REPAIRABLE SYSTEMS

Repair is the process of returning equipment to its operating state after failure. It may be more illuminating to speak about renewal, rather than repair. Repair of modular equipment may simply be the replacement of failed units by operable ones.

Failures may be of a different nature: failure of an embedded component that needs repair, failure of a replaceable unit that calls for a replacement, or degradation of operational parameter(s) that merely need adjustment (tuning). Repair, in the reliability engineering context, may imply an actual repair or/and the replacement of a failed unit. Henceforth, the two words *repair* and *renewal* will be used interchangeably.

Renewal can be of two main types, depending on whether a failure during system operation is catastrophic. The first type concerns failures that can be removed without catastrophic consequences. The operation itself may have to be interrupted, but it is not considered critical for the system. For instance, a radar system failure makes operation impossible for the time needed for failure detection and correction, but after this it can resume normal operation again. If a redundant unit fails, the system may not even “feel” its replacement/repair. So, this type of repair relates to systems with a continuous (noninterrupted) regime of operation.

The second type of renewal concerns those systems whose operation does not allow any interruption. For instance, an aircraft cannot be repaired in flight. It can have some redundant components giving it a limited ability to continue operation, but in a degraded state. However, this is not repair by common usage. Moreover, a significant failure leads to catastrophic consequences! On the other hand, it is possible to perform extensive repair on the ground and to restore “complete” ability to operate before the next flight.

Of course, the two cases described here do not cover all possible scenarios. An extension of a previous argument would suggest that a spacecraft may not be repaired in flight. However, the *Hubble Space Telescope* was repaired in orbit. We are forced to accept that a system is repairable in some situations and not repairable in others. A comprehensive discussion on reliability of repairable systems can be found in Ascher and Feingold (1).

RELIABILITY INDICES

An important goal in the design and analysis of a repairable system is to improve system reliability. Thus we need precise measures (indices) for reliability.

Reliability indices of repairable systems depend on the following main factors:

- System structure,
- Failure rate (not necessarily constant), and
- Repair modes (possibility of simultaneous repair, repair intensity for each failed unit).

Reliability of repairable systems is analyzed with the help of analytical models or numerical simulation. Among analytical mathematical models, one of the most widely used is Markov-type models.

MARKOV MODELING

In engineering practice, we generally apply Markov models to describe a renewable system mathematically. A Markov model describes the process of probabilistic transition of a system from one state to another. Each system state is characterized by the states of its units: operational, standby, failed, under repair, and the like. The process of state transition is characterized by the Markov property: the evolution of the transition process does not depend on “prehistory.” The practice is justified by the relative simplicity of mathematical expressions and the explicit result it yields.

While using Markov models to describe a repairable system, we implicitly assume that all distributions (time to failure, renewal time) are exponential. This assumption seems to be an obstacle for the wide applicability of Markov models. However, there is some justification for exponential distributions of time to failure, especially for electronic equipment. Of

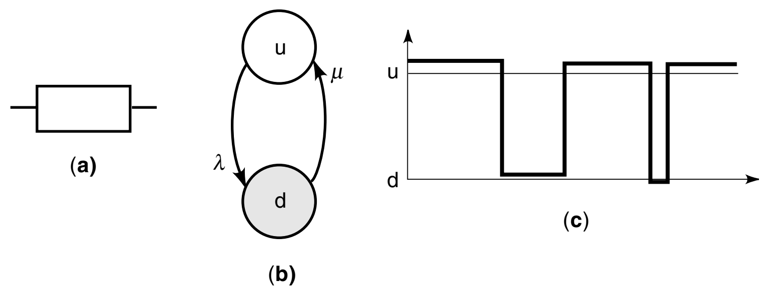

Figure 1. Structure, transition graph, and time diagram for single repairable unit.

course, such an assumption is not correct for the distribution of repair (replacement) time. Nevertheless, Markov models "work" successfully when applied to highly reliable systems.

Applying the Markov model consists of following major steps:

- Precise verbal description and complete enumeration of all possible system states and transitions between these states,
- Choice of system structure and definition of failure criteria,
- Definition of failed states on the basis of chosen failure criteria,
- Construction of the transition graph for the corresponding Markov process,
- Assignment of a transitive intensity to each arc in the transition graph.

Markov Model of Single Unit

The simplest possible repairable system is a single unit. Its (trivial) reliability block diagram is shown in Fig. 1(a). The unit can either be in the up state or in the down state. Failure implies a transition from the up state to the down state. A sample of the transition graph and the time diagram are presented in Fig. 1(b, c). Here state "u" denotes the up state, and "d" the down state. A down (failed) state is differentiated from an up state in a transition diagram by shadowing, as shown in Fig. 1(b).

A simple transition graph like the one described here admits a simple interpretation. At any given moment in time, if the system (unit) is in a given state, it can either stay in that state or make a transition to another state connected to it. Each pair of connected states is characterized by an intensity (time rate) of transition from one state to another. For the transition graph described previously, transition intensity out of state "u" is λ and that out of state "d" is μ. The transition intensities sometimes allow simple physical interpretation. In this context, the reciprocal of λ denotes the unit's MTTF (mean time to failure) and the reciprocal of μ denotes the MTTR (mean time to repair).

Let $P_u(t)$ and $P_d(t)$, respectively, be the probabilities of finding the system in up and down states at time t. It is easy to conclude that $P_u(t)$ is the availability coefficient for the unit. The probability of locating the system in the up state at time $(t + \Delta t)$ can be found from the formula of complete probability

$$P_u(t + \Delta t) = (1 - \lambda \Delta t) P_u(t) + \mu \Delta t P_d(t)$$

This leads to

$$\frac{dP_u(t)}{dt} = -\lambda P_u(t) + \mu P_d(t)$$

when the appropriate limit exists. This is the simplest example of a Kolmogorov equation for a point stochastic process. The differential equation can be solved together with a normalization condition $P_u(t) + P_d(t) = 1$ and a set of initial conditions. Note that the system of equations is consistent with the redundant differential equation for P_d because of the normalization condition, which states that the system is always located somewhere.

The system of linear differential equations arising from such transition graphs can be solved by Laplace transforms or any of the standard methods (2–4). It should be emphasized that stationary coefficients, when they exist, can be obtained without solving the differential equations. If a stationary state exists [i.e., $P_u(t) = P_u$ when $t \to \infty$], $dP_u(t)/dt$ is identically zero. This leads to

$$-\lambda P_u + \mu P_d = 0$$
$$P_u + P_d = 1$$

and

$$P_u = \frac{\mu}{\lambda + \mu}$$

Markov Model for Two Units

Reliability block diagrams for a system consisting of two identical and independent units are shown in Fig. 2(a, b). We can easily recognize these as series and parallel structures. Reliability block diagrams depict the system structure and lead to the failure criteria. However, for reliability analysis of these systems we must know not only their structures but also the regimes of their repair (renewal). We normally assume that the repair process itself is Markovian. Further description about the repair facility, whether limited or unlimited simultaneous repair is possible, is needed to describe the system

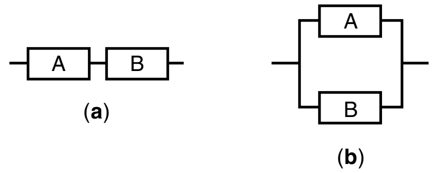

Figure 2. Structures of two unit systems.

fully. Corresponding transition graphs for these structures are given in Fig. 3(a, b).

State 2 denotes that both units are in up state (system up state), state 1 denotes a state with one failed unit (down for series, up for parallel), and state 0 denotes a state with both units failed (system down state). This brings out another point: a transition graph depends on the failure criteria, which itself is related to the system structure. For the simple examples considered previously, both states 0 and 1 are down states for the series system, and only state 0 is the down state for the parallel system.

Having drawn the transition diagrams and chosen the failure criteria, the next task is to assign transition intensities associated with each of these diagrams (which are presented by weights of the arrows). Assume that a single unit has failure rate λ, and a single repair service person restores a failed unit with rate μ. For a system of two identical and independent units, possible choices (λ_2, λ_1, μ_0, μ_1) for the failure rates and repair intensities (arrow weights) in Fig. 3(a, b) are:

- $\lambda_2 = 2\lambda$ means that both units are always operating (active),
- $\lambda_2 = \lambda$ means that a redundant unit is in a "cold" standby regime,
- $\mu_0 = 2\mu$ means that there are two repair service people working simultaneously and independently of each other,
- $\mu_0 = \mu$ means that there is a single repair service person for the entire system,
- Together with $\lambda_1 = \lambda$ and $\mu_1 = \mu$ in every case.

These simple examples give us an opportunity to demonstrate the main factors taken into account for the analysis of renewal systems. The salient features of transition diagrams for repairable system can be summarized again:

- Transition diagrams depend on the failure criteria (network structure: series, parallel).
- Failure intensities λ_i depend on redundancy type (e.g., hot standby).
- Repair intensities μ_j depend on the repair facility (limited or unlimited).

MARKOV MODEL FOR MULTIPLE UNITS

Generalization to a series or parallel system consisting of more than two units is straightforward. For more complex structures, like 'K-out-of-N' ($K \leq N$), we can use special models based on the so-called Birth-and-Death process (2,4,5). Note that the K-out-of-N system is a generalization of series and parallel systems because 1-out-of-N is a parallel system and N-out-of-N is a series system. In general cases, the only possibility is to compile a transition graph that describes the system operation and use numerical methods for obtaining the solution.

Limitations of Markov Modeling

The main disadvantage of applying Markov models in a real-world situation is the implicit assumption of exponential distributions. Although these assumptions are not critical for some commonly quoted (stationary) reliability indices like the availability coefficient or the mean time between failure (MTBF), they are very essential for others like the probability of failure free operation (PFFO) and mean time to failure. The readers are again referred to the article QUANTITATIVE MEASURES OF RELIABILITY for details.

The assumption of exponential distributions may be dropped by adopting a semi-Markov model. Unfortunately, attempts to apply semi-Markov models for reliability problems have not been very fruitful. Lack of appropriate data justifying the use of semi-Markov models, and some unjustified assumptions that still remain in this approach form the main counterarguments. It remains an area of academic interest. A better approach is the application of renewal processes, or point recurrent processes (6,7) and its special class, alternating renewal processes. The latter may be conveniently interpreted in reliability terms as alternating intervals of up and down states.

OTHER MATHEMATICAL AND NUMERICAL APPROACHES

One of the most advanced methodologies in modern reliability theory is the asymptotic analysis of renewal systems (2,4). This approach is grounded on certain limit theorems for point stochastic processes. There are two fundamental asymptotic

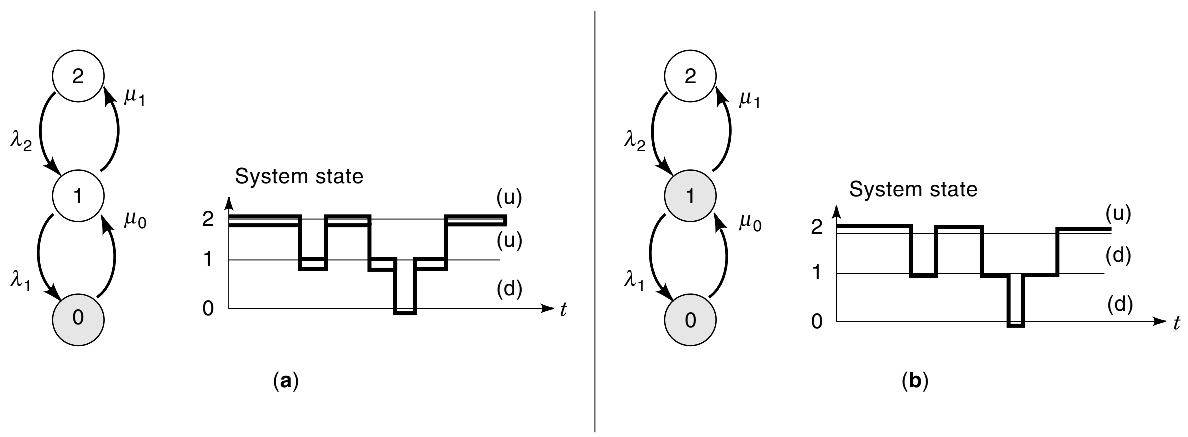

Figure 3. Transition graphs and time diagrams for series and parallel systems.

theorems in point stochastic process theory. The Renyi–Kninchine–Ososkov theorem (8–10) states that the procedure of "thinning" any point process (under a suitable normalization condition) asymptotically leads to a Poisson process. The Grigelionis–Pogozhev theorem (11,12) states that a superposition of point stochastic processes (under some not-so-restrictive conditions) also results in a Poisson process asymptotically. The first theorem is effectively used for reliability analysis of highly reliable redundant systems where system failures are "rare events". The second one is a background for the use of a Poisson process for the description of failure process of multicomponent renewable series systems. Discussion on application of these approaches to reliability theory can be found in Ref. 4. We also find heuristic methods for analyzing renewable systems there.

For a fairly complex renewable systems like communication network, analytical results are difficult to obtain. Monte Carlo simulation can be recommended (5,13) for such systems. For highly reliable systems, whose Monte Carlo modeling takes too much computer time and demands huge computer memory, accelerated methods of modeling have been developed (4) by various authors.

DATA COLLECTION

We need to pay careful attention to collection and analysis of field data. Effective reliability analysis of a repairable system demands developed engineering intuition and experience owing to many details. This is the proven way to move from reliability theory to real-world engineering applications. We remind the readers of the GIGO (Garbage In Garbage Out) principle in a jocular vein!

There are several ways of reliability data collection. Because the availability coefficient of a repairable system is one of the main reliability indices, statistical data collection begin as special tests to confirm the required availability level. Tracking the history of each individual failure is important because system maintenance (spare supply, operation monitoring, preventive maintenance, etc.) are based on current reliability data. For these purposes, reliability data must be supplied with all relevant information: environmental condition at the time of failure, level of loading, regime of its use (hot or cold), and so on. Individual failure report should then be consolidated to obtain statistical summary, which might be used for reliability analysis of newly designed system of a similar type.

There are two main ways to collect reliability data. First is recording the failure history for each type of unit (time between current and previous failure plus additional related information). Statistical inference based on the unit data gives objective information about the units of this type. This information is usually collected by unit vendors. Another type is tracking each repairable unit: from warehouse (as a spare) to installation, then to failure, repair and back to sparing or installation. This records the individual unit behavior placed in a particular set of circumstances and is useful for recognizing possible weak points in the system.

Data on mean time to repair may be obtained from special control experiments or from real usage. The MTTR value is often assigned to a unit on the basis of previous engineering experience.

Formal reliability analysis of a complex system is only as useful as the model and the approximations used. If the ultimate goal of reliability analysis of a repairable system is to ensure some overall reliability threshold for the minimum cost, we need to understand which components are more important in the reliability sense. Sometimes increased component level reliability is more effective than a subsystem redundancy (3). Given a concrete set of objective functions, ingenious analysis and judicious use of redundancy can deliver a reliable (fault tolerant) system with inexpensive and less reliable components. In our opinion, RAID (Redundant Array of Inexpensive Disks) is one such example (14).

BIBLIOGRAPHY

1. H. Ascher and H. Feingold, Repairable system reliability: Modeling, inference, misconceptions and their causes, in D. B. Owen (ed.), *Lecture Notes in Statistics,* Vol. 7, New York: Marcel Dekker, 1984.
2. B. Gnedenko, Yu. Belyaev, and A. Solovyev, *Mathematical Methods in Reliability Theory,* New York: Academic Press, 1969.
3. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing: Probability Models,* New York: Holt, Rinehart and Winston, 1975.
4. B. Gnedenko and I. Ushakov, in J. Falk (ed.), *Probabilistic Reliability Engineering,* New York: Wiley, 1995.
5. I. Ushakov (ed.), *Handbook of Reliability Engineering,* New York: Wiley, 1994.
6. D. R. Cox, *Renewal Theory,* Methuen Monographs on Applied Probability and Statistics, London: Methuen, 1962.
7. W. L. Smith, Renewal theory and its ramifications, *J. Roy. Stat. Soc. Ser. B,* **20** (2): 243–302, 1958.
8. A. Renyi, Poisson-folyamat egy jemllem zese., *Proc. Inst. Math., Hungarian Academy of Sciences,* **1** (4): 11–24, 1956.
9. A. Ya. Khinchin, On Poisson stream of random events, *Theory Probab. Appl.,* **1**: 1–23, 1956.
10. A. Ososkhov, A limit theorem for flows of similar events, *Theory Probab. Appl.,* **1** (2): 89–101, 1956.
11. B. Grigelionis, On the convergence of sums of step stochastic processes to a Poisson process, *Theory Probab. Appl.,* **8** (2): 27–31, 1963.
12. B. Progozhev, Evaluation of deviation of the equipment failure flow from a Poisson process, in B. N. Bruevich and B. Gnedenko (eds.), *Cybernetics in the Service of Communism,* Vol. 2 (in Russian), Moscow: Energia, 1964.
13. B. Gnedenko, I. Pavlov, and I. Ushakov, in S. Chakravarty (ed.), *Statistical Reliability Engineering,* New York: Wiley (forthcoming).
14. D. Patterson, G. Gibson, and R. Katz, *A Case for Redundant Arrays of Inexpensive Disks (RAID),* University of Berkeley Report no. UCB/CSD 87/391.

Igor Ushakov
Sumantra Chakravarty
QUALCOMM, Inc.