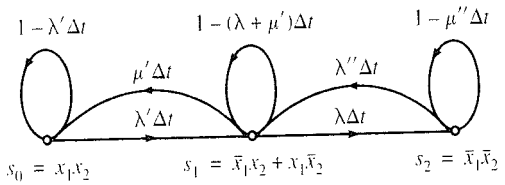
$$\text{MTTF} = \frac{\lambda + \mu' + \lambda'}{(\lambda\lambda' + \lambda''\mu' + \lambda\lambda'')} \tag{3.70}$$

When $c = 1$, $\lambda'' = 0$, and we see that Eq. (3.70) reduces to Eq. (3.66). The effect of coverage on the MTTF is evaluated in Table 3.5 by making appropriate substitutions for λ', λ'', and μ'. Notice what a strong effect the coverage factor has on the MTTF of the systems with repair. For two parallel and two standby systems, $c = 0.90$—more than half the MTTF. Practical values for c are hard to find in the literature and are dependent on design. Sieworek [1992, p. 288] comments, "a typical diagnostic program, for example, may detect only 80–90% of possible faults." Bell [1978, p. 91] states that static testing of PDP-11 computers at the end of the manufacturing process was able to find 95% of faults, such as solder shorts, open-circuit etch connections, dead components, and incorrectly valued resistors. Toy [1987, p. 20] states, "realistic coverages range between 95% and 99%." Clearly, the value of c should be a major concern in the design of repairable systems.

A more detailed treatment of coverage can be found in the literature. See Bouricius and Carter [1969, 1971]; Dugan [1989, 1996]; Kaufman and Johnson [1998]; and Pecht [1995].

3.8.5 Availability Models

In some systems, it is tolerable to have a small amount of downtime as the system is rapidly repaired and restored. In such a case, we allow repair out



where $\lambda' = 2\lambda$ for an ordinary system
$\lambda' = \lambda$ for a standby system
$\mu' = \mu$ for one repairman
$\mu' = k_1\mu$ for more than one repairman ($k_1 > 1$)
$\mu'' = \mu$ for one repairman
$\mu'' = 2\mu$ for two repairmen
$\mu'' = k_2\mu$ for more than one repairman ($k_2 > 1$)

Figure 3.16 Markov availability graph for two identical parallel elements.

of the system down state, and the model of Fig. 3.16 is obtained. Note that Fig. 3.14 and Fig. 3.16 only differ in the repair branch from state s_2 to state s_1. Using the same techniques that we used above, one can show that the equations for this model become

$$(s+\lambda')P_{s_0}(s) - \mu'P_{s_1}(s) = 1 \tag{3.71a}$$

$$-\lambda'P_{s_0}(s) + (s+\mu'+\lambda)P_{s_1}(s) - \mu''P_{s_2}(s) = 0 \tag{3.71b}$$

$$-\lambda P_{s_1}(s) + (s+\mu'')P_{s_2}(s) = 0 \tag{3.71c}$$

See Shooman [1990, Section 6.10] for more information.

The solution follows the same procedure as before. In this case, the sum of the probabilities for states 0 and 1 is not the reliability function but the availability function: $A(t)$. In most cases, $A(t)$ does not go to 0 as $t \to \infty$, as is true with the $R(t)$ function. $A(t)$ starts at 1 and, for well-designed systems, decays to a steady-state value close to 1. Thus a lower bound on the availability function is the steady-state value. A simple means for solving for the steady-state value is to formulate the differential equations for the Markov model and set all the time derivatives to 0. The set of equations now becomes an algebraic set of equations; however, the set is not independent. We obtain an independent set of equations by replacing any of these equations by the equation—the sum of all the state probabilities = 1. The algebraic solution for the steady-state availability is often used in practice. An even simpler procedure for computing the steady-state availability is to apply the final value theorem to the transformed expression for $A(s)$. This method is used in Section 4.9.2.

This chapter and Chapter 4 are linked in many ways. The technique of voting reliability joins parallel and standby system reliability as the three most common techniques for fault tolerance. Also, the analytical techniques involving Markov models are used in both chapters. In Chapter 4, a comparison is made of the reliability and availability of parallel, standby, and voting systems; in addition, some of the Markov modeling begun in this chapter is extended in Chapter 4 for the purpose of this comparison. The following chapter also has a more extensive discussion of the many shortcuts provided by Laplace transforms.

3.9 RAID SYSTEMS RELIABILITY

3.9.1 Introduction

The reliability techniques discussed in Chapter 2 involved coding to detect and correct errors in data streams. In this chapter, various parallel and standby techniques have been introduced that significantly increase the reliability of various systems and components. This section will discuss a newly developed technology for constructing computer secondary-storage systems that utilize the techniques of both Chapters 2 and 3 for the design of reliable, compact, high-performance storage systems. The generic term for such memory system technology is *redundant disk arrays* [Gibson, 1992]; however, it was soon changed to redundant array of inexpensive disks (RAID), and as technology evolved so that the quality and capacity of small disks rapidly increased, the word "inexpensive" was replaced by "independent." The term "array," when used in this context, means a collection of many disks organized in a specific fashion to improve speed of data transfer and reliability. As the RAID technology evolved, cache techniques (the use of small, very high-speed memories to accelerate processing by temporarily retaining items expected to be needed again soon) were added to the mix. Many varieties of RAID have been developed and more will probably emerge in the future. The RAID systems that employ cache techniques for speed improvement are sometimes called cached array of inexpensive disks (CAID) [Buzen, 1993]. The technology is driven by the variety of techniques available for connecting multiple disks, as well as various coding techniques, alternative read-and-write techniques, and the flexibility in organization to "tune" the architecture of the RAID system to match various user needs.

Prior to 1990, the dominant technology for secondary storage was a group of very large disks, typically 5–15, in a cabinet the size of a clothes washer. Buzen [1993] uses the term single large expensive disk (SLED) to refer to this technology. RAID technology utilizes a large number, typically 50–100, of small disks the size of those used in a typical personal computer. Each disk drive is assumed to have one actuator to position reads or writes, and large and small drives are assumed to have the same I/O read- or write-time. The bandwidth (BW) of such a disk is the reciprocal of the read-time. If data is broken into "chunks" and read (written) in parallel chunks to each of the n small disks in a RAID array, the effective BW increases. There is some "overhead" in implementing such a parallel read-write scheme, however, in the limit:

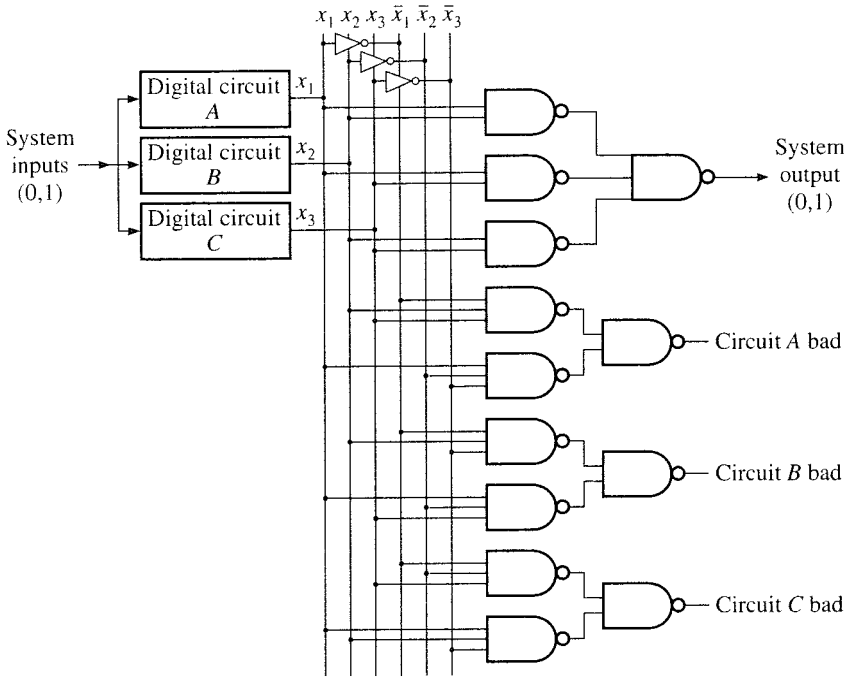


Figure 4.10 Circuit that realizes the four switching functions given in Table 4.5 for a TMR majority voter and error detector.

Fig. 4.10. The reader should realize that this circuit, with 13 NAND gates and 3 inverters, is only for a single bit output. For a 32-bit computer word, the circuit will have 96 inverters and 416 NAND gates. In Appendix B, Fig. B7, we show that the integrated circuit failure rate, λ, is *roughly* proportional to the square root of the number of gates, $\lambda \sim \sqrt{g}$, and for our example, $\lambda \sim \sqrt{512} = 22.6$. If we assume that the circuit on which we are voting should have 10 times the failure rate of the voter, the circuit would have 51,076 or about 50,000 gates. The implication of this computation is clear: One should not employ voters to improve the reliability of small circuits because the voter reliability may wipe out most of the intended improvement. Clearly, it would also be wise to consult an experienced logic circuit designer to see if the 512-gate circuit just discussed could be simplified by using other technology, semicustom gate circuits, available microelectronic chips, and so forth.

The circuit given in Fig. 4.10 could also be used to solve the chip test problem mentioned in Section 4.4.1. If the entire circuit of Fig. 4.10 were on a single IC, the outputs "circuit *A*, *B*, *C* bad" would allow initial testing and subsequent monitoring of the IC.

4.7 *N*-MODULAR REDUNDANCY WITH REPAIR

4.7.1 Introduction

In Chapter 3, we argued that as long as the operating system possesses redundancy, the addition of repair raises the reliability. One might ask at the outset why *N*-modular redundancy should be used with repair when ordinary parallel or standby redundancy with repair is very effective in achieving highly reliable and available systems. The answer to this question involves the coupling device reliability that was explored in Chapter 3. To be specific, suppose that we wish to compare the reliability of two parallel systems with that of a TMR system. Both systems fail if two of the elements fail, but in the TMR case, there are three systems that could fail; thus the probability of failure is higher. However, in general, the coupler in a parallel system will be more complex than a TMR voter, so a comparison of the two designs requires a detailed evaluation of coupler versus voter reliability. Analysis of TMR system reliability and availability can be found in Siewiorek [1992, p. 335] and in Toy [1987].

4.7.2 Reliability Computations

One might expect that it would be most efficient to seek a general solution for the reliability and availability of a system with *N*-modular redundancy and repair, then specify that $N = 3$ for a TMR system, $N = 5$ for 5-level voting, and so on. A moment's thought, however, suggests quite a different approach. The conventional solution for the reliability and availability of a system with repair involves making a Markov model and solving it much as was done in Chapter 3. In the process, the Laplace transform was computed, and a partial fraction expansion was used to find the individual exponential terms in the solution. For the case of repair, in general the repair rates couple the n states, and solution of the set of n first-order differential equations leads to the solution of an nth-order differential equation. If one applies Laplace transform theory, solution of the nth-order differential equation is "transformed into" a simpler sequence of steps. However, one step involves the solution for the roots of an nth-order polynomial.

Unfortunately, *closed-form* solutions exist only for first- through fourth-order polynomials, and solution procedures for cubic and quadratic polynomials are lengthy and seldom used. We learned in high-school algebra the formula for the roots of a quadratic equation (polynomial). A somewhat more complex solution exists for the solution of a cubic, which is listed in various handbooks [Iyanaga, p. 1396], and also for a fourth-order equation [Iyanaga, p. 1396].

A brief historical note about the origin of closed-form solutions is of interest. The formula for the third-order equation is generally attributed to Giordamo Cardano (also known as Jerome Cardan) [Cardano, 1545; Cardan, 1963]; however, he obtained the solution from Nicolo Tartaglia, and apparently it was discovered by Scipio Ferreo in circa 1505 [Hall, 1957, pp. 480–481]. Ludovico Ferrari, a pupil of Cardan, developed the formula for the fourth-order equation.

Neils Henrik Abel developed a proof that no closed-form solution exists for $n \geq 5$ [Iyanaga, p. 1].

The conclusion from the foregoing information on polynomial roots is that we should start with TMR and other simpler systems if we wish to use algebraic solutions. Numerical solutions are always possible for higher-order equations, and the mathematical software discussed in Appendix D expedites such an approach; however, the insight of an analytical solution is generally lacking. Another approach is to use simplifications and approximations such as those discussed in Appendix B (Sections B8.2 and B8.3). We will use the tried and true three-step engineering approach:

1. Represent the main features of the system by a low-order model that is amenable to closed-form solution.
2. Add further effects one at a time that complicate the model; study the effect (if necessary, use simplifying assumptions and approximations or numerical results computed over a range of parameters).
3. Put all the effects into a comprehensive model and solve numerically.

Our development begins by studying the reliability and availability of a TMR system, assuming that the design is truly TMR or that we are using a TMR model as step one in our solution approach.

4.7.3 TMR Reliability

Markov Model. We begin the analysis of voting systems with repair by analyzing the reliability of a TMR system. The Markov reliability diagram for a TMR system composed of a voter, V, and three digital subsystems x_1, x_2, and x_3 is given in Fig. 4.11. It is assumed that the xs are identical and have the same failure rate, λ, and that the voter does not fail.

If we compare Fig. 4.11 with the model given in Fig. 3.14 of Chapter 3, we see that they are essentially the same, only with different parameter values (transition rates). There are three states in both models: repair occurs from state s_1 to s_0, and state s_2 is an absorbing state. (Actually, a complete model for Fig. 4.11 would have a fourth state, s_3, which is reached by an additional failure from state s_2. However, we have included both states in state s_2 since either two or three failures both represent system failure. As a rule, it is almost always easier to use a Markov model with fewer states even if one or more of the states represent combined states. State s_2 is actually a combined state, also known as a merged state, and a complete discussion of the rules for merging appears in Shooman [1990, p. 529]. One could decompose the third state in Fig. 4.11 into $s_2 = \bar{x}_1\bar{x}_2x_3 + \bar{x}_1x_2\bar{x}_3 + x_1\bar{x}_2\bar{x}_3$ and $s_3 = \bar{x}_1\bar{x}_2\bar{x}_3$ by reformulating the model as a more complex four-state model. However, the four-state model is not needed to solve for the upstate probabilities P_{s_0} and P_{s_1}. Thus the simpler three-state model of Fig. 4.11 will be used.)

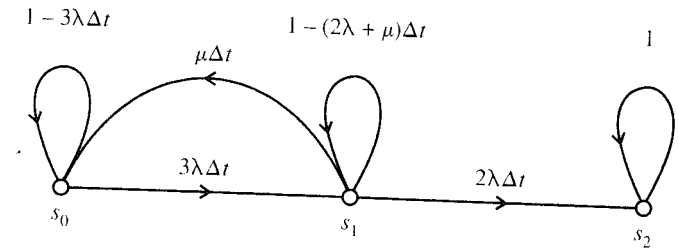


Figure 4.11 A Markov reliability model for a TMR system with repair.

In the TMR model of Fig. 4.11, there are three ways to experience a single failure from s_0 to s_1 and two ways for failures to move the system state from s_1 to s_2. Figure 3.14 of Chapter 3 uses failure rates of λ' and λ in the model; by substituting appropriate values, the model could hold for two parallel elements or for one on-line and one standby element. One can save repeating a lot of analysis and solution by realizing that the solution given in Eqs. (3.62)–(3.66) will also hold for the model of Fig. 4.11 if we let $\lambda' = 3\lambda$ (three ways to go from state s_1 to state s_2); $\lambda = 2\lambda$ (two ways to go from state s_2 to state s_3); and $\mu' = \mu$ (single repairman in both cases). Substituting these values in Eqs. (3.65) yields

$$P_{s_0}(s) = \frac{s + 2\lambda + \mu}{s^2 + (5\lambda + \mu)s + 6\lambda^2} \tag{4.25a}$$

$$P_{s_1}(s) = \frac{3\lambda}{s^2 + (5\lambda + \mu)s + 6\lambda^2} \tag{4.25b}$$

$$P_{s_2}(s) = \frac{6\lambda}{s[s^2 + (5\lambda + \mu)s + 6\lambda^2]} \tag{4.25c}$$

Note that as a check, we sum Eqs. (4.25a–c) and obtain the value $1/s$, which is the transform of unity. Thus the three equations sum to 1, as they should.

One can add the equations for P_{s_0} and P_{s_1} to obtain the reliability of a TMR system with repair in the transform domain.

$$R_{\text{TMR}}(s) = \frac{s + 5\lambda + \mu}{s^2 + (5\lambda + \mu)s + 6\lambda^2} \tag{4.26a}$$

The denominator polynomial factors into $(s + 2\lambda)$ and $(s + 3\lambda)$, and partial fraction expansion yields

$$R_{\text{TMR}}(s) = \frac{\left(\dfrac{3\lambda+\mu}{\lambda}\right)}{s+2\lambda} - \frac{\left(\dfrac{2\lambda+\mu}{\lambda}\right)}{s+3\lambda} \tag{4.26b}$$

Using transform #4 in Table B6 in Appendix B, we obtain the time function:

$$R_{\text{TMR}}(t) = \left(3+\frac{\mu}{\lambda}\right)e^{-2\lambda t} - \left(2+\frac{\mu}{\lambda}\right)e^{-3\lambda t} \tag{4.26c}$$

One can check the above result by letting $\mu = 0$ (no repair), which yields $R_{\text{TMR}}(t) = 3e^{-2\lambda t} - 2e^{-3\lambda t}$, and if $p = e^{-\lambda t}$, this becomes $R_{\text{TMR}} = 3p^2 - 2p^3$, which of course agrees with the result previously computed [see Eq. (4.2)].

Initial Behavior. The complete solution for the reliability of a TMR system with repair is given in Eq. (4.26c). It is useful to practice with the simplifying effects of initial behavior, final behavior, and MTTF solutions on this simple problem before they are applied later in this chapter to more complex models where the simplification is needed. One can evaluate the effects of repair on the initial behavior of the TMR system simply by using the transform for t^n, which is discussed in Appendix B, Section B8.3. We begin with Eq. (4.26a), where division of the denominator into the numerator using polynomial long division yields for the first three terms:

$$R_{\text{TMR}}(s) = \frac{1}{s} - \frac{6\lambda^2}{s^3} + \frac{6\lambda^2(5\lambda+\mu)}{s^4} - \cdots \tag{4.27a}$$

Using inverse transform no. 5 of Table B6 of Appendix B yields

$$\mathcal{L}\left\{\frac{1}{(n-1)!}t^{n-1}e^{-at}\right\} = \frac{1}{(s+a)^n} \tag{4.27b}$$

Setting $a = 0$ yields

$$\mathcal{L}\left\{\frac{1}{(n-1)!}t^{n-1}\right\} = \frac{1}{(s)^n} \tag{4.27c}$$

Using the transform in Eq. (4.27c) converts Eq. (4.27a) into the time function, which is a three-term polynomial in t (the first three terms in the Taylor series expansion of the time function).

$$R_{\text{TMR}}(t) = 1 - 3\lambda^2t^2 + \lambda^2(5\lambda+\mu)t^3 \cdots \tag{4.27d}$$

We previously studied the first two terms in the Taylor series expansion of

the TMR reliability expansion in Eq. (4.15). In Eq. (4.27d), we have a three-term solution, and one can compare Eqs. (4.15) and (4.27b) by calculating an additional third term in the expansion of Eq. (4.15). The expansions in Eq. (4.15) are augmented by including the cubic terms in the expansions of the bracketed terms, that is, $-4\lambda^3t^3/3$ in the first bracket and $+\lambda^3t^3/3$ in the second bracket. Carrying out the algebra adds a third term, and Eq. (4.15) becomes expanded as follows:

$$R_{\text{TMR}}(3\text{–}2) = 1 - 3\lambda^2t^2 + 5\lambda^3t^3 \tag{4.27e}$$

Thus the first three terms of Eq. (4.15) and Eq. (4.27d) are identical for the case of no repair, $\mu = 0$. Equation (4.27d) is larger (closer to unity) than the expanded version of Eq. (4.15) because of the additional term $+\lambda^2\mu t^3$ that is significant for large values of repair rate; we therefore see that repair improves the reliability. However, we note that repair only affects the cubic term in Eq. (4.27d) and not the quadratic term. Thus, for *very* small t, repair does not affect the initial behavior; however, from the above solution, we can see that it is beneficial for small and modest size t.

A numerical example will illustrate the improvement in initial reliability due to repair. Let $\mu = 10\lambda$; then the third term in Eq. (4.27d) becomes $+15\lambda^3t^3$ rather than $+5\lambda^3t^3$ with no repair. One can evaluate the increase due to $\mu = 10\lambda$ at one point in time by letting $t = 0.1/\lambda$. At this point in time, the TMR reliability without repair is equal to 0.975; with repair, it is 0.985. Further comparisons of the effects of repair appear in the problems at the end of the chapter.

The approximate analysis of this section led to a useful evaluation of the effects of repair through the computation of the power series expansion of the time function for the model with repair. This approximate result avoids the need to factor the denominator polynomial in the Laplace transform solution, which was found to be a stumbling block in obtaining a complete closed solution for higher-order systems. The next section will discuss the mean time to failure (MTTF) as another approximate solution that also avoids polynomial factoring.

Mean Time to Failure. As we saw in the preceding chapter, the computation of MTTF greatly simplifies the analysis, but it is not without pitfalls. The MTTF computes the "area under the reliability curve" (see also Section 3.8.3). Thus, for a single element with a reliability function of $e^{-\lambda t}$, the area under the curve yields $1/\lambda$; however, the MTTF calculation for the TMR system given in Eq. (4.11) yields a value of $5/6\lambda$. This implies that a single element is better than TMR, but we know that TMR has a higher reliability than a single element (see also Siewiorek [1992, p. 294]). The explanation of this apparent contradiction is simple if we examine the $n = 0$ and $n = 1$ curves in Fig. 4.4. In the region of primary interest, $0 < \lambda t < 0.69$, TMR *is* superior to a single element, but in the region $0.69 < \lambda t < \infty$ (not a region of primary interest),

the single element has a superior reliability. Thus, in computing the integral between $t = 0$ and $t = \infty$, the long tail controls the result. The lesson is that we should not trust an MTTF comparison without further study unless there is a significant superiority or unless the two reliability functions have the same shape. Clearly, if the two functions have the same shape, then a comparison of the MTTF values should be definitive. Graphing of reliability functions in the high-reliability region should always be included in an analysis, especially with the ready availability, power, and ease provided by software on a modern PC. One can also easily integrate the functions in question by using an analysis program to compute MTTF.

We now apply the simple method given in Appendix B, Section B8.2 to evaluate the MTTF by letting s approach zero in the Laplace transform of the reliability function—Eq. (4.26a). The result is

$$\text{MTTF} = \frac{5 + \mu/\lambda}{6\lambda} \tag{4.28}$$

To evaluate the effect of repair, let $\mu = 10\lambda$. The MTTF without repair increases from $5/6\lambda$ to $16/6\lambda$—a threefold improvement.

Final Behavior. The Laplace transform has a simple theorem that allows us to easily calculate the final value of a time function based on its transform. (See Appendix B, Table B7, Theorem 7.) The final-value theorem states that the value of the time function $f(t)$ as $t \to \infty$ is given by $sF(s)$ (the transform multiplied by s) as $s \to 0$. Applying this to Eq. (4.26a), we obtain

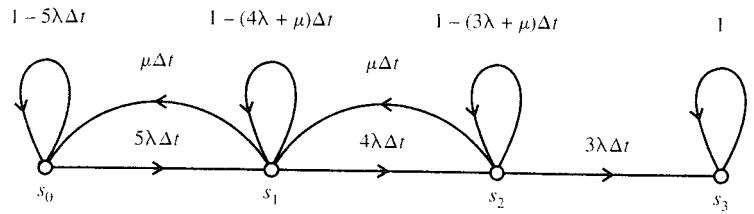
$$\lim_{s \to 0} \{sR_{\text{TMR}}\} = \lim_{s \to 0} \frac{s(s + 5\lambda + \mu)}{s^2 + (5\lambda + \mu)s + 6\lambda^2} = 0 \tag{4.29}$$

A little thought shows that this is the correct result since all reliability functions go to zero as time increases. However, when we study the availability function later in this chapter, we will see that the final value of the availability is nonzero. This value is an important measure of system behavior.

4.7.4 *N*-Modular Reliability

Having explored the analysis of the reliability of a TMR system with repair, it would be useful to develop general expressions for the reliability, MTTF, and initial behavior for N-modular systems. This task is difficult and probably unnecessary since most practical systems have 3- or 5-level majority voting. (An intermediate system with 4-level voting used by NASA in the Space Shuttle will be discussed later in this chapter.) The main focus of this section will therefore be the analysis.

Markov Model. We begin the analysis of 5-level modular reliability with



Zero failures	One failure	Two failures	Three or more failures
$s_0 = x_1 x_2 x_3 x_4 x_5$	$s_1 = \bar{x}_1 x_2 x_3 x_4 x_5 + x_1 \bar{x}_2 x_3 x_4 x_5 + x_1 x_2 \bar{x}_3 x_4 x_5 + x_1 x_2 x_3 \bar{x}_4 x_5 + x_1 x_2 x_3 x_4 \bar{x}_5$	$s_2 = \bar{x}_1 \bar{x}_2 x_3 x_4 x_5 + \bar{x}_1 x_2 \bar{x}_3 x_4 x_5 + \bar{x}_1 x_2 x_3 \bar{x}_4 x_5 + \bar{x}_1 x_2 x_3 x_4 \bar{x}_5$ + (6 more terms)	$s_3 = \bar{x}_1 \bar{x}_2 x_3 x_4 x_5 + \bar{x}_1 \bar{x}_2 x_3 \bar{x}_4 x_5 + \bar{x}_1 \bar{x}_2 x_3 x_4 \bar{x}_5 + \bar{x}_1 x_2 \bar{x}_3 x_4 x_5$ + (12 more terms)

Figure 4.12 A Markov reliability model for a 5-level majority voting system with repair.

repair by formulating the Markov model given in Fig. 4.12. We follow the same approach used to formulate the Markov model given in Fig. 4.11. There are, however, additional states. (Actually, there is one additional state that lumps together three other states.)

The Markov time-domain differential equations are written in a manner analogous to that used in developing Eqs. (3.62a–c). The notation $\dot{P}_s = dP_s/dt$ is used for convenience, and the following equations are obtained:

$$\dot{P}_{s_0}(t) = -5\lambda P_{s_0}(t) + \mu P_{s_1}(t) \tag{4.30a}$$
$$\dot{P}_{s_1}(t) = 5\lambda P_{s_0}(t) - (4\lambda + \mu)P_{s_1}(t) + \mu P_{s_2}(t) \tag{4.30b}$$
$$\dot{P}_{s_2}(t) = 4\lambda P_{s_1}(t) - (3\lambda + \mu)P_{s_2}(t) \tag{4.30c}$$
$$\dot{P}_{s_3}(t) = 3\lambda P_{s_2}(t) \tag{4.30d}$$

Taking the Laplace transform of the preceding equations and incorporating the initial conditions $P_{s_0}(0) = 1$, $P_{s_1}(0) = P_{s_2}(0) = P_{s_3}(0) = 0$ leads to the transformed equations as follows:

$$(s + 5\lambda)P_{s_0}(s) - \mu P_{s_1}(s) = 1 \tag{4.31a}$$
$$-5\lambda P_{s_0}(s) + (s + 4\lambda + \mu)P_{s_1}(s) - \mu P_{s_2}(s) = 0 \tag{4.31b}$$
$$-4\lambda P_{s_1}(s) + (s + 3\lambda + \mu)P_{s_2}(s) = 0 \tag{4.31c}$$
$$3\lambda P_{s_2}(s) + sP_{s_3}(s) = 0 \tag{4.31d}$$

Equations (4.31a–d) can be solved by a variety of means for the probabilities $P_{s_0}(t)$, $P_{s_1}(t)$, $P_{s_2}(t)$, and $P_{s_3}(t)$. One technique based on Cramer's rule is to formulate a set of determinants associated with the equations. Each of the probabilities becomes a ratio of two of the determinants: a numerator deter-

minant divided by a denominator determinant. The denominator determinant is the same for each ratio; it is generally denoted by Δ and is the determinant of the coefficients of the equations. (One can develop the form of these equations in a more elaborate fashion using matrix theory; see Shooman [1990, pp. 239–243].) A brief inspection of Eqs. (4.31a–d) shows that the first three are uncoupled from the last and can be solved separately, simplifying the algebra (this will always be true in a Markov model with repair when the last state is an absorbing one). Thus, for the first three equations,

$$\Delta = \begin{vmatrix} s+5\lambda & -\mu & 0 \\ -5\lambda & s+4\lambda+\mu & -\mu \\ 0 & -4\lambda & s+3\lambda+\mu \end{vmatrix} \tag{4.32}$$

The numerator determinants in the solution are similar to the denominator determinants; however, one column is replaced by the right-hand side of the Eqs. (4.31a–d); that is,

$$\Delta_1 = \begin{vmatrix} 1 & -\mu & 0 \\ 0 & s+4\lambda+\mu & -\mu \\ 0 & -4\lambda & s+3\lambda+\mu \end{vmatrix} \tag{4.33a}$$

$$\Delta_2 = \begin{vmatrix} s+5\lambda & 1 & 0 \\ -5\lambda & 0 & -\mu \\ 0 & 0 & s+3\lambda+\mu \end{vmatrix} \tag{4.33b}$$

$$\Delta_3 = \begin{vmatrix} s+5\lambda & -\mu & 1 \\ -5\lambda & s+4\lambda+\mu & 0 \\ 0 & -4\lambda & 0 \end{vmatrix} \tag{4.33c}$$

In terms of this group of determinants, the probabilities are

$$P_{s_0}(s) = \frac{\Delta_1}{\Delta} \tag{4.34a}$$

$$P_{s_1}(s) = \frac{\Delta_2}{\Delta} \tag{4.34b}$$

$$P_{s_2}(s) = \frac{\Delta_3}{\Delta} \tag{4.34c}$$

The reliability of the 5-level modular redundancy system is given by

$$R_{5\,\mathrm{MR}}(t) = P_{s_0}(t) + P_{s_1}(t) + P_{s_2}(t) \tag{4.35}$$

Expansion of the denominator determinant yields the following polynomial:

$$\Delta = s^3 + (12\lambda + 2\mu)s^2 + (47\lambda^2 + 8\lambda\mu + \mu^2)s + 60\lambda^3 \tag{4.36a}$$

Similarly, expanding the other determinants yields the following polynomials:

$$\Delta_1 = s^2 + (7\lambda + 2\mu)s + 12\lambda^2 + 3\lambda\mu + \mu^2 \tag{4.36b}$$

$$\Delta_2 = 5\lambda(s + 3\lambda + \mu) \tag{4.36c}$$

$$\Delta_3 = 20\lambda^2 \tag{4.36d}$$

Substitution in Eqs. (4.34a–c) and (4.35) yields the transform of the reliability function:

$$R_{5\,\mathrm{MR}}(s) = \frac{s^2 + (12\lambda + 2\mu)s + 47\lambda^2 + 8\lambda\mu + \mu^2}{s^3 + (12\lambda + 2\mu)s^2 + (47\lambda^2 + 8\lambda\mu + \mu^2)s + 60\lambda^3} \tag{4.37}$$

As a check, we compute the probability of being in the fourth state $P_{s_3}(s)$ from Eq. (4.31d) as

$$P_{s_3}(s) = \frac{3\lambda P_{s_2}(s)}{s} = \frac{60\lambda^3}{s\Delta} \tag{4.38}$$

Adding Eq. (4.37) to Eq. (4.38) and performing some algebraic manipulation yields $1/s$, which is the transform of unity. Thus the sum of all the state probabilities adds to unity as it should and the results check.

Initial Behavior. As in the preceding section, we can model the initial behavior by expanding the transform Eq. (4.37) into a series in inverse powers of s using polynomial division. The division yields

$$R_{5\,\mathrm{MR}}(s) = \frac{1}{2} - \frac{60\lambda^3}{s^4} + \frac{60\lambda^3(12\lambda + 2\mu)}{s^5} - \cdots \tag{4.39a}$$

Applying the inverse transform of Eq. (4.27c) yields

$$R_{5\,\mathrm{MR}}(s) = 1 - 10\lambda^3 t^3 + 2.5\lambda^3(12\lambda + 2\mu)t^4 \cdots \tag{4.39b}$$

We can compare the gain due to 5-level modular redundancy with repair to that of TMR with repair by letting $\mu = 10\lambda$ and $t = 0.1/\lambda$, as in Section 4.7.3, which gives a reliability of 0.998. Without repair, the reliability would be 0.993. These values should be compared with the TMR reliability without repair, which is equal to 0.975, and TMR with repair, which is 0.985. Since it is difficult to compare reliabilities close to unity, we can focus on the unreliabilities with repair. The 5-level voting has an unreliability of 0.002; the TMR, 0.015. Thus, the change in voting from 3-level to 5-level has reduced the unre-

TABLE 4.6 Comparison of the MTTF for Several Voting and Parallel Systems with Repair

System	MTTF Equation	$\mu = 0$	$\mu = 10$	$\mu = 100$
TMR with repair	$\frac{5 + \frac{\mu}{\lambda}}{6\lambda}$	$\frac{0.83}{\lambda}$	$\frac{2.5}{\lambda}$	$\frac{17.5}{\lambda}$
5MR with repair	$\frac{47 + 8\frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2}{60\lambda^3}$	$\frac{0.78}{\lambda}$	$\frac{3.78}{\lambda}$	$\frac{180.78}{\lambda}$
Two parallel	$\frac{3\lambda + \mu}{2\lambda^2}$	$\frac{1.5}{\lambda}$	$\frac{6.5}{\lambda}$	$\frac{51.5}{\lambda}$
Two standby	$\frac{2\lambda + \mu}{\lambda^2}$	$\frac{2}{\lambda}$	$\frac{12}{\lambda}$	$\frac{102}{\lambda}$

liability by a factor of 7.5. Further comparisons of the effects of repair appear in the problems at the end of this chapter.

Mean Time to Failure Comparison. The MTTF for 5-level voting is easily computed by letting s approach 0 in the transform equation, which yields

$$\text{MTTF}_{5\,\text{MR}} = \frac{47\lambda^2 + 8\lambda\mu + \mu^2}{60\lambda^3} \tag{4.40}$$

This MTTF is compared with some other systems in Table 4.6. The table shows, as expected, that 5MR is superior to TMR when repair is present. Note that two parallel or two standby elements appear more reliable. Once reduction in reliability due to the reliability of the coupler and coverage is included and compared with the reduction due to the reliability of the voter, this advantage may disappear.

Initial Behavior Comparison. The initial behavior of the systems given in Table 4.6 is compared in Table 4.7 using Eqs. (4.27d) and (4.39b) for TMR and 5MR systems. For the case of two ordinary parallel and two standby systems, we must derive the initial behavior equation by adding Eqs. (3.65a) and (3.65b) to obtain the transform of the reliability function that holds for both parallel and standby systems.

$$R(s) = P_{s_0}(s) + P_{s_1}(s) = \frac{s + \lambda + \lambda' + \mu'}{s^2 + (\lambda + \lambda' + \mu')s + \lambda\lambda'} \tag{4.41}$$

For an ordinary parallel system, $\lambda' = 2\lambda$ and $\mu' = \mu$, and substitution into Eq. (4.41), long division of the denominator into the numerator, and inversion of

TABLE 4.7 Comparison of the Initial Behavior for Several Voting and Parallel Systems with Repair

System	Initial Reliability Equation, $\mu = 10\lambda$	Value of t at which $R = 0.999$
TMR with repair	$1 - 3(\lambda t)^2 + 15(\lambda t)^3$	$\frac{0.0192}{\lambda}$
5MR with repair	$1 - 10(\lambda t)^3 + 80(\lambda t)^4$	$\frac{0.057}{\lambda}$
Two parallel	$1 - (\lambda t)^2 + 4.33(\lambda t)^3$	$\frac{0.034}{\lambda}$
Two standby	$1 - 0.5(\lambda t)^2 + 2(\lambda t)^3$	$\frac{0.045}{\lambda}$

the transform (as was done previously) yields

$$R_{\text{parallel}}(t) = 1 - (\lambda t)^2 + \lambda^2(3\lambda + \mu)t^3/3 \tag{4.42a}$$

For a standby system, $\lambda' = \lambda$ and $\mu' = \mu$, and substitution into Eq. (4.41), long division, and inversion of the transform yields

$$R_{\text{standby}}(t) = 1 - (\lambda t)^2/2 + \lambda^2(2\lambda + \mu)t^3/6 \tag{4.42b}$$

Equations (4.42a) and (4.42b) appear in Table 4.7 along with Eqs. (4.27d) and (4.39b), where $\mu = 10\lambda$ has been substituted.

Table 4.7 shows that the length of time the reliability takes to decay from 1 to 0.999, which makes it clearly a high-reliability region. For the TMR system, the duration is $t = 0.0192\lambda$; for the 5-level voting system, $t = 0.057\lambda$. Thus the 5-level system represents an increase of nearly 3 over the 3-level system. One can better appreciate these numerical values if typical values are substituted for λ. The length of a year is 8,766 hours, which is often approximated as 10,000 hours. A high-reliability computer may have an MTTF($1/\lambda$) of about 10 years, or approximately 100,000 hours. Substituting this value for t shows that the reliability of a TMR system with a repair rate of 10 times the failure rate will have a reliability exceeding 0.999 for about 1,920 hours. Similarly, a 5-level voting system will have a reliability exceeding 0.999 for about 5,700 hours. In the case of the parallel and standby systems, the high-reliability region is longer than in a TMR system, but is less than in a 5-level voter system.

Higher-Level Voting. One could extend the above analysis to cover higher-level voting systems; for example, 7-level and 9-level voting. Even though it is easy to replicate many different copies of a logic circuit on a chip at low

cost, one seldom goes beyond the 3-level or 5-level voting system, although the foregoing methods could be used to solve for the reliability of such higher-level systems.

If one fabricates a very large scale integrated circuit (VLSI) with many circuits and a voter, an interesting question arises. There is a yield problem with complex chips caused by imperfections. With so much redundancy, how can one be sure that the chip does not contain such imperfections that a 5-level voter system with imperfections is really equivalent to a 4- or 3-level voter system? In fact, a 5-level voter system with two failed circuits is actually inferior to a 3-level voter. One more failure in the former will result in three failed and two good circuits, and the voter believes the failed three. In the case of a 3-level voter, a single failure will still leave the remaining two good circuits in control. The solution is to provide internal test inputs on an IC voter system so that the components of the system can be tested. This means that extra pins on the chip must be dedicated to test points. The extra outputs in Fig. 4.10 could provide these test points, as was discussed in Section 4.6.2.

The next section discusses the effect of voter reliability on N-modular redundancy. Note that we have not discussed the effects of coverage in a TMR system. In general, the simple nature of a voter catches almost all failures, and coverage is not significant in modeling the system.

4.8 N-MODULAR REDUNDANCY WITH REPAIR AND IMPERFECT VOTERS

4.8.1 Introduction

The analysis of the preceding section did not include two imperfections in a voting system: the reliability of the voter itself and also the concept of coverage. In the case of parallel and standby systems, which were treated in Chapter 3, coverage made a considerable difference in the reliability. The circuit that detected failures of the active system and switched to the standby (hot or cold) element in a parallel or standby system is reasonably complex and will have a significant failure rate. Furthermore, it will have the problem that it cannot detect all faults and will sometimes fail to switch when it should or switch when it should not. In the case of a voter, the concept and the resulting circuit is much simpler. Thus one might be justified in assuming that the voter does not have a coverage problem and so reduce our evaluation to the reliability of a voter and how it affects the system reliability. This can then be contrasted with the reliability of a coupler and a parallel system (introduced in Section 3.5).

4.8.2 Voter Reliability

We begin our discussion of voter reliability by considering the reliability of a TMR system as shown in Fig. 4.1 and the reliability expression given in

Eq. (4.19). In Section 4.5, we asked how small the voter reliability, p_v, can be so that the gains of TMR still exceed the reliability of a single circuit. The analysis was given in Eqs. (3.34) and (3.35). Now, we perform a similar analysis for a TMR system with an imperfect voter. The computation proceeds from a consideration of Eq. (4.19). If the voter were perfect, $p_v = 1$, then the reliability would be computed as

$$R_{\text{TMR}} = 3p_c^2 - 2p_c^3 \tag{4.43a}$$

If we include an imperfect voter, this expression becomes

$$R_{\text{TMR}} = 3p_v p_c^2 - 2p_v p_c^3 = p_v(3p_c^2 - 2p_c^3) \tag{4.43b}$$

If we assume constant-failure rates for the voter and the circuits in the TMR configuration, then for the voter we have $p_v = e^{-\lambda_v t}$, and for the TMR circuits, $p = e^{-\lambda t}$. If we use a three-term approximation for the exponential and substitute into Eq. (4.43b), one obtains an expression for the initial reliability, as follows:

$$R_{\text{TMR}} = \left(1 - \lambda_v t + \frac{(\lambda_v t)^2}{2!} - \frac{(\lambda_v t)^3}{3!}\right) \times \left[3\left(1 - 2\lambda_v t + \frac{(2\lambda t)^2}{2!} - \frac{(2\lambda t)^3}{3!}\right) - 2\left(1 - 3\lambda t + \frac{(3\lambda t)^2}{2!} - \frac{(3\lambda_v t)^3}{3!}\right)\right] \tag{4.44a}$$

Expanding the preceding equation and retaining only the first four terms yields

$$R_{\text{TMR}} = 1 - \lambda_v t + \frac{(\lambda_v t)^2}{2} - 3(\lambda t)^2 \tag{4.44b}$$

Furthermore, we are mainly interested in the cases where $\lambda_v < \lambda$; thus we can omit the third term (which is a second-order term in λ_v) and obtain

$$R_{\text{TMR}} = 1 - \lambda_v t - 3(\lambda t)^2 \tag{4.44c}$$

If we want the effect of the voter to be negligible, we let $\lambda_v t < 3(\lambda t)^2$,

$$\frac{\lambda_v}{\lambda} < 3\lambda t \tag{4.45}$$

One can compare this result with that given in Eq. (3.35) for two parallel systems by setting $n = 2$, yielding

$$\frac{\lambda_c}{\lambda} < \lambda t \tag{3.35}$$

The approximate result is that the coupler must have a failure rate three times smaller than that of the voter for the same decrease in reliability.

One can examine the effect of repair on the above results by examining Eq. (4.27d) and Eq. (4.42). In both cases, the effect of the repair rate does not appear until the cubic term is encountered. The above comparisons only involved the linear and quadratic terms, so the effect of repair would only become apparent if the repair rate were very large and the time interval of interest were extended.

4.8.3 Comparison of TMR, Parallel, and Standby Systems

Another advantage of voter reliability over parallel and standby reliability is that there is a straightforward scheme for implementing voter redundancy (e.g., Fig. 4.8). Of course, one can also make redundant couplers for parallel or standby systems, but they may be more complex than redundant voters.

It is easy to make a simple model for Fig. 4.8. Assume that the voters fail so that their outputs are stuck-at-zero or stuck-at-one and that voter failures do not corrupt the outputs of the circuits that feed the voters (e.g., A_1, B_1, and C_1). Assume just a single stage (A_1, B_1, and C_1) and a single redundant voter system (V_1, V_1', and V_1''). The voter works if two or three of the three voters work. Thus this is the same formula for TMR systems, and the reliability of the system becomes

$$R_{\text{TMR}} \times R_{\text{voter}} = (3p_c^2 - 2p_c^3) \times (3p_v^2 - 2p_v^3) \tag{4.46}$$

It is easy to evaluate the advantages of redundant voters. Assume that $p_c = 0.9$ and that the voter is 10 times as reliable: $(1 - p_c) = 0.1$, $(1 - p_v) = 0.01$, and $p_v = 0.99$. With a single voter, $R = 0.99[3(0.9)^2 - 2(0.9)^3] = 0.99 \times 0.972 = 0.962$. In the case of a redundant voter, we have $[3(0.99)^2 - 2(0.99)^3] \times [3(0.9)^2 - 2(0.9)^3] = 0.999702 \times 0.972 = 0.9717$. The redundant voter is thus significant; if the voter is less reliable, voter redundancy is even more effective. Assume that $p_v = 0.95$; for a single voter, $R = 0.95\ [3(0.9)^2 - 2(0.9)^3] = 0.95 \times 0.972 = 0.923$. In the case of a redundant voter, we have $[3(0.95)^2 - 2(0.95)^3] \times [3(0.9)^2 - 2(0.9)^3] = 0.99275 \times 0.972 = 0.964953$.

The foregoing calculations and discussions were performed for a TMR circuit with a single voter or redundant voters. It is possible to extend these computations to the subsystem level for a system such as that depicted in Fig. 4.8. In addition, one can repair a failed component of a redundant voter; thus one can use the analysis techniques previously derived for TMR and 5MR systems where the systems and voters can both be repaired. However, repair of voters really begs a larger question: How will we modularize the system architecture?

Assume one is going to design the system architecture with redundant voters and voting at a subsystem level. If the voters are to be placed on a single chip along with the circuits, then there is no separate repair of a voter system—only repair of the circuit and voter subsystem. The alternative is to make a separate chip for the N circuits and a separate chip for the redundant voter. The proper strategy to choose depends on whether there will be scheduled downtime for the system during which testing and replacement can occur and also whether the chips have sufficient test points. No general conclusion can be reached; the system architecture should be critiqued with these issues in mind.

4.9 AVAILABILITY OF N-MODULAR REDUNDANCY WITH REPAIR AND IMPERFECT VOTERS

4.9.1 Introduction

When repair is present in a system, it is often possible for the system to fail and be down for a short period of time without serious operational effects. Suppose a computer used for electronic funds transfers is down for a short period of time. This is not catastrophic if the system is designed so that it can tolerate brief outages and perform the funds transfers at a later time period. If the system is designed to be self-diagnostic, and if a technician and a replacement plug in boards are both available, the machine can be restored quickly to operational status. For such systems, availability is a useful measure of system performance, as with reliability, and is the probability that the system is up at any point in time. It can be measured during operation by recording the downtimes and operating times for several failure and repair cycles. The availability is given by the ratio of the sum of the uptimes for the system divided by the sum of the uptimes and the downtimes. (Formally, this ratio becomes the availability in the limit as the system operating time approaches infinity.) The availability $A(t)$ is the probability that the system is up at time t, which can be written as a sum of probabilities:

$$\begin{aligned} A(t) = {} & P(\text{no failures}) + P(\text{one failure + one repair}) \\ & + P(\text{two failures + two repairs}) \\ & + \cdots + P(n \text{ failures} + n \text{ repairs}) + \cdots \end{aligned} \tag{4.47}$$

Availability is always higher than reliability, since the first term in Eq. (4.47) is the reliability and all the other terms are positive numbers. Note that only the first few terms in Eq. (4.47) are significant for a moderate time interval and higher-order terms become negligible. Thus one could evaluate availability analytically by computing the terms in Eq. (4.47); however, the use of the Markov model simplifies such a computation.

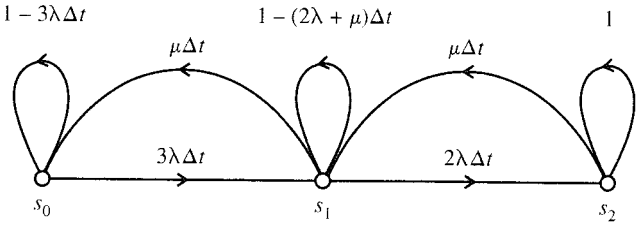


Figure 4.13 A Markov availability model for a TMR system with repair.

4.9.2 Markov Availability Models

A brief introduction to availability models appeared in Section 3.8.5; such computations will continue to be used in this section, and availabilities for TMR systems, parallel systems, and standby systems will be computed and compared. As in the previous section, we will make use of the fact that the Markov availability model given in Fig. 3.16 will hold with minor modifications (see Fig. 4.13). In Fig. 3.16, the value of λ' is either one or two times λ, but in the case of TMR, it is three times λ. For the second transmission between s_1 and s_2 for the TMR system, there are two possibilities of failure; thus the transition rate is 2λ. Since there is only one repairman, the repair rate is μ.

A set of Markov equations can be written that will hold for two in parallel and two in standby, as well as for TMR. The algorithm used in the preceding chapter will be employed. The terms 1 and Δt are deleted from Fig. 4.13. The time derivative of the probability of being in state s_0 is set equal to the "flows" from the other nodes; for example, $-\lambda' P_{s_0}(t)$ is from the self-loop and $\mu' P_{s_1}(t)$ is from the repair branch. Applying the algorithm to the other nodes and using algebraic manipulation yields the following:

$$\dot{P}_{s_0}(t) + \lambda' P_{s_0}(t) = \mu' P_{s_1}(t) \tag{4.48a}$$

$$\dot{P}_{s_1}(t) + (\lambda + \mu') P_{s_1}(t) = \lambda' P_{s_0}(t) + \mu'' P_{s_2}(t) \tag{4.48b}$$

$$\dot{P}_{s_2}(t) + \mu'' P_{s_2}(t) = \lambda P_{s_1}(t) \tag{4.48c}$$

$$P_{s_0}(0) = 1 \qquad P_{s_1}(0) = P_{s_2}(0) = 0 \tag{4.48d}$$

The appropriate values of parameters for this set of equations is given in Table 4.8. A complete solution of these equations is given in Shooman [1990, pp. 344–347]. We will use the Laplace transform theorems previously introduced to simplify the solution.

The Laplace transforms of Eqs. (4.48a–d) become

TABLE 4.8 Parameters of Eqs. (4.48a–d) for Various Systems

System	λ	λ'	μ'	μ''
Two in parallel	λ	2λ	μ	μ
Two standby	λ	λ	μ	μ
TMR	2λ	3λ	μ	μ

$$(s + \lambda')P_{s_0}(s) - \mu' P_{s_1}(s) = 1 \tag{4.49a}$$

$$-\lambda' P_{s_0}(s) + s(s + \lambda + \mu')P_{s_1}(s) - \mu'' P_{s_2}(s) = 0 \tag{4.49b}$$

$$-\lambda P_{s_1}(s) + (s + \mu'')P_{s_2}(s) = 0 \tag{4.49c}$$

In the case of a system composed of two in parallel, two in standby, or TMR, the system is up if it is in state s_0 or state s_1. The availability is thus the sum of the probabilities of being in one of these two states. If one uses Cramer's rule or a similar technique to solve Eqs. (4.49a–c), one obtains a ratio of polynomials in s for the availability:

$$A(s) = P_{s_0}(s) + P_{s_1}(s) = \frac{s^2 + (\lambda + \lambda' + \mu' + \mu'')s + (\lambda'\mu'' + \mu'\mu'')}{s[s^2 + (\lambda + \lambda' + \mu' + \mu'')s + (\lambda\lambda' + \lambda'\mu'' + \mu'\mu'')]} \tag{4.50}$$

Before we begin applying the various Laplace transform theorems to this availability function, we should discuss the nature of availability and what sort of analysis is needed. In general, availability always starts at 1 because the system is always assumed to be up at $t = 0$. Examination of Eq. (4.47) shows that initially near $t = 0$, the availability is just the reliability function that of course starts at 1. Gradually, the next term P(one failure and one repair) becomes significant in the availability equation; as time progresses, other terms in the series contribute. Although the overall effect based on the summation of these many terms is hard to understand, we note that they generally lead to a slow decay of the availability function to some steady-state value that is reasonably close to 1. Thus the initial behavior of the availability function is not as important as that of the reliability function. In addition, the MTTF is not always a significant measure of system behavior. The one measure of interest is the final value of the availability function. If the availability function for a particular system has an initial value of unity at $t = 0$ and decays slowly to a steady-state value close to unity, this system must always have a high value of availability, in which case the final value is a lower bound on the availability. Examining Table B7 in Appendix B, Section B8.1, we see that the final value and initial value theorems both depend on the limit of $sF(s)$ [in our case, $sA(s)$] as s approaches 0 and ∞. The initial value is when s approaches ∞. Examination of Eq. (4.50) shows that multiplication of $A(s)$ by s results in a cancellation of

TABLE 4.9 Comparison of the Steady-State Availability, Eq. (4.50) for Various Systems

System	Eq. (4.50)	$\mu = \lambda$	$\mu = 10\lambda$	$\mu = 100\lambda$
Two in parallel	$\frac{\mu(2\lambda + \mu)}{2\lambda^2 + 2\lambda\mu + \mu^2}$	0.6	0.984	0.9998
Two standby	$\frac{\mu(\lambda + \mu)}{\lambda^2 + \lambda\mu + \mu^2}$	0.667	0.991	0.9999
TMR	$\frac{\mu(3\lambda + \mu)}{6\lambda^2 + 3\lambda\mu + \mu^2}$	0.4	0.956	0.9994

the multiplying s term in the denominator. As s approaches infinity, both the numerator and denominator polynomials approach s^2; thus the ratio approaches 1, as it should. However, to find the final value, we let s approach zero and obtain the ratio of the two constant terms given in Eq. (4.51).

$$A(\text{steady state}) = \frac{(\lambda'\mu'' + \mu'\mu'')}{(\lambda\lambda' + \lambda'\mu'' + \mu'\mu'')} \tag{4.51}$$

The values of the parameters given in Table 4.8 are substituted in this equation, and the steady-state availabilities are compared for the three systems noted in Table 4.9.

Clearly, the Laplace transform has been of great help in solving for steady-state availability and is superior to the simplified time-domain method: (a) let all time derivatives equal 0; (b) delete one of the resulting algebraic equations; (c) add the equation's sum of all probabilities to equal 1; and (d) solve (see Section B7.5).

Table 4.9 shows that the steady-state availability of two elements in standby exceeds that of two parallel items by a small amount, and they both exceed the TMR system by a greater margin. In most systems, the repair rate is much higher than the failure, so the results of the last column in the table are probably the most realistic. Note that these steady-state availabilities depend only on the ratio μ/λ. Before one concludes that the small advantages of one system over another in the table are significant, the following factors should be investigated:

- It is assumed that a standby element cannot fail when it is in standby. This is not always true, since batteries discharge in standby, corrosion can occur, insulation can break down, etc., all of which may significantly change the comparison.
- The reliability of the coupling device in a standby or parallel system is more complex than the voter reliability in a TMR circuit. These effects on availability may be significant.
- Repair in any of these systems is predicated on knowing when a system has failed. In the case of TMR, we gave a simple logic circuit that would detect which element has failed. The equivalent detection circuit in the case of a parallel or standby system is more complex and may have poorer coverage.

Some of these effects are treated in the problems at the end of this chapter. It is likely, however, that the detailed design of comparative systems must be modeled to make a comprehensive comparison.

A simple numerical example will show the power of increasing system availability using parallel and standby system configurations. In Section 3.10.1, typical failure and repair information for a circa-1985 transaction-processing system was quoted. The time between failures of once every two weeks translates into a failure rate $\lambda = 1/(2 \times 168) = 2.98 \times 10^{-3}$ failures/hour, and the time to repair of one hour becomes a repair rate $\mu = 1$ repair/hour. These values were shown to yield a steady-state availability of 0.997—a poor value for what should be a highly reliable system. If we assume that the computer system architecture will be configured as a parallel system or a standby system, we can use the formulas of Table 4.9 to compute the expected increase in availability. For an ordinary parallel system, the steady-state availability would be 0.999982; for a standby system, it would be 0.9999911. Both translate into unavailability values $\bar{A} = 1 - A$ of 1.8×10^{-5} and 8.9×10^{-6}. The unavailability of the single system would of course be 3×10^{-3}. The steady-state availability of the Stratus system was discussed in Section 3.10.2 and, based on claimed downtime, was computed as 0.9999905, which is equivalent to an unavailability of 95×10^{-7}. In Section 3.10.1, the Tandem unavailability, based on hypothetical goals, was 4×10^{-6}. Comparison of these four unavailability values yields the following: (a) for a single system, $3{,}000 \times 10^{-6}$; (b) for a parallel system, 18×10^{-6}; (c) for a standby system, 8.9×10^{-6}; (d) for a Stratus system, 9.5×10^{-6}; and (e) for a Tandem system, 4×10^{-6}. Also compare the Bell Labs' ESS switching system unavailability goals and demonstrated availability of 5.7×10^{-6} and 3.8×10^{-6}. (See Table 1.4.) Of course, more definitive data or complete models are needed for detailed comparisons.

4.9.3 Decoupled Availability Models

A simplified technique can be used to compute the steady-state value of availability for parallel and TMR systems. Availability computations really involve the evaluation of certain conditional probabilities. Since conditional probabilities are difficult to deal with, we introduced the Markov model computation technique. There is a case in which the dependent probabilities become independent and the computations simplify. We will introduce this case by focusing on the availability of two parallel elements.

Assume that we wish to compute the steady-state availability of two parallel elements, A and B. The reliability is the probability of no system failures in interval 0 to t, which is the probability that either A or B is good,

$P(A_g + B_g) = P(A_g) + P(B_g) - P(A_gB_g)$. The subscript "$g$" means that the element is good, that is, has not failed. Similarly, the availability is the probability that the system is up at time t, which is the probability that either A or B is up, $P(A_{up} + B_{up}) = P(A_{up}) + P(B_{up}) = P(A_{up}B_{up})$. The subscript "up" means that the element is up, that is, is working at time t. The product terms in each of the above expressions, $P(A_gB_g) = P(A_g)P(B_g|A_g)$ and $P(A_{up}B_{up}) = P(A_{up})P(B_{up}|A_{up})$ are the conditional probabilities discussed previously. If there are two repairmen—one assigned to component A and one assigned to component B—the events $(B_g|A_g)$ and $(B_{up}|A_{up})$ become *decoupled*, that is, the events are independent. The coupling (dependence) comes from the repairmen. If there is only one repairman and element A is down and being repaired, then if element B fails, it will take longer to restore B to operation; the repairman must first finish fixing A before working on B. In the case of individual repairmen, there is no wait for repair of the second element if two items have failed because each has its own assigned repairman. In the case of such decoupling, the dependent probabilities become independent and $P(B_g|A_g) = P(B_g)$ and $P(B_{up}|A_{up}) = P(B_{up})$. This represents considerable simplification; it means that one can compute $P(B_g)$, $P(A_g)$, $P(B_{up})$, and $P(A_{up})$ separately and substitute into the reliability or availability equation to achieve a simple solution. Before we apply this technique and illustrate the simplicity of the solution, we should comment that because of the high cost, it is unlikely that there will be two separate repairmen. However, if the repair rate is much larger than the failure rate, $\mu \gg \lambda$, the decoupled case is approached. This is true since repairs are relatively fast and there is only a small probability that a failed element A will still be under repair when element B fails. For a more complete discussion of this decoupled approximation, consult Shooman [1990, pp. 521–529].

To illustrate the use of this approximation, we calculate the steady-state availability of two parallel elements. In the steady state,

$$A(\text{steady state}) = P(A_{ss}) + P(B_{ss}) - P(A_{ss})P(B_{ss}) \tag{4.52}$$

The steady-state availability for a single element is given by

$$A_{ss} = \frac{\mu}{\lambda + \mu} \tag{4.53}$$

One can verify this formula by reading the derivation in Appendix B, Sections B7.3 and B7.4, or by examining Fig. 3.16. We can reduce Fig. 3.16 to a single element model by setting $\lambda = 0$ to remove state s_2 and letting $\lambda' = \lambda$ and $\mu' = \mu$. Solving Eqs. (3.71a, b) for $P_{s_0}(t)$ and applying the final value theorem (multiply by s and let s approach 0) also yields Eq. (4.53). If A and B have identical failure and repair rates, substitution of Eq. (4.53) into Eq. (4.52) for both A_{ss} and B_{ss} yields

$$A_{ss} = \frac{2\mu}{\lambda + \mu} - \left(\frac{\mu}{\lambda + \mu}\right)^2 = \frac{\mu(2\lambda + \mu)}{(\lambda + \mu)^2} \tag{4.54}$$

If we compare this result with the exact one in Table 4.9, we see that the numerator is the same and the denominator differs only by a coefficient of two in the λ^2 term. Furthermore, since we are assuming that $\mu \gg \lambda$, the difference is very small.

We can repeat this simplification technique for a TMR system. The TMR reliability equation is given by Eq. (4.2), and modification for computing the availability yields

$$A(\text{steady state}) = [P(A_{ss})]^2[3 - P(A_{ss})] \tag{4.55}$$

Substitution of Eq. (4.53) into Eq. (4.55) gives

$$A(\text{steady state}) = \left(\frac{\mu}{\lambda + \mu}\right)^2 \left(3 - \frac{2\mu}{\lambda + \mu}\right) = \left(\frac{\mu}{\lambda + \mu}\right)^2 \left(\frac{3\lambda + \mu}{\lambda + \mu}\right) \tag{4.56}$$

There is no obvious comparison between Eq. (4.56) and the exact TMR availability expression in Table 4.9. However, numerical comparison will show that the formulas yield nearly equivalent results.

The development of approximate expressions for a standby system requires some preliminary work. The Poisson distribution (Appendix A, Section A5.4) describes the probabilities of success and failure in a standby system. The system succeeds if there are no failures or one failure; thus the reliability expression is computed from the Poisson distribution as

$$R(\text{standby}) = P(0 \text{ failures}) + P(1 \text{ failure}) = e^{-\lambda t} + \lambda t e^{-\lambda t} \tag{4.57}$$

If we wish to transform this equation in terms of the probability of success p of a single element, we obtain $p = e^{-\lambda t}$ and $\lambda t = -\ln p$. (See also Shooman [1990, p. 147].) Substitution into Eq. (4.57) yields

$$R(\text{standby}) = p(1 - \ln p) \tag{4.58}$$

Finally, substitution in Eq. (4.58) of the steady-state availability from Eq. (4.53) yields an approximate expression for the availability of a standby system as follows:

$$A(\text{steady state}) = \left[\frac{\mu}{\lambda + \mu}\right]\left[1 - \ln\left(\frac{\mu}{\lambda + \mu}\right)\right] \tag{4.59}$$

Comparing Eq. (4.59) with the exact expression in Table 4.9 is difficult because of the different forms of the equations. The exact and approximate

expressions are compared numerically in Table 4.10. Clearly, the approximations are close to the exact values. The best way to compare availability numbers, since they are all so close to unity, is to compare the differences with the unavailability $1 - A$. Thus, in Table 4.10, the difference in the results for the parallel system is $(0.99990197 - 0.99980396)/(1 - 0.99980396) = 0.49995$, or about 50%. Similarly, for the standby system, the difference in the results is $(0.999950823 - 0.999901)/(1 - 0.999901) = 0.50326$, which is also 50%. For the TMR system, the difference in the results is $(0.999707852 - 0.999417815)/(1 - 0.999417815) = 0.498819$—again, 50%. The reader will note that these results are good approximations, all approximations yield a slightly higher result than the exact value, and all are satisfactory for preliminary calculations. It is recommended that an exact computation be made once a design is chosen; however, these approximations are always useful in checking more exact results obtained from analysis or a computer program.

The foregoing approximations are frequently used in industry. However, it is important to check their accuracy. The first reference known to the author of such approximations appears in Calabro [1962, pp. 136–139].

4.10 MICROCODE-LEVEL REDUNDANCY

One can employ redundancy at the microcode level in a computer. Microcode consists of the elementary instructions that control the CPU or microprocessor—the heart of modern computers. Microinstructions perform such elementary operations as the addition of two numbers, the complement of a number, and shift left or right operations. When one structures the microcode of the computing chip, more than one algorithm can often be used to realize a particular operation. If several equivalent algorithms can be written, each one can serve the same purpose as the independent circuits in the N-modular redundancy. If the algorithms are processed in parallel, there is no reduction in computing speed except for the time to perform a voting algorithm. Of course, if all the algorithms use some of the same elements, and if those elements are faulty, the computations are not independent. One of the earliest works on microinstruction redundancy is Miller [1967].

4.11 ADVANCED VOTING TECHNIQUES

The voting techniques described so far in this chapter have all followed a simple majority voting logic. Many other techniques have been proposed, some of which have been implemented. This section introduces a number of these techniques.

4.11.1 Voting with Lockout

When N-modular redundancy is employed and N is greater than three, additional considerations emerge. Let us consider a 4-level majority voter as an

TABLE 4.10 Comparison of the Exact and Approximate Steady-State Availability Equations for Various Systems

System	Exact, Eq. (4.50)	Approximate, Eqs. (4.54), (4.56), and (4.59)	Exact, $\mu = 100\lambda$	Approximate, $\mu = 100\lambda$
Two in parallel	$\dfrac{\mu(2\lambda + \mu)}{2\lambda^2 + 2\lambda\mu + \mu^2}$	$\dfrac{\mu(2\lambda + \mu)}{(\lambda + \mu)^2}$	0.99980396	0.99990197
Two standby	$\dfrac{\mu(\lambda + \mu)}{\lambda^2 + \lambda\mu + \mu^2}$	$\left(\dfrac{\mu}{\lambda + \mu}\right)\left[1 - \ln\left(\dfrac{\mu}{\lambda + \mu}\right)\right]$	0.999901	0.999950823
TMR	$\dfrac{\mu(3\lambda + \mu)}{6\lambda^2 + 3\lambda\mu + \mu^2}$	$\left(\dfrac{\mu}{\lambda + \mu}\right)^2\left(\dfrac{3\lambda + \mu}{\lambda + \mu}\right)$	0.9994417815	0.999707852