Module PE.PAS.U16.5

Markov models for reliability analysis

U16.1 Introduction

We have seen in modules U11 and U12 methods of analysis for nonrepairable and repairable components, respectively, while modules U14 and U15 provided methods of analysis for nonrepairable systems. Markov models provide us the most general method of all, applicable to nonrepairable and repairable components as well as nonrepairable and repairable systems. It is especially with respect to repairable systems that the method becomes attractive as no other method deals with this type of system with the same degree of effectiveness and simplicity.

The reader would do well to review Section U12.1 on random processes in module U12 before proceeding. Here, we remind the reader that a random process is a collection of random variables indexed by a parameter (typically time) such that the random variables are ordered in a particular sequence. We recall that the indexing parameter may be discrete, resulting in a discrete-time process, or continuous, resulting in a continuous-time process. In addition, the state space, i.e., the values assumed by the random variables comprising the process, may be discrete, resulting in a discrete-state process, or continuous, resulting in a continuous-state process. Formal terminology exists which relate to Markov processes, as follows [1].

1. Discrete-time Markov chain: a discrete-time/discrete state Markov process.
2. Continuous-time Markov chain: a continuous-time/discrete state Markov process.
3. Discrete-time Markov process: a discrete-time/continuous state Markov process.

4. Continuous-time Markov process: a continuous-time/continuous state Markov process.

In this module, we only deal with #2 of the above, and we therefore use the term "Markov chain" to refer to it. An implication here is that we only study Markov processes that have discrete states as this is the approach that taken in the development of most power system reliability evaluation procedures.

U16.2 Markov properties

The formal definition of a Markov chain is as follows [2,3]:

Definition: The random process {X(t), t ≥0} is a continuous-time Markov chain if for all s≥0, t ≥0 and nonnegative integers i, j, x(u), 0≤u<s,

$$P[X(t+s) = j \mid X(s) = i, \quad X(u) = x(u), \quad 0 \le u < s]$$
$$= P[X(t+s) = j \mid X(s) = i]$$

Interpretation: Assume that s represents the *present* of the random process. Then:

- the conditioning event in the above definition

  $$X(s)=i, \quad X(u)=x(u), \quad 0\le u<s$$

  expresses the present and past of the random process.
- the left-hand-side is the conditional probability of the "future" random variable X(t+s)

Thus, the definition indicates that the conditional distribution of the "future" X(t+s) given the present X(s) and the past X(u) depends only on the present and is independent of the past.

This means that the present "summarizes" the entire history of the process, i.e., all of the information contained in the values taken by the random variables of the past are contained in the random variable of the present. Thus, we say that a Markov process is a "memoryless" process.

Example [4]: Consider that the demand for electric power monitored at a low-voltage bus of a transformer station on any given day can be classified as either high or low (1 or 0). It has been observed that if the demand is high on a certain day, the probability that it will be high the next day is 0.75. If the demand is low, on the other hand, the probability that it will be low the next day is 0.5. Note: the probabilities depend on only today's demand, and not yesterday's demand. Therefore, the process describing the state of the demand (1 or 0) from day-to-day is Markov.

A key concept in dealing with Markov processes is the notion of states. In general, a Markov process may have any number of states. For example, it is typical in determining appropriate maintenance intervals to model component states based on how much deterioration the component has incurred. Different approaches here include:

- Two states: S1 (working) and S2 (failed)
- Three states: S1 (working well), S2 (failed), S3 (working with deterioration)
- Four states: S1 (working well) S2 (failed), S3(working with minor deterioration), S4 (working with major deterioration)

Since many deterioration processes for most components are gradual, it is clear that we may have a large number of deterioration states. In addition, we may also provide for a number of states associated with different levels of maintenance.

We note that, in general, the recognition of different states implies that the component may reside in any of them. Therefore, it is possible that the component, while residing in one state, may transition to another state. If we consider a certain time t and a certain time interval Δt, then there is a probability for each pair of states for which a transition is possible (including the pairs

comprised of two identical states). We call this probability the transition probability.

For states j and k, we denote this probability as $p_{jk}(t,\Delta t)$ such that:

$$p_{jk}(t,\Delta t) = P[X(t+\Delta t) = k \mid X(t) = j]$$

Fig. U16.1 illustrates a 4-state Markov model together with the transition probabilities. The fact that no transitions are shown between states 1 and 4 or between states 2 and 3 indicates that the transition probabilities for these states are zero. In the diagram, the dependence of the transition probabilities on t and Δt is implied.

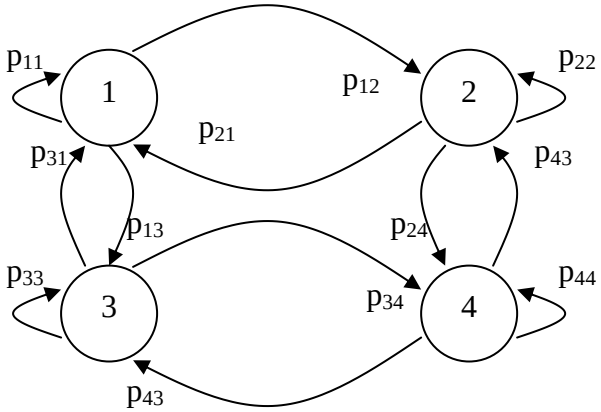

Fig. U16.1: Illustration of 4-state Markov model

If the probability $P[X(t+s) = j \mid X(s) = i]$ is independent of s (implying that transition probabilities are the same no matter what "present" time s we choose, then the Markov chain is said to be *stationary* or *homogeneous*.

Although stationary Markov chains have transition probabilities independent of time, the transition probabilities do depend on the *time interval* of interest. Thus, for stationary Markov processes, it is appropriate to denote the transition probabilities as $p_{jk}(\Delta t)$ rather than $p_{jk}(t,\Delta t)$.

We consider only stationary Markov processes in our treatment.

Can we draw any conclusions regarding the distribution associated with the “times to transition”?

Given that there are possibly a number of other states to which we may transition, it is appropriate to think of the transition time as the amount of time it stays in a certain state before transiting to another state.

Let’s call these times the transition times, and denote the transition time from state j as T_j. How is T_j distributed?

To answer this question [3], suppose that a Markov chain enters state j at some time, say time t=0, and suppose that the process does not leave state j (that is, a transition does not occur) during the interval (0,10) minutes. What is the probability that the process will not leave state j during the 5 minutes after the 10 minutes is up, i.e, during the interval (10,15)?

Now since we know that the process is in state j at time t=10, by the memoryless property, we also know the probability that it remains in state j during the first 5 minutes is the same as the probability that it remained in state j during the interval (0,5), i.e.,

$$P(T_j > 15 \mid T_j > 10) = P(T_j > 5 \mid T_j > 0) = P(T_j > 5)$$

And the same basic reasoning leads us to conclude:

$$P(T_j > s + t \mid T_j > s) = P(T_j > t)$$

for all s≥0, t≥0, i.e., the distributions on transition times depend only on the time interval but not on the time itself. We have seen this kind of property before with what we called time to failure (in the case of nonrepairable components) and interevent times (in the case of repairable components), and we know that the only distribution which provides this property for a random variable is the exponential distribution (see module 10).

Thus, all transition times for a Markov process are exponentially distributed!

In other words [3], a Markov chain is a random process that moves from state to state such that the amount of time it spends in each state, before proceeding to the next state, is exponentially distributed.

A final word on Markov chains… the states must be mutually exclusive, i.e., the process cannot reside in two or more states at the same time. This should be self-evident from all discussion so far.

U16.3 Relation to Poisson processes

Module U12 provides an introduction to random (stochastic) processes, and we saw that the renewal process was one example, where all interevent times are identically and independently distributed (but with arbitrary distribution). We also saw that the Poisson process is a special case of the renewal process where the interevent times are exponentially distributed.

Noting our conclusion in the last section that a Markov process has exponentially distributed transition times, it is natural to inquire about the relationship between a Markov and a Poisson process. We make the following observations:

- A Poisson process is a counting process, i.e., the "state" of a Poisson process at a particular time is characterized by the number of events that have occurred up until that time. Thus, we see that a Poisson process has an infinite number of states 0, 1, 2, …k,…, ∞, and that any particular transition always takes the process from one state k to the next state k+1. Further, the time between transitions must be exponential.
- A Markov process may transition from any state to any other state and is not constrained to only transition from state k to state k+1, and the time between transitions must be exponential.

So we infer from the above that a Poisson process is a special case of a Markov process, such that a Markov process will be Poisson if

the transitions are constrained to always occur from state k to state k+1.

U16.4 Solving for state probabilities

Examine Fig. U16.1 again, and note that the indicated probabilities are those of transiting from one state to another in the next time interval Δt. What are typically of interest, however, are the probabilities that the random process will be in any particular state at a give time t. We denote these *state probabilities* at $p_j(t)$, contained in the row vector:

$$\underline{p}(t) = \begin{bmatrix} p_1(t) & \cdots & p_n(t) \end{bmatrix}$$

where, for any particular time t, it must be the case that:

$$\sum_{j=1}^{n} p_j(t) = 1$$

We aim to obtain these probabilities in this section, in four steps, by developing:

1. Transition intensities
2. Two system matrices
3. The system differential equation
4. The solution procedure

Our development in this section is adapted from [5].

U16.4.1 Transition intensities

We indicated above that for stationary Markov processes, the transition probability is a function of the time interval of interest. If we select the time interval Δt to be very small, then we can assume that the transition probability is a linear function of the time interval, i.e.,

$$p_{jk}(\Delta t) = P[X(t+\Delta t) = k \mid X(t) = j] \approx \lambda_{jk}\Delta t \qquad \text{(U16.1)}$$

where the constant of proportionality is λ_{jk} and is referred to as the state j to state k transition intensity,

defined as:

$$\lambda_{jk} = \lim_{\Delta t \to 0} \frac{p_{jk}(\Delta t)}{\Delta t}, \quad j \neq k \qquad \text{(U16.2)}$$

(The transition intensity from a working state to a failed state is equivalent to the hazard function used in modules U11 and U12). Likewise, we may consider $p_{jj}(\Delta t)$, the probability that the random process will remain in state j (no transition) within the next time interval Δt. This is given by:

$$p_{jj}(\Delta t) = P[X(t+\Delta t) = j \mid X(t) = j]$$

Now it is tempting, by analogy, to equate this to $\lambda_{jj}\Delta t$, in terms of the "transition intensity from state j to state j." However, it will prove more convenient later to define the complement, i.e., the "transition intensity from state j to any other state k, $k \neq j$." We denote this as λ_j, where, for small Δt, $\lambda_j \Delta t \approx 1 - p_{jj}(\Delta t)$, so that

$$p_{jj}(\Delta t) = P[X(t+\Delta t) = j \mid X(t) = j] \approx 1 - \lambda_j \Delta t \qquad \text{(U16.3)}$$

where the constant of proportionality is the

state j transition intensity,

defined as:

$$\lambda_j = \lim_{\Delta t \to 0} \frac{1 - p_{jj}(\Delta t)}{\Delta t} \qquad \text{(U16.4)}$$

For stationary Markov processes, to which we have confined ourselves, the transition intensities are constant for all time and may therefore be called transition rates as well, providing the clear indication that

- λ_{jk} may be interpreted as the expected number of transitions per unit time from state j to state k, and
- λ_j may be interpreted as the expected number of transitions per unit time from state j to any other state.

Now consider that the random process is in state j at some particular moment in time. In the next time interval Δt, there are two kinds of events that can occur. Either the process will not transition or it will transition to one of the other states, so that if we add the probabilities of all possible events, we obtain 1, i.e.,

$$p_{jj}(\Delta t) + \sum_{j \neq k} p_{jk}(\Delta t) = 1$$

$$\rightarrow 1 - p_{jj}(\Delta t) = \sum_{j \neq k} p_{jk}(\Delta t)$$

Substitution into (U16.4) results in:

$$\lambda_j = \lim_{\Delta t \to 0} \frac{\sum_{j \neq k} p_{jk}(\Delta t)}{\Delta t}$$

But by equation (U16.2), each of the terms in the above summation, when divided by Δt, results in λ_{jk}, so that

$$\lambda_j = \sum_{j \neq k} \lambda_{jk} \qquad \text{(U16.5)}$$

U16.4.2 Two system matrices

The development of section U16.4.1 provides that we may specify two matrices in terms of the two different sets of parameters identified.

The *stochastic transitional probability matrix* (according to Billinton [6]) or the *transition probability matrix* (according to Endrenyi [5]) is comprised of what we have called the transition probabilities $p_{jk}(\Delta t)$, given by:

$$\underline{P}(\Delta t) = \begin{bmatrix} p_{11}(\Delta t) & \cdots & p_{1n}(\Delta t) \\ \vdots & \vdots\cdots & \vdots \\ p_{n1}(\Delta t) & \cdots & p_{nn}(\Delta t) \end{bmatrix}$$

Note that the sum of the row elements is 1. This matrix is useful for determining the state probabilities at time $(t+\Delta t)$ if we know the state probabilities at time t. We can see this by observing that the probability of being in state j at time $(t+\Delta t)$ is equal to

- The probability of:
  - being in state j at time t and
  - not making a transition to any other state in Δt

  This is $p_j(t)p_{jj}(\Delta t)$
- Plus the probability of being in any state k at time t and transiting to state j in Δt. This is $\sum_{j\neq k} p_k(t)p_{kj}(\Delta t)$.

Thus, we see that:

$$p_j(t+\Delta t) = p_j(t)\,p_{jj}(\Delta t) + \sum_{j\neq k} p_k(t)\,p_{kj}(\Delta t) \quad \text{(U16.6)}$$

We can write this in matrix form as

$$\underline{p}(t+\Delta t) = \underline{p}(t)\underline{P}(\Delta t) \quad \text{(U16.7)}$$

We call the above relation the state transition relation.

Example: Consider the state diagram for a single repairable component, as illustrated in Fig. U16.2, where state 1 is working and state 2 is failed (and being repaired). Write the state transition

relation for this state diagram and express it in terms of the transition intensities.

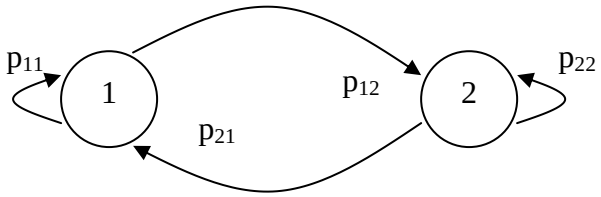

Fig. U16.2: Two state system

The state transition relation is given by:

$$\begin{bmatrix} p_1(t+\Delta t) & p_2(t+\Delta t) \end{bmatrix} = \begin{bmatrix} p_1(t) & p_2(t) \end{bmatrix} \begin{bmatrix} p_{11}(\Delta t) & p_{12}(\Delta t) \\ p_{21}(\Delta t) & p_{22}(\Delta t) \end{bmatrix}$$

Using (U16.1) and (U16.3) (which indicate that $p_{jk}(\Delta t) \approx \lambda_{jk}\Delta t$ and $p_{jj}(\Delta t) \approx 1-\lambda_j\Delta t$), we may express the above in terms of the state transition intensities as:

$$\begin{bmatrix} p_1(t+\Delta t) & p_2(t+\Delta t) \end{bmatrix} = \begin{bmatrix} p_1(t) & p_2(t) \end{bmatrix} \begin{bmatrix} 1-\lambda_1\Delta t & \lambda_{12}\Delta t \\ \lambda_{21}\Delta t & 1-\lambda_2\Delta t \end{bmatrix} \quad \text{(U16.8)}$$

The second system matrix that we may specify, comprised of the transition intensities λ_{jk} and λ_j, is called the *transition intensity matrix*, given by:

$$\underline{A} = \begin{bmatrix} -\lambda_1 & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & -\lambda_2 & \cdots & \lambda_{2n} \\ \vdots & \vdots\cdots & \vdots\cdots & \vdots \\ \lambda_{n1} & \cdots & \cdots & \lambda_{nn} \end{bmatrix}$$

From (U16.5), we see that the elements in any row of $\underline{A}$ must add to zero. This means that the determinant of $\underline{A}$ is zero, and thus the rows of $\underline{A}$ are not independent.

Because the elements of the transition intensity matrix A are given by $p_{jk}(\Delta t) \approx \lambda_{jk}\Delta t$ and $p_{jj}(\Delta t) \approx 1-\lambda_j\Delta t$ for small Δt, we can say that, whereas the transition probability matrix P describes the behavior of the random process in arbitrary time intervals t, the transition intensity matrix A describes the behavior of the random process for very small intervals of time [2].

Example: Obtain the transition intensity matrix for the two-state model of Fig. U16.2.

$$\underline{A} = \begin{bmatrix} -\lambda_1 & \lambda_{12} \\ \lambda_{21} & -\lambda_2 \end{bmatrix} \qquad \text{(U16.9)}$$

Note that the determinant is $\lambda_1\lambda_2-\lambda_{12}\lambda_{21}$, but because $\lambda_1 = \lambda_{12}$ and $\lambda_2 = \lambda_{21}$ (since there is only one other state to which a state may transition), we see that the determinant is zero.

The transition intensity matrix is related to the transition probability matrix according to:

$$\underline{A} = \lim_{\Delta t \to 0} \frac{\underline{P}(\Delta t) - \underline{I}}{\Delta t}$$

where I is the (square) identity matrix of dimension equal to that of A and P(Δt). The above relation also enables expression of P(Δt) in terms of A, if Δt is small, according to:

$$\underline{P}(\Delta t) = \underline{A}\Delta t + \underline{I}$$

We will see in the next subsection that the transition intensity matrix is important in setting up the system differential equations.

U16.4.3 The system differential equations

Reconsider (U16.6), repeated below for convenience:

$$p_j(t+\Delta t) = p_j(t)\,p_{jj}(\Delta t) + \sum_{j\neq k} p_k(t)\,p_{kj}(\Delta t)$$

Note that it is written in terms of the transition probabilities. Let's rewrite it in terms of transition intensities by using (U16.1) and (U16.3) (which indicate that $p_{jk}(\Delta t)\approx\lambda_{jk}\Delta t$ and $p_{jj}(\Delta t)\approx 1-\lambda_j\Delta t$):

$$p_j(t+\Delta t) = p_j(t)[1-\lambda_j\Delta t] + \sum_{j\neq k} p_k(t)\lambda_{kj}\Delta t$$

Bringing $p_j(t)$ over to the left-hand-side, and then dividing both sides by Δt, we obtain:

$$\frac{p_j(t+\Delta t)-p_j(t)}{\Delta t} = -p_j(t)\lambda_j + \sum_{j\neq k} p_k(t)\lambda_{kj}$$

In the limit as Δt goes to zero, we recognize the left-hand-side as a derivative, so that:

$$\dot{p}_j(t) = -p_j(t)\lambda_j + \sum_{j\neq k} p_k(t)\lambda_{kj}$$

We recognize the right-hand-side as the j^{th} row of a matrix product consisting of the row vector $\underline{p}(t)$ and the transition intensity matrix $\underline{A}$. Therefore, we have that:

$$\underline{\dot{p}}(t) = \underline{p}(t)\underline{A} \qquad \text{(U16.10)}$$

This is a first-order differential equation. If we were to discretize this equation so as to integrate it numerically on a computer, we would find that the value at step (k+1) (i.e., $t=(k+1)\Delta t$) depends only on the value at step k (i.e., $t=k\Delta t$). This is a property of all first order differential equations, and it is no coincidence, because it also reflects that the system is Markovian! A reasonable conclusion that we may draw here is that all Markovian systems

may be modeled using first order differential equations, a conclusion that is analytically pleasing.

U16.4.4 Solution procedure

As indicated at the end of the previous subsection, we may obtain the solution to (U16.10) through numerical integration on a computer. Alternatively, we may solve it directly. LaPlace Transforms provides a convenient tool for this purpose.

Taking the LaPlace Transform of (U16.10), we obtain:

$$\underline{\tilde{p}}(s)s - \underline{p}(0) = \underline{\tilde{p}}(s)\underline{A} \qquad \text{(U16.11)}$$

where the tilde over a function represents the LaPlace transform of that function. Note that in (U16.11) that $\underline{p}(0)$ represents the initial condition of the system and is obtained by assuming that the random process is in one particular state k, and not in any other state $j \neq k$, such that $p_k(0)=1.0$ and $p_j(0)=0 \ \forall \ j \neq k$.

Rearranging (U16.10), we have:

$$\underline{\tilde{p}}(s)s\underline{I} - \underline{\tilde{p}}(s)\underline{A} = \underline{p}(0)$$

where the insertion of the identity matrix $\underline{I}$,

$$\underline{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots\cdots & \vdots\cdots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}$$

does not change the first term but shows clearly the transition to the next step, which is to factor according to:

$$\underline{\tilde{p}}(s)\left[s\underline{I} - \underline{A}\right] = \underline{p}(0)$$

Now if we post-multiply both sides by [s$\underline{I}$-$\underline{A}$]$^{-1}$, we get:

$$\underline{\tilde{p}}(s) = \underline{p}(0)\left[s\underline{I} - \underline{A}\right]^{-1} \qquad \text{(U16.12)}$$

and we may use tables of LaPlace transforms to obtain $\underline{p}$(t) from (U16.12).

Remark: Recall that we said that $\underline{A}$ is singular, which means that it has a zero determinant, and as a result, it cannot be inverted. However, this does not mean that the matrix [s$\underline{I}$-$\underline{A}$] is also singular. In fact, [s$\underline{I}$-$\underline{A}$] is a matrix with properties that depend on the LaPlace variable s, and therefore its singularity (or nonsingularity) depends on the value of s. In system theory, the values of s that make the matrix [s$\underline{I}$-$\underline{A}$] singular are called the eigenvalues of the system characterized by the differential equation (U16.10). When $\underline{A}$ is singular, as it is in our case, it implies that the system has a zero eigenvalue. A system having a zero eigenvalue will have a constant term in the time-domain expressions of the state variables, which in our case are the elements of $\underline{p}$(t). For stable systems (systems for which the time-domain expressions of the state variables are bounded from above and below as time increases), this constant term is the steady-state, or long-term value. In Markov modeling, we call these steady-state values the *long-run* [5] or the *limiting state* [6] probabilities. Random processes for which long-run probabilities exist are said to be *ergodic*.

Example: For the two-state system illustrated in Fig. U16.2, determine the state probabilities assuming that the system is initially working (that is, it is initially in state 1).

The fact that the system is initially working means that $\underline{p}$(0)=[1 0]. The transition intensity is given by (U16.9), repeated here for convenience:

$$\underline{A} = \begin{bmatrix} -\lambda_1 & \lambda_{12} \\ \lambda_{21} & -\lambda_2 \end{bmatrix}$$

Then, (U16.12) becomes:

$$\underline{\tilde{p}}(s) = \begin{bmatrix} 1 & 0 \end{bmatrix} \left[s \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -\lambda_1 & \lambda_{12} \\ \lambda_{21} & -\lambda_2 \end{bmatrix} \right]^{-1}$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s+\lambda_1 & -\lambda_{12} \\ -\lambda_{21} & s+\lambda_2 \end{bmatrix}^{-1}$$

(U16.13)

The inversion operation is carried out as usual for a 2×2 matrix, according to:

$$\begin{bmatrix} s+\lambda_1 & -\lambda_{12} \\ -\lambda_{21} & s+\lambda_2 \end{bmatrix}^{-1}$$

$$= \frac{1}{(s+\lambda_1)(s+\lambda_2) - \lambda_{12}\lambda_{21}} \begin{bmatrix} s+\lambda_2 & \lambda_{12} \\ \lambda_{21} & s+\lambda_1 \end{bmatrix}$$

Recalling for this two-state model that $\lambda_1 = \lambda_{12}$ and $\lambda_2 = \lambda_{21}$, the inverse becomes:

$$\frac{1}{s^2 + s(\lambda_1 + \lambda_2) + \lambda_1\lambda_2 - \lambda_{12}\lambda_{21}} \begin{bmatrix} s + \lambda_{21} & \lambda_{12} \\ \lambda_{21} & s + \lambda_{12} \end{bmatrix}$$

$$= \frac{1}{s^2 + s(\lambda_{12} + \lambda_{21})} \begin{bmatrix} s + \lambda_{21} & \lambda_{12} \\ \lambda_{21} & s + \lambda_{12} \end{bmatrix}$$

$$= \frac{1}{s(s + \lambda_{12} + \lambda_{21})} \begin{bmatrix} s + \lambda_{21} & \lambda_{12} \\ \lambda_{21} & s + \lambda_{12} \end{bmatrix}$$

and (U16.13) becomes

$$\underline{\tilde{p}}(s) = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix}}{s(s + \lambda_{12} + \lambda_{21})} \begin{bmatrix} s + \lambda_{21} & \lambda_{12} \\ \lambda_{21} & s + \lambda_{12} \end{bmatrix}$$

(U16.14)

The individual state probabilities then become:

$$\tilde{p}_1(s) = \frac{s + \lambda_{21}}{s(s + \lambda_{12} + \lambda_{21})} \qquad \text{(U16.15)}$$

and

$$\tilde{p}_2(s) = \frac{\lambda_{12}}{s(s + \lambda_{12} + \lambda_{21})} \qquad \text{(U16.16)}$$

Using partial fraction expansion, we obtain:

$$\tilde{p}_1(s) = \frac{\frac{\lambda_{21}}{\lambda_{12}+\lambda_{21}}}{s} + \frac{\frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}}}{s+\lambda_{12}+\lambda_{21}}$$

$$\tilde{p}_2(s) = \frac{\frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}}}{s} - \frac{\frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}}}{s+\lambda_{12}+\lambda_{21}}$$

Taking the inverse LaPlace transform of the above, we get:

$$p_1(t) = \frac{\lambda_{21}}{\lambda_{12}+\lambda_{21}} + \frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}} e^{-(\lambda_{12}+\lambda_{21})t} \qquad \text{(U16.17)}$$

$$p_2(t) = \frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}} - \frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}} e^{-(\lambda_{12}+\lambda_{21})t} \qquad \text{(U16.18)}$$

We may obtain the long-run or limiting state probabilities (steady-state values of $p_1(t)$ and $p_2(t)$) if we let $t \to \infty$. This results in:

$$p_{1\infty} = \frac{\lambda_{21}}{\lambda_{12}+\lambda_{21}} \qquad \text{(U16.19)}$$

$$p_{2\infty} = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}} \qquad \text{(U16.20)}$$

Figure U16.3 illustrates the time-domain response for the 2-state system [7].

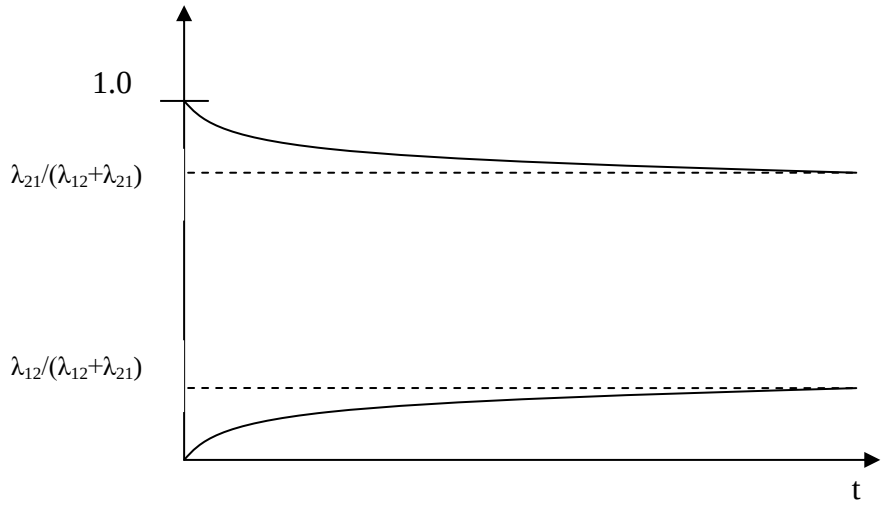

Fig. U16.3: Transient Response [7]

It is of interest to compare the above results with Example U12.2 in module 2, where we used an alternating renewal process to characterize a repairable component, with failure time and repair time exponentially distributed with parameters λ and μ, respectively. In that case, we found that the system availability was given by:

$$A(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t} \qquad \text{(U16.21)}$$

In comparing the two results, note that:

- The "working" state of Example U12.2 is the same as the state 1 here, and the "failed" state of Example U12.2 is the same as the state 2 here.
- The time to failure of Example U12.2 is the same as the time to transition from state 1 to state 2, and the time to repair is the same as the time to transition from state 2 to state 1.
- The parameter λ of Example U12.2 characterizes the exponentially distributed failure time, which is characterized by λ_{12} here. The parameter μ of Example U12.2 characterizes the exponentially distributed repair time, which is characterized by λ_{21} here.

Making appropriate substitutions into U16.22, we have:

$$A(t) = \frac{\lambda_{21}}{\lambda_{12} + \lambda_{21}} + \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}} e^{-(\lambda_{12} + \lambda_{21})t}$$

Comparing the above with (U16.17) and (U16.18), we see that it is identical to (U16.17), which gives the probability of being in state 1 at any time t. Since state 1 is the working state, its probability is interpreted as the probability that the component is working, which is exactly the definition of availability.

It is consoling to see that the Markov approach to analyzing a repairable component agrees with the approach based on Renewal theory. However, it raises the question: Why develop duplicate methods to do the same thing? We will see the answer to this question in the next section.

U16.5 Obtaining long-run probabilities

In many cases, it is of interest to obtain only the long-run probabilities. In such cases, it is unnecessary to solve the differential equation of (U16.10), which is

$$\underline{\dot{p}}(t) = \underline{p}(t)\underline{A}$$

because a long-run probability is a steady-state value and therefore we know in advance that all derivatives are zero, i.e.,

$$\underline{0} = \underline{p}_{\infty}\underline{A} \qquad \text{(U16.22)}$$

where $\underline{p}_{\infty} = \underline{p}(t=\infty)$. Therefore, the problem of obtaining the long-run probabilities requires that we merely solve linear algebraic equations rather than differential equations. This makes us very happy ☺.

Let's try it on our example.

Example: Obtain the long-run probabilities for the two-state model by solving the algebraic equations.

Equation (U16.22) is:

$$\begin{bmatrix}0 & 0\end{bmatrix} = \begin{bmatrix}p_{1\infty} & p_{2\infty}\end{bmatrix}\begin{bmatrix}-\lambda_{12} & \lambda_{12} \\ \lambda_{21} & -\lambda_{21}\end{bmatrix}$$

Extracting the equations yields:

$$0 = -p_{1\infty}\lambda_{12} + p_{2\infty}\lambda_{21}$$

$$0 = p_{1\infty}\lambda_{12} - p_{2\infty}\lambda_{21}$$

Here, we see that we have a problem, as multiplying either equation by -1 results in two identical equations, and given that there are two unknowns, the system of equations is under-constrained, and therefore there is no unique solution. In fact, if we know our linear system theory, we would have seen this coming since we have already recognized that $\underline{A}$ is singular.

But all is not lost. The issue is quickly resolved using common sense. There are only 2 possible outcomes for any particular time t: either the random process is in state 1 or it is in state 2. There are no other possibilities, and this is also true when t=∞. Therefore, we have $p_{1,\infty}+ p_{2,\infty}=1$, giving us another linear equation which is independent from any of the (in this case just one) equations from (U16.22). Therefore, we can write:

$$0 = -p_{1_\infty}\lambda_{12} + p_{2_\infty}\lambda_{21}$$

$$1 = p_{1_\infty} + p_{2_\infty}$$

or

$$\begin{bmatrix}0 & 1\end{bmatrix} = \begin{bmatrix}p_{1_\infty} & p_{2_\infty}\end{bmatrix}\begin{bmatrix}-\lambda_{12} & 1\\ \lambda_{21} & 1\end{bmatrix}$$

Post-multiplying both sides by the inversion of the 2×2 matrix results in:

$$\frac{-\begin{bmatrix}0 & 1\end{bmatrix}}{\lambda_{12}+\lambda_{21}}\begin{bmatrix}1 & -1\\ -\lambda_{21} & -\lambda_{12}\end{bmatrix} = \begin{bmatrix}p_{1_\infty} & p_{2_\infty}\end{bmatrix}$$

which is:

$$\begin{bmatrix}p_{1_\infty} & p_{2_\infty}\end{bmatrix} = \begin{bmatrix}\frac{\lambda_{21}}{\lambda_{12}+\lambda_{21}} & \frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}}\end{bmatrix}$$

The LaPlace transform that we have applied in the example provided so far is useful for Markovian systems that do not have a large number of states. However, the power of the Markov modeling approach is that it can handle even very large systems quite well, as ultimately, all that is required is the solution of a set

of first-order differential equations, assuming the time response is desired. If all that we desire is the steady-state response, then we need only deploy an algebraic solver for linear systems. Here, for example, the LU decomposition approach commonly used in large power system analysis computer packages is a very appropriate tool.

U16.6 State Frequencies and Durations

The material from this section is adapted from [5]. Two parameters of any Markovian system that are often of interest are the state frequencies and the state durations.

- State frequency: We denote the frequency of state j as f_j. This is the expected number of stays in (or arrivals into, or departures from) state j per unit time, computed over a long period.
- State duration: We denote the duration of state j as T_j. This is the expected amount of time per stay the random process is in state j, computed over a long period.

By these definitions, both the state frequency and the state duration are means taken over the long-term behavior of the process.

16.6.1 Frequencies and Durations for the 2-state model

These notions are illustrated in Fig. U16.4 [7] for a 2-state system.

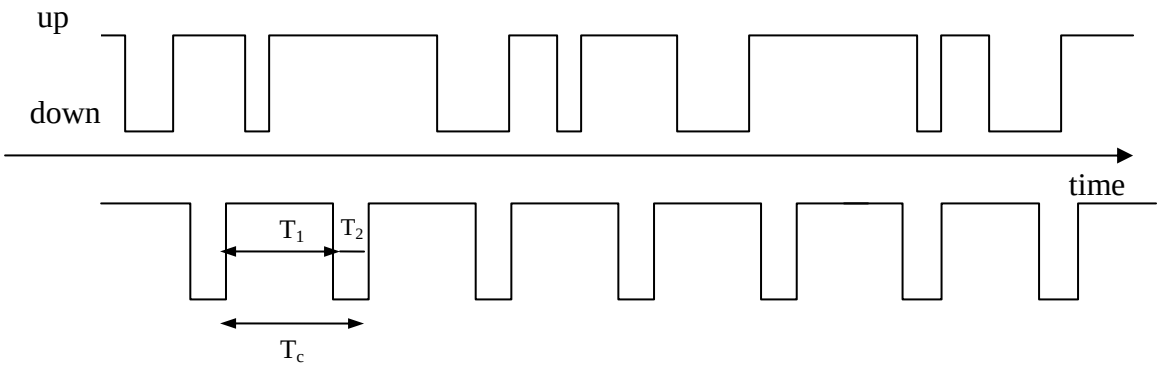

Fig. U16.4: History of a 2-state process [7]

Fig. U16.4, the upper history is actual, whereas the lower history is one in which all up-times T_1 are the same and all down-times T_2 are the same, implying that all cycle times T_c are also the same throughout the history of the process, and are equal to the mean up-time, mean down-time, and mean cycle time, respectively, of the actual process history. These times are precisely the state durations for this 2-state process.

In obtaining the corresponding frequencies, one should note first that the frequency of a state is the number of times of visiting that state in a certain time period divided by the amount of time in that period (where the time period is taken from the beginning of an up-state to the end of a down state). In Fig. U16.4, we could just count the number of up states (or down states) in the total time interval, which would be 7), and divide by the amount of time (which is not shown). This method may be applied to the upper diagram (actual history) or to the lower diagram, as long as the time interval is long enough.

It is important to note that the result is the same independent of whether we count up states or down states.

But an easier, and equivalent method is to count the number of up (or down) states in a single cycle (which is 1) on the lower diagram and divide by the cycle time T_c, i.e., $1/T_c$. Clearly, we obtain the same answer for frequency of state 1 and for frequency of state 2, i.e.,

$$f_1=f_2=1/T_c \qquad \text{(U16.23)}$$

Consider $f_1=1/T_c$, and multiply both sides by T_1, resulting in

$$T_1f_1=T_1/T_c$$

The right-hand-side is the ratio of the mean duration in state 1 to the mean cycle time, and we recognize this ratio as the long-run probability that the process is in state 1, $p_{1,\infty}$. Therefore,

$$T_1f_1=p_{1,\infty}$$

$$\rightarrow f_1=p_{1,\infty}/T_1 \qquad \text{(U16.24)}$$

Likewise,

$$f_2=p_{2,\infty}/T_2 \qquad \text{(U16.25)}$$

16.6.2 Frequencies and durations for the general case

Equations (U16.24) and (U16.25) suggest that we can compute frequencies based on

$$f_j=p_{j,\infty}/T_j \qquad \text{(U16.26)}$$

But our development was based on a 2-state model. Is it reasonable to infer the general result?

This question is easily answered if we think of any general Markov model as being comprised of a single state, call it state j, and all other states. The conceptualization of this model is given in Fig. U16.5 [5].

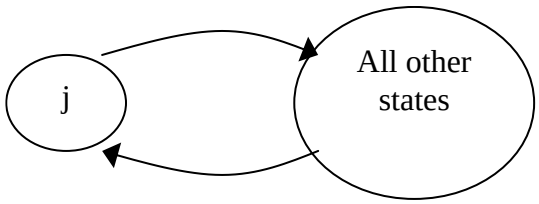

Fig. U16.5: Any Markov model conceptualized as 2 states [5]

Denote f_j and T_j as the frequency and duration for state j, respectively. We also use T_a as the duration for all other states. Then the cycle time is $T_c=T_j+T_a$.

We now have a situation that is described in precisely the same terms as our simple 2-state model, and the same kind of analysis follows, i.e.,

$$f_j=1/T_c \qquad \text{(U16.27)}$$

$$T_j\ f_j=T_j/T_c$$

$$T_j\ f_j=p_{j,\infty}$$

$$f_j= p_{j,\infty}/T_j \qquad \text{(U16.28)}$$

16.6.3 Relation of frequencies and durations to transition intensities

We desire to relate the frequencies and durations to the transition intensities, as they represent indices that decision-makers often use.

To do this, we introduce the concept of frequency of transfer from state j to state k. This frequency, f_{jk}, is defined as the expected number of direct transfers from state j to state k per unit time.

From this definition, and noting that the expected number of transfers in a time interval is equal to the probability of 1 transfer if the time interval is very small ($E[N_A]=N*P[A]$ with very small time interval➔N=1), we may write that:

$$\begin{aligned} f_{jk} &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P[\{X(t+\Delta t)=k\} \cap \{X(t)=j\}] \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P[\{X(t+\Delta t)=k\} \mid \{X(t)=j\}] P[X(t)=j] \\ &= \left(\lim_{\Delta t \to 0} \frac{1}{\Delta t} P[\{X(t+\Delta t)=k\} \mid \{X(t)=j\}] \right) P[X(t)=j] \end{aligned}$$

We recognize the limit, by (U16.1) and (U16.2), as λ_{jk}, and the term P[X(t)=j] as the long-run state probability $p_{j,\infty}$. Therefore:

$$f_{jk}=\lambda_{jk}\, p_{j,\infty} \qquad \text{(U16.29)}$$

One may also at this point distinguish f_{jk} from λ_{jk} by noting that whereas λ_{jk} is conditioned by the system being in state j, f_{jk} is not. Therefore,

- whereas λ_{jk} gives the expected number of transfers from state j to state k per unit time, *given the system is in state j*,
- f_{jk} gives the expected number of transfers from state j to state k per unit time, *with no condition specified*.

Therefore, the f_{ik} will always be less than λ_{jk}, unless $p_{j,\infty}=1.0$, which could be the case in a two-state process with $\lambda_{kj}=0$ (0 repair time, or instant renewal).

Now, from our definitions,

- f_j is the expected number of stays in (or arrivals into, or departures from) state j per unit time, and
- f_{jk} is the expected number of transitions (or departures from) state j to state k per unit time,

Then it follows that

$$f_j = \sum_{k \neq j} f_{jk} \qquad \text{(U16.30)}$$

Substitution of (U16.29) into (U16.30) results in:

$$f_j = \sum_{k \neq j} \lambda_{jk} p_{j,\infty} = p_{j,\infty} \sum_{k \neq j} \lambda_{jk} \qquad \text{(U16.31)}$$

Equating (U16.31) to (U16.28), we have:

$$f_j = \frac{p_{j,\infty}}{T_j} = p_{j,\infty} \sum_{k \neq j} \lambda_{jk} \qquad \text{(U16.32)}$$

Solving for T_j, we obtain:

$$T_j = \frac{1}{\sum_{k \neq j} \lambda_{jk}} \qquad \text{(U16.33)}$$

which tells us that the mean duration of stays in any given state equals the reciprocal of the total rate of departures from that state [5].

16.6.4 Relation of frequencies and durations to MTTF, MTTR, MTBF

It is worthwhile recalling that we have used durations for a 2-state model in module U12, where we saw that MTTF=1/λ and MTTR=1/μ, where λ and μ were the failure and repair rates

(intensities), respectively. These expressions are consistent with (U16.33) where MTTF=T_1, MTTR=T_2, $\lambda=\lambda_{12}$, $\mu=\lambda_{21}$, and there is only one term in the summation of (U16.33).

It is also useful to consider (U16.27), $f_j=1/T_c$, for the 2-state model. In this case, $T_c=T_1+T_2$=MTTF+MTTR, so that

$$f_1 = f_2 = \frac{1}{MTTF + MTTR}$$

We can also define here the mean-time-between-failures as

$$\text{MTBF=MTTF+MTTR}$$

so that clearly,

$$f_1 = f_2 = \frac{1}{MTBF}$$

Here we see that the MTBF is the entire cycle time for the 2 state model, i.e., MTBF=$T_c=T_1+T_2$. We also see, once again, that the state 1 frequency is the same as the state 2 frequency, since the number of times it leaves from state 1 is the same as the number of times it enters into state 2.

U16.7 Analysis of a multistate model

We consider a three-phase transformer comprised of three single-phase transformers with one spare single phase transformer [5], characterized by the following information:

- The failure rates of the single phase transformers, including the spare, are the same, denoted by λ.
- Only a single repairman is available (so only 1 transformer may be repaired at a time), and it is his job to first get the spare installed as soon as one transformer fails, and second, after installing the spare, to repair the failed transformer.
- The transition rate for installing the spare is γ and that of repairing a failed transformer is μ.

- Define failure as having fewer than 3 transformers up (ignoring the possibility of operating them open-delta). As soon as fewer than three transformers are up, it is assumed that the whole bank is removed from service and no further failures can occur.

Our goal is to compute the long-run probabilities. The states are identified according to the following reasoning:

- The initial state is assumed to be 3 units up, 1 spare. We denote this as state 1.
- As a result of our definition of failure, the only failure state is one in which 2 units are up with no spare. Denote this as state 4.
- There are two other states:
    - From state 1, one unit may fail, leaving the system with 2 units up and 1 spare. This state we denote as state 2. Because there are three banks that may fail, the transition rate from state 1 to state 2 is 3λ. Also, we note that from state 4, the repairman may repair a failed transformer so that the system transitions from 4 back to 2.
    - From state 2, the repairman can install the spare, leading to a state where 3 units are up, with no spare.
    - From state 3, the system may either return to state 1 via repair of a failed transformer or it may transition to state 4 as a result of failure of one of the transformers.

The state space diagram for this system is illustrated in Fig. U16.6.

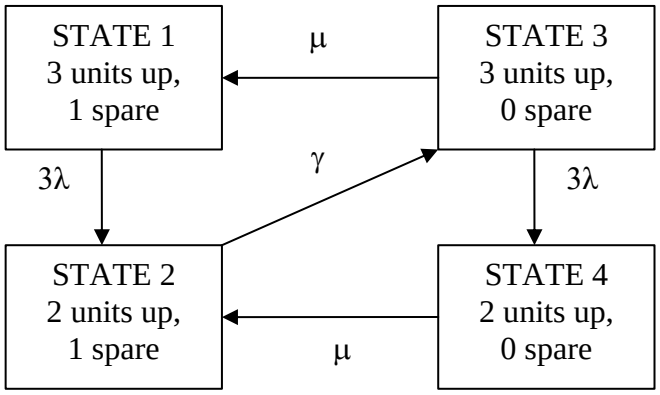

Fig. U16.6: State Space Diagram for Transformer Bank with Spare

Our first step is to set up the transition intensity matrix. This is:

$$\underline{A} = \begin{bmatrix} -3\lambda & 3\lambda & 0 & 0 \\ 0 & -\gamma & \gamma & 0 \\ \mu & 0 & -(3\lambda+\mu) & 3\lambda \\ 0 & \mu & 0 & -\mu \end{bmatrix}$$

Recall that we may obtain long-run probabilities from (U16.22), repeated here for convenience:

$$\underline{0} = \underline{p}_{\infty}\,\underline{A}$$

or,

$$\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} p_{1,\infty} & p_{2,\infty} & p_{3,\infty} & p_{4,\infty} \end{bmatrix} \begin{bmatrix} -3\lambda & 3\lambda & 0 & 0 \\ 0 & -\gamma & \gamma & 0 \\ \mu & 0 & -(3\lambda+\mu) & 3\lambda \\ 0 & \mu & 0 & -\mu \end{bmatrix}$$

Also recall that we know that $\underline{A}$ is singular, so we need to eliminate one of the equations and replace it with

$$p_{1,\infty}+p_{2,\infty}+p_{3,\infty}+ p_{4,\infty}=1$$

This amounts to replacing one column in A with a column of "ones," simultaneous with replacing the corresponding zero in the row vector of the left-hand-side with a 1, resulting in:

$$\begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix}=\begin{bmatrix}p_{1,\infty} & p_{2,\infty} & p_{3,\infty} & p_{4,\infty}\end{bmatrix}\begin{bmatrix}-3\lambda & 1 & 0 & 0\\ 0 & 1 & \gamma & 0\\ \mu & 1 & -(3\lambda+\mu) & 3\lambda\\ 0 & 1 & 0 & -\mu\end{bmatrix}$$

Lets solve it for several different sets of data. We specify the data in terms of failure rate per year for each transformer, repair time in hours, and the install time in hours. It is important to convert all of the data into the same units, according to the following:

λ=failure rate per year/8760

μ=1/(repair time in hours)

γ=1/(install time in hours)

Results are indicated in Table U16.1.

Table U16.1: Results for Transformer Example

Failure Rate (per yr)	Repair Time (hrs)	Install Time (hrs)	$p_{1,\infty}$	$p_{2,\infty}$	$p_{3,\infty}$	$p_{4,\infty}$
0.1	1000	50	0.9641	0.0017	0.0330	0.0011
1	1000	50	0.6744	0.0155	0.2310	0.0791
1	1000	10	0.6829	0.0031	0.2339	0.0801
1	200	50	0.9162	0.0168	0.0628	0.0043
1	200	10	0.9286	0.0034	0.0636	0.0044

Note that states 2 and 4 are the ones that we would want to avoid, whereas states 1 and 3 are operable. The results indicate that failure rates of 1 per year will cause us to see a lot of down time, even if we minimize the install time, unless we can minimize repair time.

U16.8 Combining states

It is often convenient, and possible, to reduce the number of states in a Markov model without losing accuracy in the results associated with the retained and original results. However, it is required, in doing so, that the resulting model remain Markovian, i.e., that the transition times remain exponentially distributed.

Two definitions are necessary first:

- Internal states are the states to be merged.
- External states are the remaining states to be retained.

We will simply state the necessary condition, when combining states, such that the model remains Markovian. Proofs are given in a variety of textbooks, and in a paper [8] that is convenient to power system engineers. The condition is [5]:

> A group of (internal) states can be merged if the transition intensities to any external states are the same from each internal state.

Note that the condition puts a requirement on the transition intensities out of the internal states but not on the transition intensities into the internal states. We call this the *merging condition*.

Consider that we detect a group of states satisfying the condition. Let's denote the group of states as J such that all states j to be merged are internal states and satisfy $j \in J$. The situation is illustrated in Fig. U16.8.

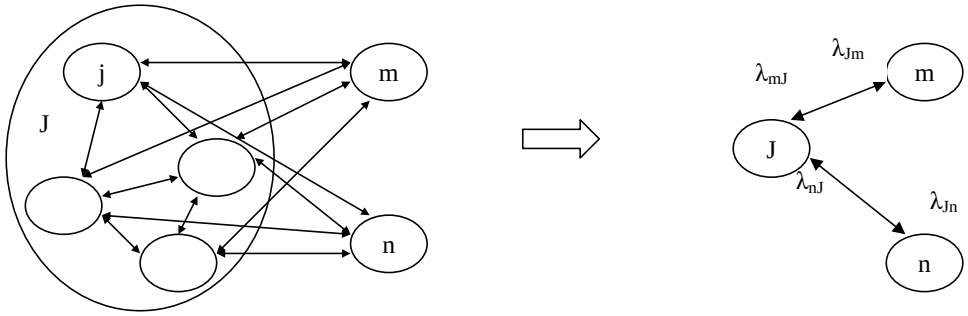

Fig. U16.8: Combining States

An obvious step to take when merging states is to maintain both transitions into the group of internal states and transitions from the group of internal states to external states. This is done in Fig. U16.8 since states m and n had transitions into and from the states in group J before and after the merging.

The question becomes, however:

What should be the combined state J probability, frequency, and transition intensities, denoted by p_J, f_J, λ_{kJ}, and λ_{Jk}, respectively?

In denoting the new transition intensities, k may represent any external state to which an internal state is connected. In Fig. U16.8, k=m and k=n.

16.8.1 State Probability

The events corresponding to the random process being in any of the internal states $j \in J$ are mutually exclusive, as a result of this being a Markov process (which requires states to be mutually exclusive, i.e., the process cannot reside in more than one state at a time).

As a result, the state probability for combined state J is given by:

$$p_{J,\infty} = \sum_{j \in J} p_{j,\infty} \qquad \text{(U16.34)}$$

16.8.2 State frequency

The frequency of the new state J, f_J, is the total of the frequencies of leaving an internal state j for an external state k, and therefore

$$f_J = \sum_{k \notin J} \sum_{j \in J} f_{jk} \qquad \text{(U16.35)}$$

By (U16.29), $f_{jk}=\lambda_{jk}\, p_{j,\infty}$, the above may be expressed as

$$f_J = \sum_{k \notin J} \sum_{j \in J} \lambda_{jk}\, p_{j,\infty} = \sum_{j \in J} \sum_{k \notin J} \lambda_{jk}\, p_{j,\infty}$$

$$= \sum_{j \in J} p_{j,\infty} \sum_{k \notin J} \lambda_{jk} \qquad \text{(U16.36)}$$

Note that the first summation cannot be replaced by eq. (U16.34) since the second summation is different for each term in the first summation.

16.8.3 Transition intensities

The transition intensities λ_{kJ}, and λ_{Jk} are computed on the basis of two requirements on characterizing indices:

- the frequencies of transfer from external state k to J, f_{kJ}, must be the same as that from external state k to all the internal states j before their combination, meaning:

  $$f_{kJ} = \sum_{j \in J} f_{kj}$$

  and replacing the frequencies on both sides by $p\lambda$, we get:

$$p_{k,\infty}\lambda_{kJ} = \sum_{j\in J} p_{k,\infty}\lambda_{kj}$$
$$\Rightarrow p_{k,\infty}\lambda_{kJ} = p_{k,\infty}\sum_{j\in J}\lambda_{kj} \qquad \text{(U16.37)}$$
$$\Rightarrow \lambda_{kJ} = \sum_{j\in J}\lambda_{kj}$$

- the transitions from J to k must be the same as

$$f_{Jk} = \sum_{j\in J} f_{jk}$$

and again replacing the frequencies on both sides by pλ, we get:

$$p_{J,\infty}\lambda_{Jk} = \sum_{j\in J} p_j\lambda_{jk} \qquad \text{(U16.38)}$$

where, this time, we cannot remove p_j from the summation as we did in (U16.37), so, in solving for λ_{Jk}, we obtain:

$$\lambda_{Jk} = \frac{\sum_{j\in J} p_{j,\infty}\lambda_{jk}}{p_{J,\infty}} = \frac{\sum_{j\in J} p_{j,\infty}\lambda_{jk}}{\sum_{j\in J} p_{j,\infty}} \qquad \text{(U16.39)}$$

If the merging condition is satisfied, that all the λ_{jk} are the same, then it is just a constant within the summation and can therefore be factored from the numerator of (U16.39), resulting in:

$$\lambda_{Jk} = \lambda_{jk} \qquad \text{(U16.40)}$$

Note the difference between eq. (U16.40) and the merging condition. The merging condition requires λ_{jk} to be the same for all j in J in order to merge, whereas eq. (U16.40) indicates what the resulting transition rate from the merged state should be after merger.

Therefore, (U16.37) and (U16.40) result in the following general rule for merging states [10]:

> Rule 1: When two (internal) states have identical transition rates to common external states, those two states can be merged into one. Entry rates are added. Exit rates remain the same.

It should also be mentioned that a state diagram representing a Markov model may also include states which have entry transitions but no exit transitions, i.e., once that state is reached, the process will never depart. Such states have 1.0 transition probability to themselves and are called absorbing states. For example, catastrophic failure states are often modeled as absorbing states. If a state diagram consists of multiple absorbing states, then they may be combined into a single state. Obviously, absorbing states satisfy the merging condition (which is captured in rule 1), since they have no transitions to other states. So we arrive at a version of rule 1 that is specialized for the case of absorbing states.

> Rule 2a: Multiple absorbing states may be merged into a single state. Entry rates from the common external states are added.

Another special case of rule 1 which deserves mention is when a non-absorbing state makes transition only to an absorbing state. In this case, the two states still satisfy rule 1 because they have no transitions to *any external states* (note that rule 2a covers the case when two states have no transitions to *any state*). They may therefore be combined according to rule 2b:

> Rule 2b: An absorbing state may be combined with a nonabsorbing state if the nonabsorbing state may transition only to the absorbing state and no external state. In this case, entry rates from the common external states are added.

The above three rules preserve model integrity, and calculations for external states will be the same as what would have been obtained without the model simplifications.

One final comment should be made here. It is typical that the states are classified according to some criteria. The most basic and most common criterion is whether the state is a success state or a failure state, but it is possible and not uncommon to have more than just two classifications. For example, there may be multiple failure modes making it desirable to classify failure states according to which failure mode it corresponds. In such a case, one would typically not want to combine states of different classes since there was apparently some original motivation to distinguish between them, and once they are combined, it is no longer possible to do so. So we have a final rule:

> Rule 3: Two states satisfying rules 1, 2a, or 2b should be combined only if they are of the same state classification.

Example [11]: Use our rules to simplify the state diagram of Fig. U16.9.

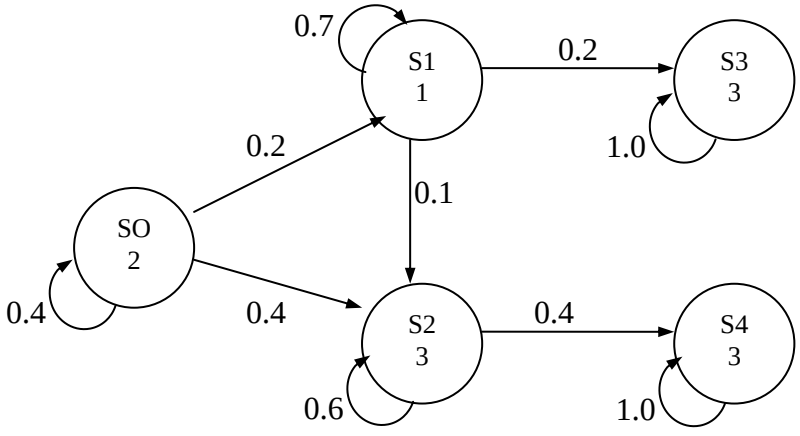

Fig. U16.9: Example for combining states

In Fig. U16.9, there are two identifiers within each state circle. The top one indicates the state ID, i.e., S0, S1, …. The bottom one

indicates the class to which the state belongs; in this case, possible classes are 1, 2, or 3.

One recognizes states S3 and S4 immediately as absorbing states and therefore, according to rule 2a, they are candidates for combining. We may do so without violating rule 3 since the two states are both of class 3 and therefore of the same class. There are not common entry states so there is no need to add entry rates. The resulting model following combination of states S3 and S4 is given in Fig. U16.10.

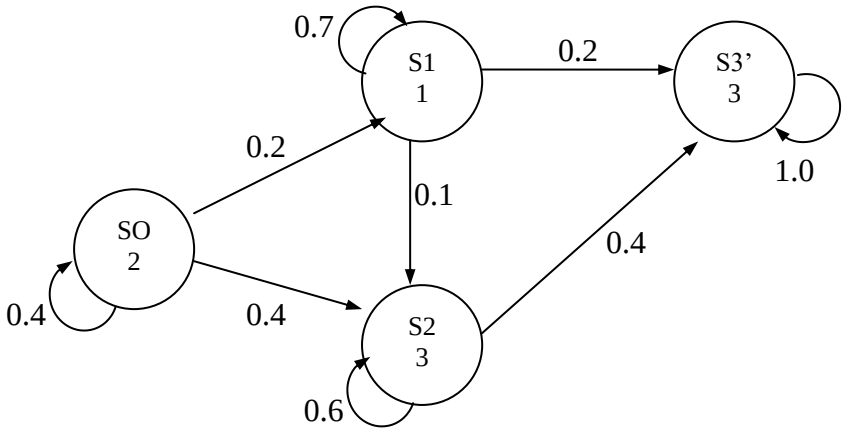

Fig. U16.10: After combining states S3 and S4 into S3'

We note from Fig. U16.10 that S2 has transition only to S3', which is an absorbing state, and therefore, according to rule 2b, they are candidates for combining. We may do so without violating rule 3 since the two states are both of class 3 and therefore of the same class. There is a common entry state, S1, so we must add the corresponding entry rates from S1: 0.2+0.1=0.3. The resulting model following combination of states S2 and S3' is given in Fig. U16.11, which represents the most simplified model for this example.

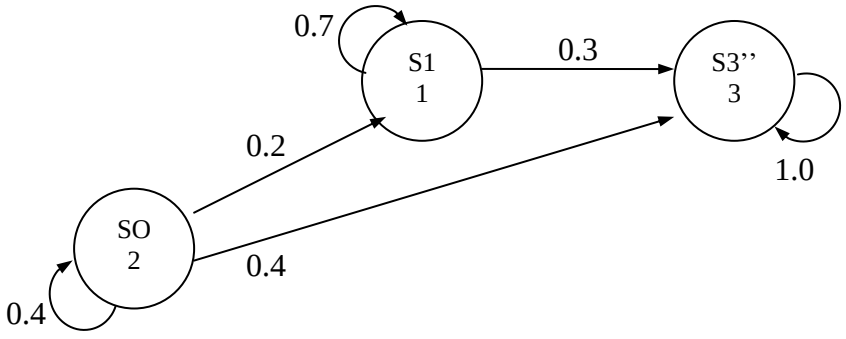

Fig. U16.11: After combining states S2 and S3' into S3''

U16.9 Construction procedure for Markov models

The most important aspect to developing good Markov models, and probably to developing good models of any type, is to understand the physical system being modeled. Without a thorough and proper understanding of the system, further model development is not typically fruitful, although the effort of doing so may lead to increased understanding of the physical system. If the analyst finds that understanding of the physical system is lacking, then time should be taken to gain this understanding. Typically, this begins with a significant amount of reading documents that describe the system – e.g., books, papers, manuals, etc. An on-site visit to view the system is a must. Discussions with knowledgeable people, face-to-face if possible, are extremely useful. Subsequent activities stemming from these steps, that can facilitate obtaining this understanding, are described in the following subsections. There can be considerable overlap in the activities described in these subsections.

U16.9.1 Failure definition

Perhaps the most important step is to identify what it means for the overall system to fail. This can be hard, when the system has the characteristic that it may continue to operate at a lower performance level while in a degraded state. **Is the degraded state a failure, or not?** In large scale power systems, a loss of

continuity between source and load almost always constitutes a failure, and there are various other levels of refinements to power system failure criteria, e.g., in terms of circuit loading, bus voltages, etc. In fact, most power systems operate under well-defined industry-developed reliability criteria, which typically serve well to guide the answer to this question.

Such explicit criteria may not be available for smaller, more self-contained systems, and so, in such cases, one must develop the failure criteria. In this case, one suggested rule [10] is "needed functions must be accomplished in the needed time period." If the system, while in the degraded state, is able to do this, then it is not a failed state.

16.9.2 Perform a system-level FMEA analysis

Identifying all individual system components, their function(s), how they are intended to operate, and how they may fail. This step is typical of what is called a *failure modes and effects analysis* (a bottom-up approach), and there are a number of good references that are useful in guiding such effort. The end result should also include the system effects of each failure mode for each component, the criticality, or severity, of those effects, and the transition rate for each failure mode.

16.9.3 Categorize failures

Different failure modes may be categorized or classified according to their system level effect and criticality. A typical classification is whether each failure is "success" or "failure" from an overall systems point of view. In some cases, there may be different graduations between these two extremes such that there are a number of different classes of interest (see the related discussion in Section 16.8.3 above).

This step is sometimes called a "failure effects analysis" [5]. It is perhaps the most important, and often the most computationally intensive, step of the entire procedure.

The power system engineer is able to appreciate this last statement when considering all of the possible topological states of a large-scale power system such that a state is defined by the identity of the elements (circuits and generators) that are in service. If we desire to classify these states in terms of failure or working, where failure is defined as all voltages within a desired acceptable band and all flows below a desired acceptable circuit rating, then we would have to run a power flow computation for every single state.

16.9.4 Develop the Markov model

Markov model construction begins from a successful state – i.e., a state where all components are operating. Then begin identifying other states using the following rule: "**For any successful state, list all failure rate categories for all successful components**." Once all successful states and their transition to failure states have been identified, then the model is completed by drawing in all necessary repair transitions.

16.9.5 Simplify the model

Absorbing and transition states should be merged according to the rules for combining states as specified in Section 16.8.

16.9.6 Solve for the state probabilities

Develop the transition intensity matrix $\underline{A}$ and then use it to solve for the state probabilities. If only long-run state probabilities are desired, then one may simply solve (U16.22) involving only linear, algebraic equations. If one wants the transient state probabilities as well, then the differential equation (U16.10) must be solved, which may be done in one of three ways:

- LaPlace Transforms: This approach results in a rigorous analytical expression and is quite desirable. However, it is typically only tractable for models with a small number of states.

- Numerical integration: This is a certain and complete solution procedure that, if properly implemented on a computer, will provide the solution.
- Transition matrix: This approach is, although approximate, as good as the numerical integration approach if the time increment is chosen to be very small. It has the advantage of being the simplest approach of the three.

16.9.7 Computing the reliability indices

One the state probabilities are obtained, we may compute class probabilities, frequencies, and durations associated with any particular group of states. If the class of interest is that corresponding to "failure," then we obtain the failure state probability (otherwise known as the probability of failure), the failure state frequency (how frequently the system fails), and the failure state duration (how much time the system is expected to reside in the failed state).

Class Probabilities: Assume that class C contains n states $j \in C$. Then the probability of residing in one of the class C states is the probability of residing in state j_1 or j_2 or j_3 or … or j_n. Since all Markov states are mutually exclusive, this is simply the summation of the individual probabilities, as implied by (U16.34), so that

$$p_{C,\infty} = \sum_{j \in C} p_{j,\infty} \tag{U16.41}$$

Class frequency: The class frequency f_C is the frequency for which the system resides in the class of states C. It is computed by (U16.36) according to:

$$f_C = \sum_{j \in C} p_{j,\infty} \sum_{k \notin C} \lambda_{jk} \tag{U16.42}$$

One can understand f_C as the sum of the class C state probabilities, each multiplied by the rate of transitions from the respective state

to a state of another class. If there were only two classes, success W and failure F, then f_F would be the sum of the system failure state probabilities, each multiplied by the rate of transitions from the respective state to the success domain [5]. Alternatively, f_C is the sum of {the expected number of transfers from the class C states $\sum_{k \notin C} \lambda_{jk}$ (which is a sum of conditional probabilities) multiplied by the probability of being in that Class C state $p_{j,\infty}$ }.

Class mean duration: The class mean duration T_C is the expected amount of time, i.e., the mean time, for residing in states of class C. From (U16.28) we have that $T_C=p_{C,\infty}/f_C$, and on substitution of (U16.41) and (U16.42), we have:

$$T_C = \frac{p_{C,\infty}}{f_C} = \frac{\sum_{j \in C} p_{j,\infty}}{\sum_{j \in C} p_{j,\infty} \sum_{k \notin C} \lambda_{jk}} \qquad \text{(U16.43)}$$

16.9.8 Example

An interesting illustration of Markov model construction is given in the appendix. There are 4 classes of states identified. The state probabilities are obtained using the transition matrix approach. This illustration was obtained from [12], originally taken from [11].

U16.10 Truncating the state space

One can recognize an underlying philosophy behind the procedure of Section 16.9, where we identify all states that the process may possibly visit, classify them, construct the Markov model, and then compute indices associated with the states within certain classes of interest. This approach is sometimes referred to as the *state*

enumeration approach since we must enumerate all of the states in order to classify them.

In principle, there is nothing wrong with this approach. In practice, however, it suffers from the fact that in many kinds of systems, the number of states can be extremely large, and their complete enumeration can be very tedious. As a result, it is of interest to look for ways to reduce the state space. If one thinks of the state space in terms of a growing "tree" of states on a piece of paper, where the tree grows from left to right, then our task can be viewed as an effort to truncate this tree so that many states, and thus much evaluation work, are eliminated. In this sense, then, we are attempting to truncate the state space.

The fundamental concept behind state space truncation is to eliminate states that have probabilities not significantly affecting the desired reliability indices. This can save much computational time as it means that we eliminate the need to classify many of the states (and thus the need to analyze them using some computationally intensive tool such as, for example, the power flow program).

A note of care, however, is in order. The above stated concept does not necessarily mean that we simply eliminate low probability states – but rather, we eliminate the states that do not significantly affect the desired reliability indices. If our desired reliability indices are probability, frequency, and duration of system failure, then we need to eliminate states that do not significantly affect these indices. The point is that a low probability state may actually comprise a significant portion of the total system failure probability.

Reference [5] provides a very illuminating example in regards to this point, which we also describe here.

Consider a system comprised of N identical, independent components (e.g., transmission circuits) such that the availability of each component is A (long-run probability of being in the

working state) and the unavailability is 1-A (long-run probability of being in the failed state).

The probability of a state where r out of N components have failed is the product of the probability that r out of N components have failed and the probability that the remaining N-r components have not: $(1\text{-}A)^r A^{N\text{-}r}$. Since the number of ways that r out of N identical components can fail is given by N!/r!(N-r)!, the probability of residing in a state having r failed components, called an r-fold failure, is given by a binomial distribution according to:

$$p_r = \frac{N!}{r!(N-r)!}(1-A)^r A^{N-r} \qquad \text{(U16.44)}$$

Let N=35 and A=0.9. The probability distribution for r=1,…r=9 is shown in Figure U16.12.

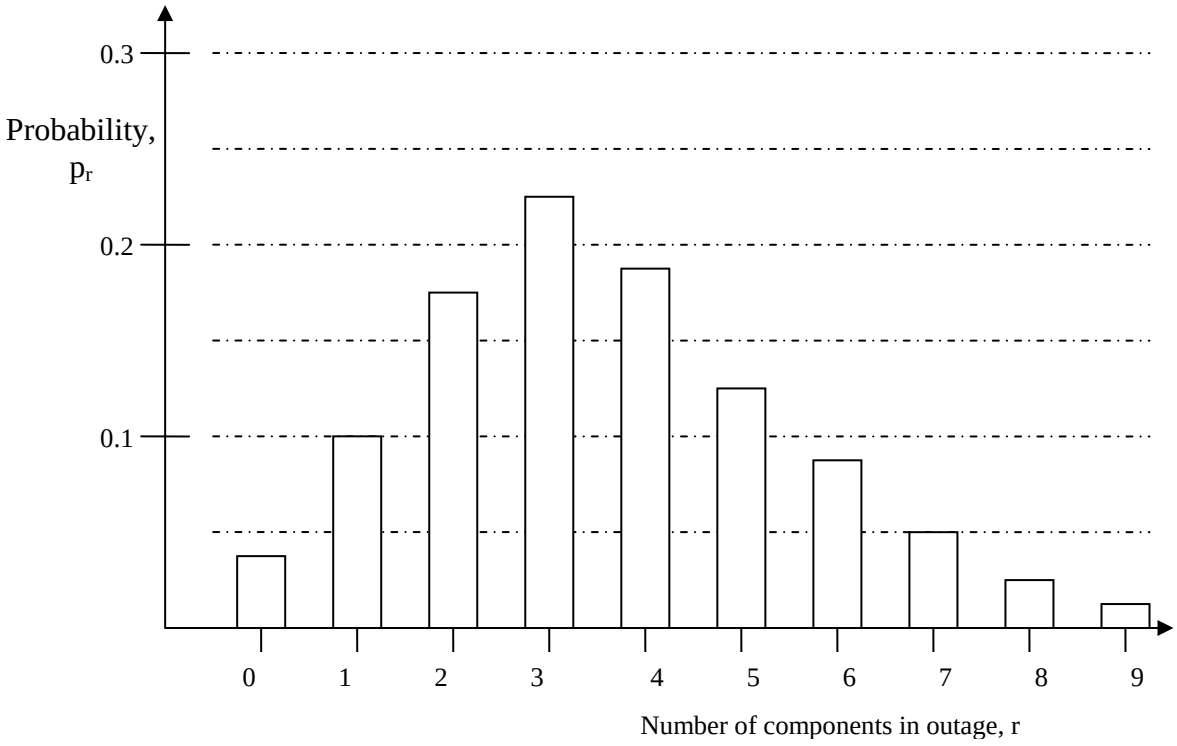

Fig. U16. 12: Probability distribution for r-fold failures [5]

It is clear that states consisting of more than about 7 failed components constitute a very small portion of the total state space probability.

We are mainly interested in, however, for each of the r-fold failure states, the percentage for which are system failures. Call this ρ_{Fr}. In order to determine this, we need to understand the system, what it means to fail, and how to detect the system failure. In a power system, for example, we would need to run the load flow program for each state to determine the failure.

Let's avoid tedious discussion of this kind of analysis and simply assume that ρ_{Fr}, the proportion of states in each set of r-fold failure states that are system failures, increases linearly between some $r=r_1$ and some $r=r_2$, such that for $r \leq r_1$, there are no system failure states and for $r \geq r_2$, all states are system failure states.

This is actually quite representative of many systems. In a power system, for example, it might very well be the case that no N-1 or N-2 contingency causes system failure, but all N-k contingencies, $k \geq 6$, do cause system failure.

Let $r_1=2$ and $r_2=6$. Then we can observe how ρ_{Fr}, the proportion of system failures in each set of r-fold failure states, according to shaded regions of the bars in Fig. U16.13, increases with r. The actual value of the probability of system failure for each set of r-fold failure states is $p_{F3}=\rho_{Fr}p_r$. We see clearly that $r_1 \leq 2$ has no system failures and $r_2 \geq 6$ has only system failures.

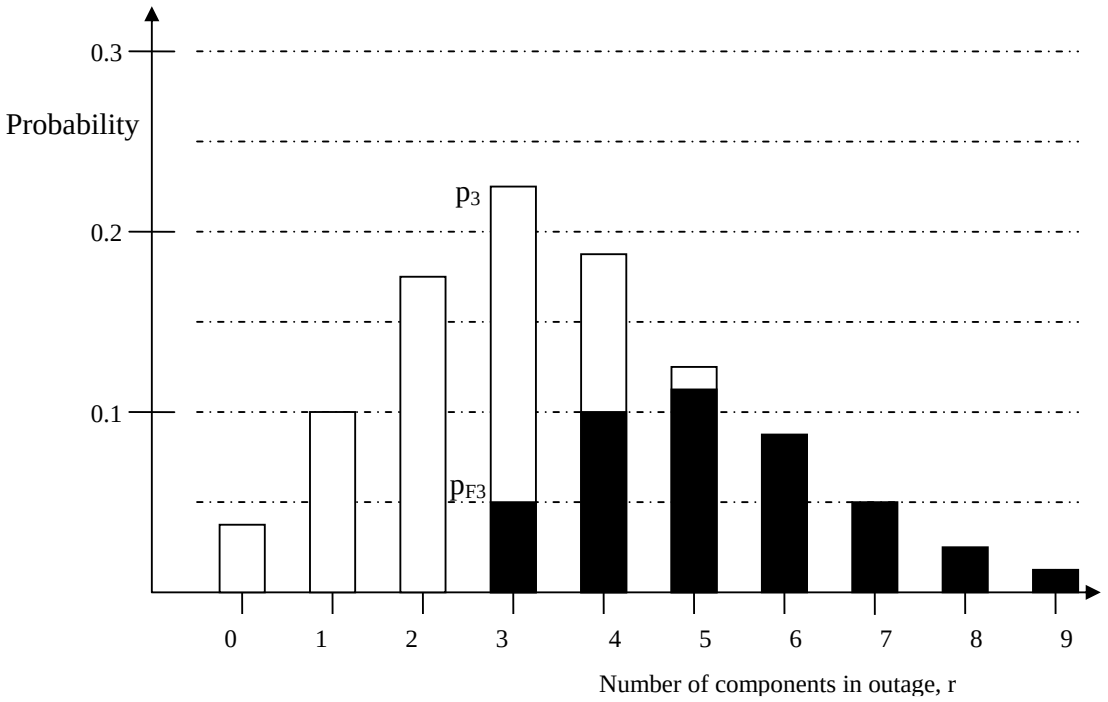

Fig. U16.13: Probability distribution for r-fold failures [5]

Fig. U16.13 illustrates that although the sets of r=8- and r=9-fold failures comprise small portions of the total state probabilities, perhaps 3%, they comprise much larger portions of the total system failure probabilities (as a percentage of the total blackened portion) – about 9%.

One common approach to state-space truncation is to analyze only failures for r=1, …, r=r_0, where r_0 is chosen so that for r>r_0, the corresponding system failure states have such low probability that they do not affect the total system failure probability. A well-known example of this in power systems is when analyses are done only for so-called N-1 (r_0=1) or possibly N-2 (r_0=2) contingencies only, excluding N-3 and higher order contingencies.

Consider indexing the spread of r-fold failures that contain both success and failure states as m=r_2-r_1. For the example of Fig. U16.12 and U16.13, r_1=2 and r_2=6, so that m=4, i.e. there are

m-1=3 levels of r-fold failures that contain both success states and failure states.

Let's define an error associated with selecting a particular r_0 for truncating the state space as:

$$\varepsilon = \frac{\sum_{r=0}^{N} p_{Fr} - \sum_{r=0}^{r_0} p_{Fr}}{\sum_{r=0}^{N} p_{Fr}} = \frac{\sum_{r_0+1}^{N} p_{Fr}}{\sum_{r=0}^{N} p_{Fr}} \qquad \text{(U16.45)}$$

Let's assume that an error of 10% (ε=0.10) is acceptable, and identify the minimum truncation level r_0 that would satisfy this error. In the example of Fig. U16.12 and U16.13, where N=35, m=4, 1-A=0.01, the minimum truncation level would be 7, since we already observed for the 8- and 9-fold failures comprise about 9% of the total system failures.

Repeating this analysis for different values of N, A, r_1,and m results in Table U16.2. In inspecting this table, one should be clear regarding the various terms, repeated here for convenience:

- r_1: highest level of r-fold failures that cannot cause system failure.
- m: indexes the spread of r-fold failures that contain both success and failure states.
- r_0: minimum truncation level necessary to achieve 10% error.

For example, inspection of the Table U16.2 element corresponding to N=35, 1-A=0.1, r_1=2, and m=4, indicates r_0=7, implying that for a system size of 35 components, each of which have unavailability of 0.1, if system failures only occur for more than 2 component failures (N-k, k>2) and always occur for 2+4=6 or more

component failures (N-k, k≥6), then one must analyze up to the 7 component failures (N-k, k≤7) to achieve 10% accuracy.

Table U16.2: Minimum truncation levels for 10% error [5]

1-A	r_1	Minimum truncation level (r_0) for 10% error									
		N=10		N=35			N=100			N=1000	
		m=4	m=8	m=4	m=8	m=16	m=4	m=8	m=16	m=16	m=160
0.1	1	4	4	6	7	7	14	14	15	-	-
	2	4	4	7	7	7	14	14	15	-	-
	3	5	-	7	8	8	14	14	15	112	113
	30	-	-	-	-	-	-	-	-	112	113
0.01	1	2	2	3	3	3	4	4	4	-	-
	2	3	3	4	4	4	5	5	5	-	-
	3	4	-	5	5	5	5	5	5	15	15
	30	-	-	-	-	-	-	-	-	33	33
0.001	1	2	2	2	2	2	2	2	2	-	-
	2	3	3	3	3	3	3	3	3	-	-
	3	4	-	4	4	4	4	4	4	5	5
	30	-	-	-	-	-	-	-	-	31	31

One can draw the following conclusions from Table U16.2:

- The level r_0 appears to be fairly insensitive to m, suggesting that the level r_0 would be fairly insensitive to other models of p_{Fr} that might be used in place of the linear one used in this analysis.
- The level r_0 is heavily influenced by system size N and component unavailability 1-A.

  - If the unavailability is high, e.g., on the order of 0.1, then large systems require very large r_0, i.e., such systems require analysis of N-k failures with k very large in order to achieve accurate results.
  - If the unavailability is low, e.g., 0.001 or less, then reasonable accuracy requires significantly lower levels of r_0, even for large systems. Fortunately, this is usually the case for bulk power transmission systems.

- The level r_0 is also influenced by r_1. In most bulk transmission systems today, under stressed, but *secure* conditions, it would be the case that r_1=1, indicating that we could only guarantee no system failures for N-1 contingencies.

It is of interest to consider the effect of truncation on computational intensity, as characterized by the number of system states to be evaluated, s. If no truncation is performed, then all system states must be evaluated, so that is $s=2^N$. If truncation is applied up to level r_0, then the number of states to be evaluated is given by

$$s = \sum_{r=0}^{r_0} \frac{N!}{r!(N-r)!} \qquad \text{(U16.46)}$$

(It can be shown that the above evaluates to 2^N if r_0=N). It is of interest to plot the dependence of s associated with a particular r_0 to achieve a desired level of accuracy against the number of components. Such plots are provided on a logarithmic scale in Fig. U16.14 [5] for systems of size 10 up to N=100. These plots were developed using a linear model for ρ_{Fr}, with r_1=2 and r_2=8.

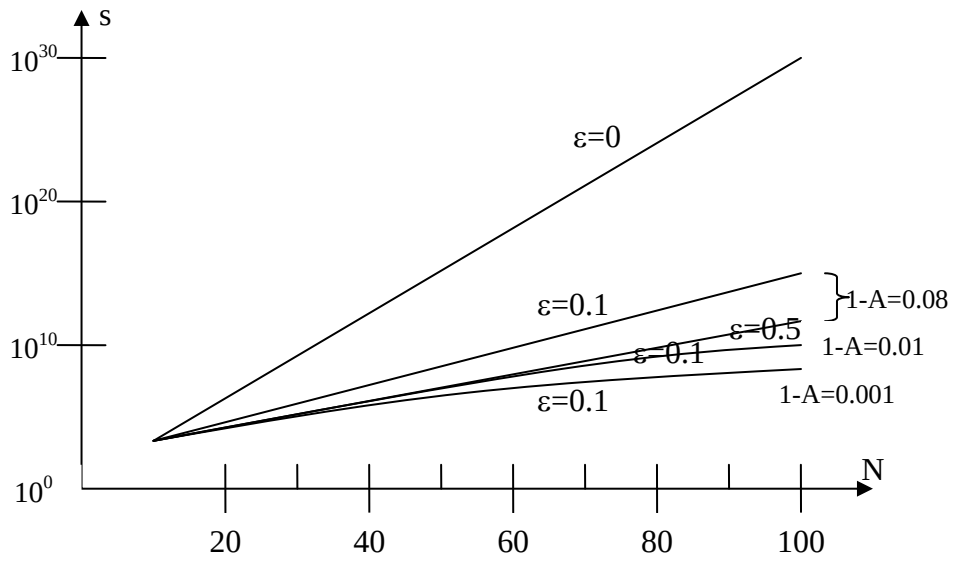

Fig. U16.14: Number of states to be evaluated, s, for a given error ε and component unavailability 1-A [5]

One observes from Fig. U16.14, that:

- The case of no error (ε=0), where $s=2^N$, is independent of (1-A).
- With (1-A)=0.08, s increases almost exponentially with N, even with high error allowed (ε=0.5).
- With lower values of (1-A), i.e., 0.01 and 0.001, the increase in s with N is no longer exponential and tends to be almost flat for large values of N.

As a final comment in this section, another method to save on computations is to restrict the states to be evaluated to the so-called minimal-cut states [5]. These are the failed states for which any repair transition (repair of any failed component in that failed state) will result in a success state. Thus, we see that the repair of any one of the components restores the system to the working condition, and as a result, such states are rightfully called minimum cut states. It is possible to show that the summation of minimum cut state probability is a lower bound on system failure probability. We investigate this further in a later module.

U16.11 Standby: effect of hidden failures/maintenance

Reference [5] provides an interesting example that illustrates the power of Markov models together with some insights regarding the effects of hidden failures and maintenance when a standby unit is available.

Consider a substation having a main transformer C and a standby transformer D, where the standby unit cannot fail when it is not in service, and it is ready for service each time it is called upon to operate (so no installation time). The state-space diagram of this system is shown in Fig. U16.15. The only system failure state is state 4.

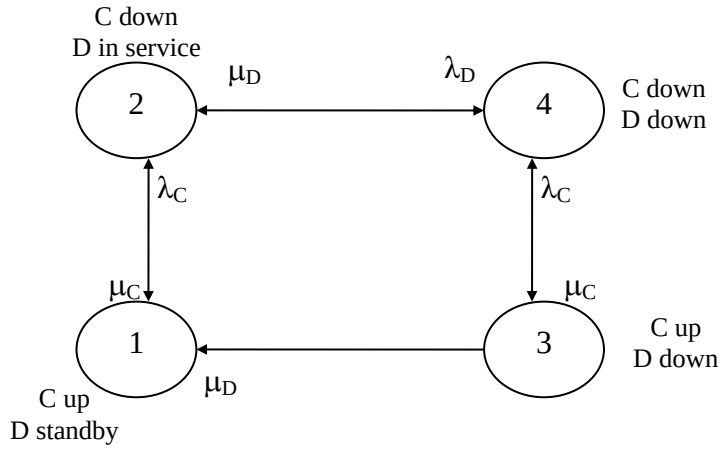

Fig. U16.15: State space diagram of substation with standby transformer [5]

The transition intensity matrix is given by:

$$\underline{A} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -\lambda_C & \lambda_C & 0 & 0 \\ \mu_C & -(\mu_C + \lambda_D) & 0 & \lambda_D \\ \mu_D & 0 & -(\mu_D + \lambda_C) & \lambda_C \\ 0 & \mu_D & \mu_C & -(\mu_D + \mu_C) \end{bmatrix} \end{array} \quad \text{(U16.46)}$$

To solve for the long-run probabilities, we use (U16.22), which is $\underline{0}=\underline{p}_\infty\underline{A}$ together with $\sum p_{\infty,i}=1$ to obtain the state probabilities. It is

easiest to do numerically on a computer, but one may also extract the equations, do some tedious algebra, and, on applying the assumptions that

a. the probability of being is the normal state (state 1) is very high, i.e., $p_1 \approx 1$, and

b. all repair rates are much greater than all failure rates, i.e., $\lambda_C, \lambda_D << \mu_C, \mu_D$,

it can be shown then the probability of system failure, which is p_4, is given by

$$P_F \approx \frac{1}{\mu_C + \mu_D} \frac{\lambda_C \lambda_D}{\mu_C} \tag{U16.47a}$$

By (U16.31), repeated here for convenience,

$$f_j = p_{j,\infty} \sum_{k \neq j} \lambda_{jk}$$

we have that the failure frequency, which is f_4, is given by

$$f_F \approx \frac{1}{\mu_C + \mu_D} \frac{\lambda_C \lambda_D}{\mu_C} (\mu_C + \mu_D) = \frac{\lambda_C \lambda_D}{\mu_C} \tag{U16.48a}$$

If the repair rate of both transformers is the same, i.e., $\mu_C = \mu_D$, then we obtain:

$$P_F \approx \frac{\lambda_C \lambda_D}{2\mu_C^2} \tag{U16.47b}$$

$$f_F \approx \frac{\lambda_C \lambda_D}{\mu_C} \tag{U16.48b}$$

It is interesting to compare the results (U16.47b) and (U16.48b) with a slightly different situation where, instead of a standby unit, both units are operating, in which case, we may have a transition from state 1 to state 3 with intensity λ_D. Analysis of this situation results in [5]

$$P_F = \frac{\lambda_C \lambda_D}{(\lambda_C + \mu_C)(\lambda_D + \mu_D)} \qquad \text{(U16.49a)}$$

$$f_F = \frac{\lambda_C \lambda_D (\mu_C + \mu_D)}{(\lambda_C + \mu_C)(\lambda_D + \mu_D)} \qquad \text{(U16.50a)}$$

If we apply here $\lambda_C, \lambda_D << \mu_C, \mu_D$, with $\mu_C = \mu_D$ we obtain

$$P_F = \frac{\lambda_C \lambda_D}{\mu_C^2} \qquad \text{(U16.49b)}$$

$$f_F = \frac{2\lambda_C \lambda_D}{\mu_C} \qquad \text{(U16.50b)}$$

where we see that the probability and frequency of failure (both units down) in the standby system are half the corresponding values in a system where the same two transformers are continuously operating in the normal state. The improvement in reliability is due to the fact that the standby unit cannot fail when it is not in service, i.e., when the system is in the normal state 1.

Let's extend our standby example illustrated in Fig. U16.15 to include the possibility that the spare transformer fails when it is not operating. Such a failure mode might be rightfully termed a *hidden*

failure since the transformer failure state would remain undetected until the transformer was needed and put into service. It is known that such situations are not uncommon for certain kinds of protection systems.

We model this situation using the following two assumptions:

- The hidden failure may occur only during the time when the main unit, C, is operating and the spare unit, D, is off-line in standby. (If the hidden failure occurred during unit D operation, then it would not be a hidden failure. Although, in actuality, the hidden failure may occur during the repair mode (state 3), it is convenient to assume that such failures are only manifested once the unit goes back into standby mode, so that a hidden failure may only occur from state 1. We denote the transition intensity of hidden failure as λ_h.
- Once the hidden failure occurs (from state 1), then the failure state is reached only on occurrence of failure to unit C, with transition intensity λ_C.

The resulting state diagram is illustrated in Fig. U16.16.

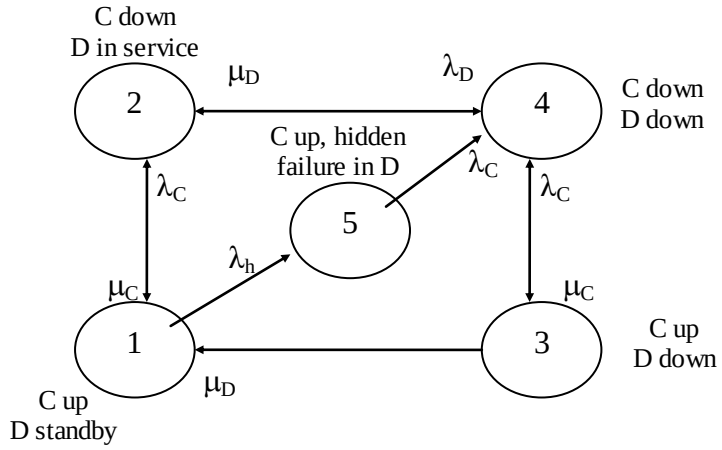

Fig. U16.16: State space diagram of substation with standby transformer and hidden failure [5]

The transition intensity matrix for the model shown in Fig. U16.16 is given as:

$$\underline{A}=\begin{bmatrix} -(\lambda_C+\lambda_h) & \lambda_C & 0 & 0 & \lambda_h \\ \mu_C & -(\mu_C+\lambda_D) & 0 & \lambda_D & 0 \\ \mu_D & 0 & -(\mu_D+\lambda_C) & \lambda_C & 0 \\ 0 & \mu_D & \mu_C & -(\mu_D+\mu_C) & 0 \\ 0 & 0 & 0 & \lambda_C & -\lambda_C \end{bmatrix}$$

Again, we solve for the long-run probabilities, using (U16.22), which is $\underline{0}=\underline{p}_\infty\underline{A}$ together with $\sum p_{\infty,i}=1$ to obtain the state probabilities. Extracting the equations, and doing some tedious algebra, and, again applying the assumptions (a) and (b) ($p_1\approx 1$ and $\lambda_C, \lambda_D << \mu_C, \mu_D$), we obtain the relations:

$$P_F \approx \frac{1}{\mu_C+\mu_D}\left(\frac{\lambda_C\lambda_D}{\mu_C}+\lambda_h\right) \qquad \text{(U16.51)}$$

$$f_F \approx \frac{\lambda_C\lambda_D}{\mu_C}+\lambda_h \qquad \text{(U16.52)}$$

The reliability of standby device can be improved by periodic inspection and preventive maintenance of it. A standby model including the effects of an “inspection/maintenance” program is illustrated in Fig. U16.17, where only the standby device is inspected and maintained. We assume that the “inspection/maintenance” activity on the spare is carried out only during the time period when the main unit is up and the spare unit is not failed, which means we may transition to an “inspection/maintenance” state from only states 1 and 5.

The difference between transitioning to a maintenance mode from state 5 and transitioning to a maintenance mode from state 1 is that in state 5, a hidden failure does in fact exist, whereas in state 1, it does not. Therefore, from state 5, we go to state 3, which is a state where the spare is "down" and must be repaired with repair rate μ_D. In contrast, from state 1, we go to state 6, which is a state where the spare is not "down" but just in maintenance and therefore returns to state 1 NOT with repair rate μ_D but with maintenance rate μ_M. It should be the case that $\mu_D << \mu_M$ (mean time to repair should be significantly greater than the mean time to maintain).

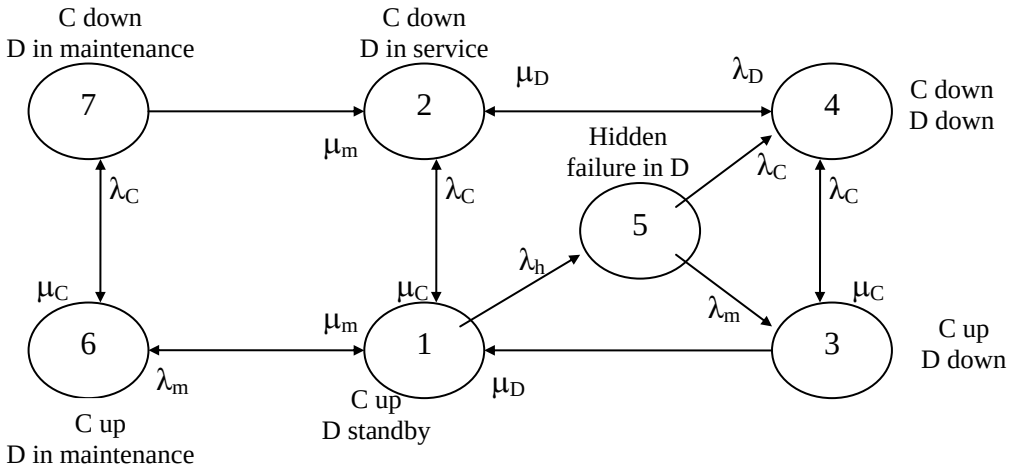

Fig. U16.17: State space diagram of substation with standby transformer, hidden failure, maintenance, and inspection [5]

In the model of Fig. U16.17, state 6 represents the state where the standby device is undergoing maintenance while the main unit is still operating. It is possible, of course, that during this maintenance, the main unit, C, may fail, in which case it transitions to state 7.

From state 7, since the main unit is failed, we transition to state 2 upon completion of the maintenance, where the spare is in service. Fig. U16.17 also illustrates that the spare has a hidden failure when it is inspected/maintained, as indicated in the transition from state 5 to state 3.

The system failure states are now 4 and 7.

The transition intensity matrix for the model shown in Fig. U16.16 is given as:

$$\underline{A}=\begin{bmatrix} -(\lambda_C+\lambda_h+\lambda_m) & \lambda_C & 0 & 0 & \lambda_h & \lambda_m & 0 \\ \mu_C & -(\mu_C+\lambda_D) & 0 & \lambda_D & 0 & 0 & 0 \\ \mu_D & 0 & -(\mu_D+\lambda_C) & \lambda_C & 0 & 0 & 0 \\ 0 & \mu_D & \mu_C & -(\mu_D+\mu_C) & 0 & 0 & 0 \\ 0 & 0 & \lambda_m & \lambda_C & -(\lambda_m+\lambda_C) & 0 & 0 \\ \mu_m & 0 & 0 & 0 & 0 & -(\mu_m+\lambda_C) & \lambda_C \\ 0 & \mu_m & 0 & 0 & 0 & \mu_C & -(\mu_m+\mu_C) \end{bmatrix} \quad \text{(U16.53)}$$

Again, we solve for the long-run probabilities, using (U16.22), which is $\underline{0}=\underline{p}_\infty\underline{A}$ together with $\sum p_{\infty,i}=1$ to obtain the state probabilities. Extracting the equations, and doing some tedious algebra, and, again applying the assumptions (a) and (b) ($p_1\approx 1$ and $\lambda_C, \lambda_D << \mu_C, \mu_D$) together with the additional assumptions that

c. The transition rate for hidden failures in the spare is much lower than the transition rate for normal failures in the main unit, i.e., $\lambda_h << \lambda_C$,

d. All repair rates are much greater than the maintenance rate, i.e., $\lambda_m << \mu_C, \mu_D$.

$$P_F = \frac{\lambda_C}{\mu_C+\mu_D}\left(\frac{\lambda_D}{\mu_C}+\frac{\lambda_h}{\lambda_C+\lambda_m}\right)+\frac{\lambda_C\lambda_m}{\mu_m(\mu_C+\mu_m)} \quad \text{(U16.54)}$$

$$f_F = \lambda_C\left(\frac{\lambda_D}{\mu_C}+\frac{\lambda_h}{\lambda_C+\lambda_m}+\frac{\lambda_m}{\mu_m}\right) \quad \text{(U16.55)}$$

Equation (U16.55), normalized by λ_h, is plotted against λ_m normalized by λ_C, for different values of $\gamma = \lambda_C T_m$.

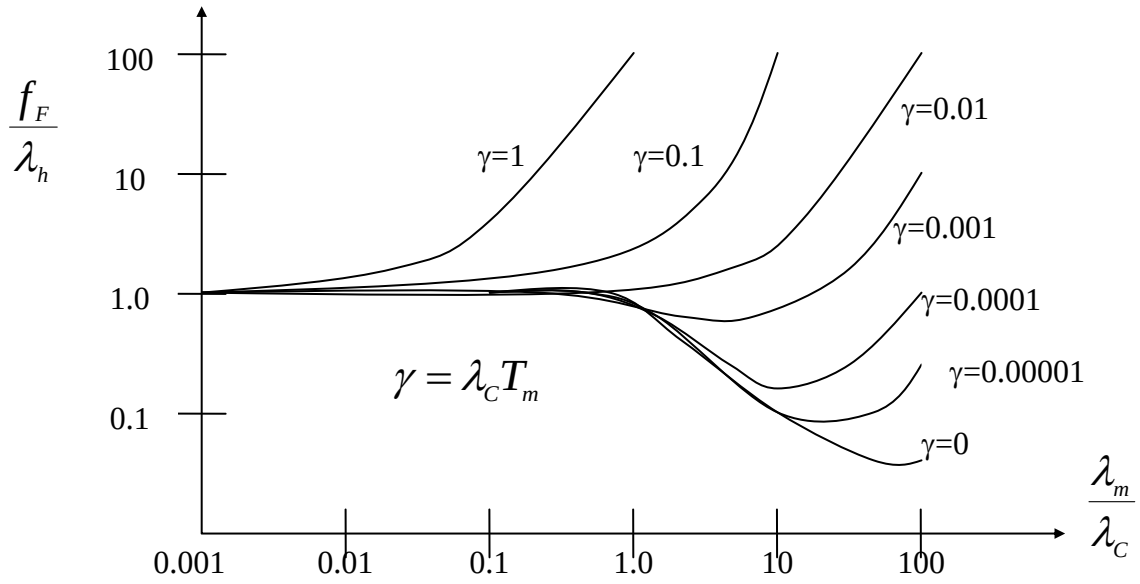

Fig. U16.18: Effect of maintenance on system failure frequency [5]

Conclusions that can be drawn from these curves are:

- Reduction of f_F can be achieved through reducing the mean duration of maintenance T_m. Since $T_m=T_6= 1/(\mu_m+\lambda_c)$, we can achieve this by increasing μ_m (we have no control over λ_c).
- At the same time, the maintenance rate λ_m must be set to an optimal value, which depends on $\gamma=\lambda_c T_m$.

Note that if T_m is large, an increase of the maintenance rate only increases the system failure frequency, implying that if you require a lot of time to perform the maintenance, performing maintenance frequently is not a good idea since the system will be exposed, for an extended period (the maintenance period) to a situation where a single component failure causes system failure.

It is evident that, in general, frequent short-term inspections (relatively high λ_m) allowing for relatively few maintenance overhauls of longer duration (low T_m) will improve system reliability.

The ability to model the influence of inspection and maintenance on failure is an important but complex subject of great current

interest. For example, the model illustrated in Fig. U16.19 allows representation of inspection, maintenance, and deterioration in a component together with associated decision-making processes. We will not have time to discuss this model further, but the interested reader can refer to [13, 14].

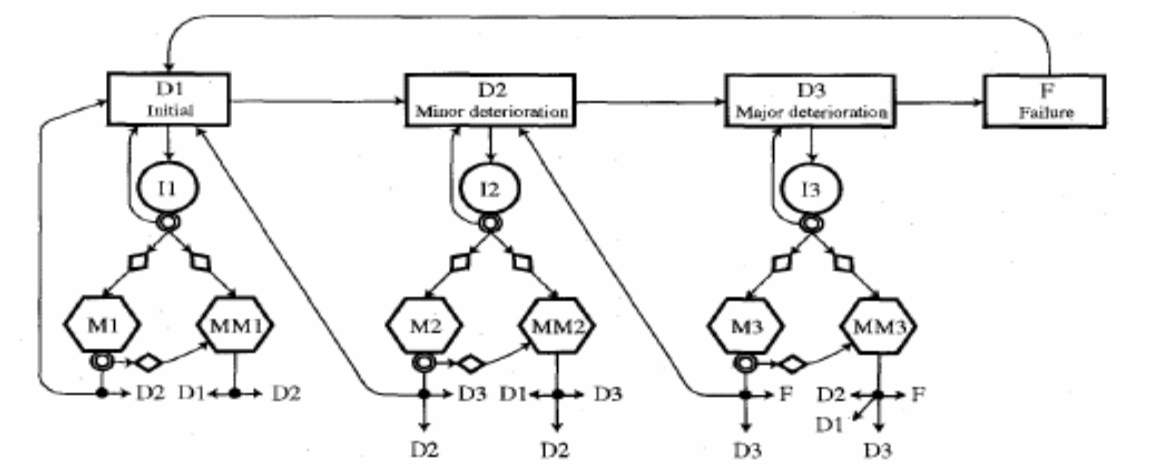

Fig. U16.19: Advanced Model for Maintenance/Inspection Effects

U16.12 First passage times

A final topic that we wish to mention in this module is that of first passage times. The first passage time is the expected value of the amount of time the process will take to transition from a given state j to another state k, under the assumption that the process begins in state j. The computation is facilitated by forcing state k in the model to be an absorbing state. The method of computing first passage times is provided in [4]. Also, reference [15] discusses this issue from a power system reliability perspective and describes in very readable terms two different methods for computing it.

References

[1] W. Blischke and D. Murthy, “Reliability: Modeling, Prediction, and Optimization,” John Wiley, New York, 2000.
[2] Y. Viniotis, “Probability and Random Processes for Electrical Engineers,” McGraw-Hill, New York, 1998.
[3] S. Ross, “Introduction to Probability Models,”Academic Press, New York, 1970.
[4] G. Anders, “Probability Concepts in Electric Power Systems,” John Wiley, New York, 1990.
[5] J. Endrenyi, “Reliability Modeling in Electric Power Systems,” John Wiley, New York, 1978.

[6] R. Billinton and R. Allan, “Reliability Evaluation of Engineering Systems: Concepts and Techniques,” Second edition, Plenum Press, New York, 1992.
[7] P. Anderson, “Power system protection,” volume III.
[8] J. Kemeny and J. Snell, “Finite Markov Chains,” Van Nostrand, Princeton, N.J., 1960.
[9] C. Singh and R. Billinton, “Frequency and Duration Concepts in System Reliability Evaluation,” IEEE Transactions on Reliability, Vol. 24, No. 1, pp 31-36, April 1975.
[10] W. Goble, “Evaluating Control Systems Reliability – Techniques and Applications,” Instrument Society of America, 1990.
[11] S. Zhao, “The influence of generator rejection system reliability on risk-based security assessment,” MS Thesis, Iowa State University, 1998.
[12] J. McCalley, W. Fu, “Chapter 6: Reliability of Special Protection Schemes,” in “Special Protection Schemes in Power Networks," by Task Force 38.02.19, chaired by Daniel Karlsson and Xavier Waymel, A technical brochure for International Conference on Large High Voltage Electric Systems (CIGRE), January, 2001.
[13] J Endrenyi, G. Anders, and A. Leite da Silva, “Probabilistic Evaluation of the Effect of Maintenance on Reliability – An Application,” IEEE Transactions on Power Systems, Vol. 13, No. 2, May 1998.
[14] G. Anders and J. Endrenyi, “Using Life Curves in the Management of Equipment Maintenance,” Proc. of the 7th International Conference on Probabilistic Methods Applied to Power Systems, Sept. 22-26, 2002, Naples, Italy.
[15] A. Leite da Silva and J. Endrenyi, “Application of First Passage Times in the Markov Representation of Electric Power Systems,” Proc. of the 4th International Conference on Probabilistic Methods Applied to Power Systems,” Rio de Janeiro, Brazil, September 1994.

Appendix Markov Model Example

The typical power plant, in which a GRS is installed, features high generation capacity and multiple generation units, interconnected to the system by two or more transmission lines. Without GRS, disturbances resulting in decreased transmission capacity may cause an out of step condition at the plant during high loading conditions. Any circuit that initiates GRS action during a forced outage condition is defined as a critical circuit. A properly designed GRS, activated by outage of any critical circuit, will trip a limited amount of generation at the plant in order to avoid out of step conditions for the remaining units [A.1-A.9]. Figure A.1 shows a portion of the IEEE Reliability Test System [A.10] together with an illustration of the GRS logic. Line 12--13 and line 13--23 are critical lines. Without GRS, outage of either of these two transmission lines may result in a plant-out-of-step condition. To improve the transient stability performance of this plant, a GRS is installed. When the GRS detects a line outage on either of these two lines, it trips promptly only one generator to keep the other two generators in service. The GRS logic is simple: when there is a fault on a critical line, the breakers on this line open; an «open» signal (high level signal) from any breaker energizes the output of the OR gate. The high level signal from the OR gate output, together with the high level arming signal, sets the AND gate output in high level, which is input to the 2 out of 3 voting scheme. When two or more of the voting scheme input signals are high, the voting scheme output signal is high; otherwise, it is low. The high level signal from the voting scheme will trip the selected generator. Here, it is assumed that breakers and the voting scheme are fully reliable. Breakers are external to GRS; so assuming they are 100% reliable helps to isolate the GRS influence. Their failure potential can be included in this

analysis if desired. The voting scheme is assumed fully reliable to simplify the illustration process.

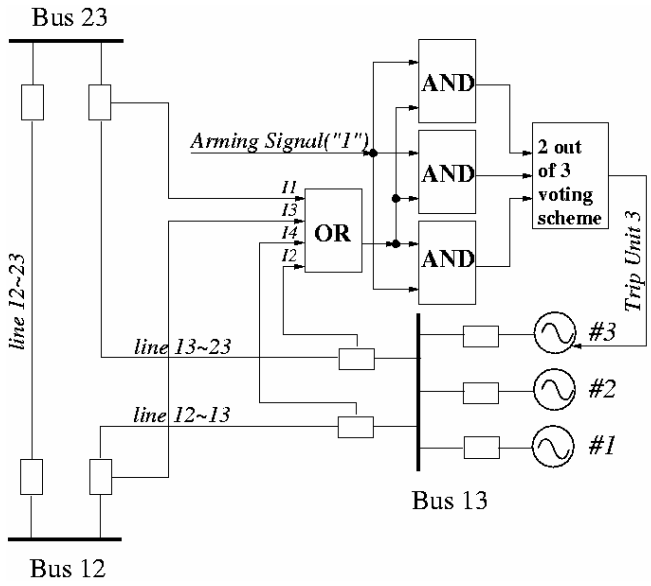

Figure A.1. GRS logic circuit and voting scheme.

1. Describe the system:

The logic diagram has been already developed as Figure A.1. There are four GRS input events, corresponding to four initiating events, as shown in Table A.1.

Signals to SPS Logic					
Event	I_1	I_2	I_3	I_4	Probability
E_1	1	1	0	0	$Pr(E_1)=Pr(F_1)Pr(\overline{F}_2)$
E_2	0	0	1	1	$Pr(E_2)=Pr(\overline{F}_1)Pr(F_2)$
E_3	0	0	0	0	$Pr(E_3)=Pr(\overline{F}_1)Pr(\overline{F}_2)$
E_4	1	1	1	1	$Pr(E_4)=Pr(F_1)Pr(F_2)$

Table A.1. Event input mapping table.

2. Complete a system level FMEA:

System states are represented by the combinations of states of all system components. Given defined modes, e.g.,

- 0 -- normal mode l.
- 1 -- failure mode 1.

- 2 -- failure mode 2.

The AND and OR gates have the following two failure modes:

- 1 -- the output of the component is "stuck" to 1.
- 2 -- the output of the component is "stuck" to 0.

Thus, the FMEA as shown in Table A.2, which also shows the assumed failure rates, is created.

Failure Mode and Effective Analysis (FMEA)			
Component	Failure mode	Failure effect	Failure rate (per day)
OR	1	Constant 1	λ_1=0.0003/365
OR	2	Constant 0	λ_2=0.02/365
AND	1	Constant 1	λ_3=0.0003/365
AND	2	Constant 0	λ_4=0.02/365

Table A.2. FMEA list for the illustration system.

3. Develop the Markov Model:

Four digits $d_1d_2d_3d_4$ are used to code the state of the system. The digit d_1 represents the state of component OR (*0*-normal, *1*-failure mode 1, *2*-failure mode 2). Digits d_2, d_3, and d_4 represent the state of the three components ANDs (*0*-normal, *1*-failure mode 1, *2*-failure mode 2). By this definition, the following 3^4=*81* states are obtained,

0000 0001 0010 0100 0002 0020 0200 1000 2000
0011 0101 0110 0021 0201 0012 0102 0210 0120
1001 1010 1100 2001 2010 2100 0022 0202 0220
1002 1020 1200 2002 2020 2200 0111 0211 0121
0112 1011 1101 1110 2011 2101 2110 0221 0212
0122 1021 1201 1012 1102 1210 1120 2021 2201
2012 2102 2210 2120 0222 1022 1202 1220 2022
2202 2220 1111 2111 1211 1121 1112 2211 2121
2112 1221 1212 1122 2212 2122 2221 1222 2222

In order to reduce the dimension of the transition matrix, the number of system states can be reduced by merging some states

as the three ANDs play the same role in the system. The criterion is: states that have identical d_1 and the same combinations of d_2, d_3 and d_4 are considered to be the same state and merged. As a result, the number of states is reduced to 30, according to the following text.

S0--0000 S1--0001, 0010, 0100 S2--0002, 0020, 0200
S3--1000 S4--2000 S5--0011, 0101, 0110
S6--0021, 0201, 0012, 0102, 0210, 0120 S7--1001, 1010, 1100
S8--2001, 2010, 2100 S9--0022, 0202, 0220 S10--1002, 1020, 1200
S11--2002, 2020, 2200 S12--0111 S13--0211, 0121, 0112
S14--1011, 1101, 1110 S15--2011, 2101, 2110 S16--0221, 0212, 0122
S17--1021, 1201, 1012, 1102, 1210, 1120
S18--2021, 2201, 2012, 2102, 2210, 2120 S19--0222
S20--1022, 1202, 1220 S21--2022, 2202, 2220 S22--1111
S23--2111 S24--1211, 1121, 1112 S25--2211, 2121, 2112
S26--1221, 1212, 1122 S27--2212, 2122, 2221 S28--1222
S29--2222

Here, $S=S0,S1,...,Sn$ represent a state space of the GRS, where Sj is a set of mutually exclusive and exhaustive states. Further, each of the above states can be classified into one of the following C_1, C_2, C_3 and C_4 categories based on the response of each system state to system input events,

- C_1 -- If the input is an active signal, then the GRS trips successfully; if the input is an inactive signal; then the GRS has a nuisance trip.
- C_2 -- If the input is an active signal, then the GRS trips successfully; if the input is an inactive signal, then the GRS does not trip.
- C_3 -- If the input is an active signal, then the GRS fails to trip; if the input is an inactive signal; then the GRS has a nuisance trip.
- C_4 -- If the input is an active signal, then the GRS fails to trip; if the input is an inactive signal, then the GRS does not trip.

For example, *S3* and *S5* are both in C_1 because when the GRS is in state *S3* or *S5*, the GRS trips successfully if the input is an active signal, and it has a nuisance trip if the input is an inactive signal. Similar thinking leads to the following:

C_1 -- S3,S5,S7,S10,S12,S13,S14,S15,S17,S22,S23,S24,S25.
C_2 -- S0, S1, S2, S6.
C_3 -- None.
C_4 -- S4,S8,S9,S11,S16,S18,S19,S20,S21,S26,S27,S28,S29.

Figure A.2 shows the preliminary Markov model for the GRS.

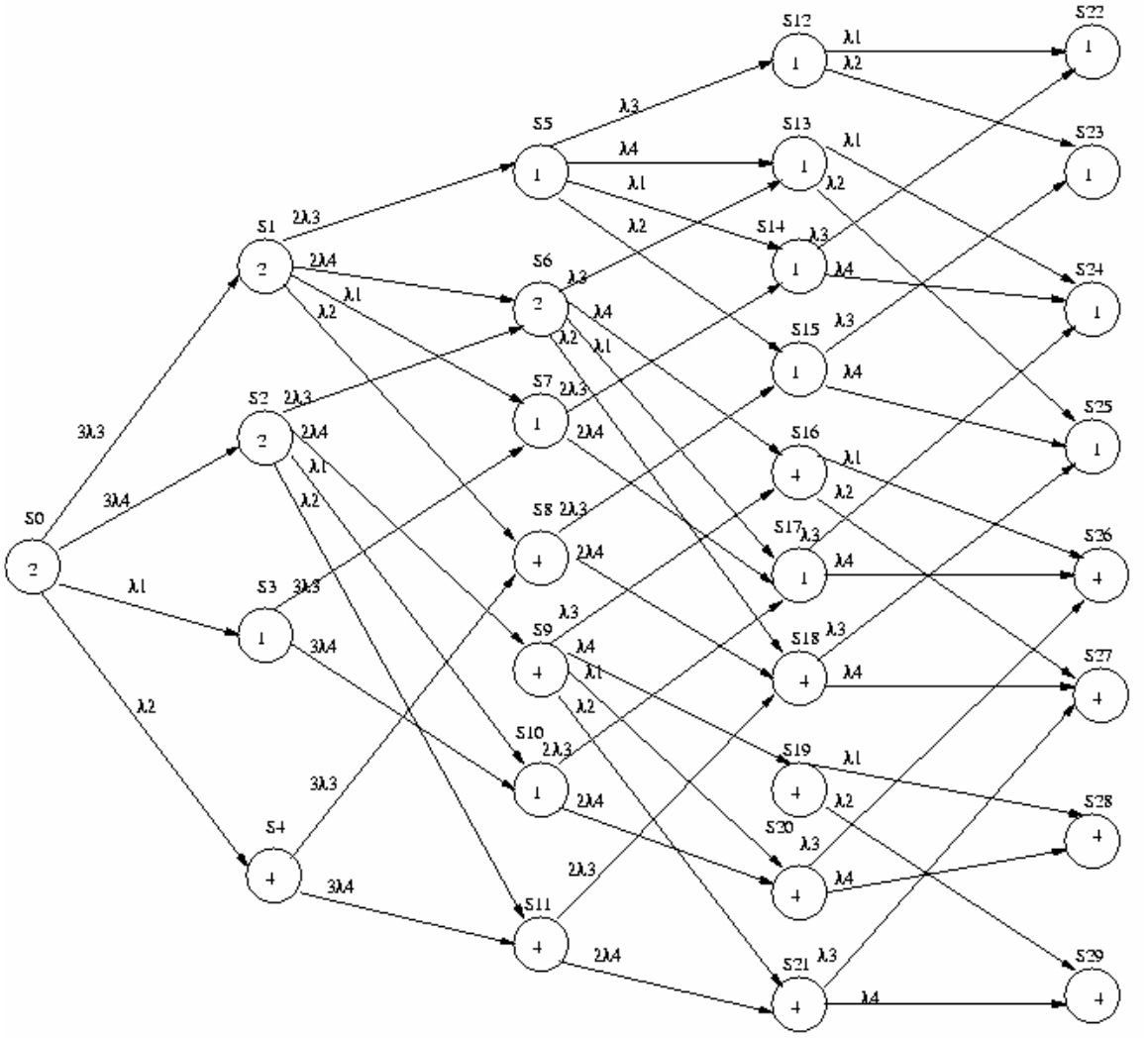

Figure A.2. The preliminary Markov model for the GRS.

4. Simplify the Markov model:

First, the following two concepts are given:

- A transition state is a state that has non-zero entry transition probability from other state(s) and non zero exit transition probability to other state(s).
- An absorbing state is a state that has a 1.0 transition probability to itself.

Then the reduction steps are as follows:

- Merge absorbing states belonging to the same class. Entry transition probabilities are added.
- For each absorbing state, eliminate all preceding states that a) are in the same class C_j as the absorbing state; b) have only one exit transition probability. Add the entry probabilities as the entry probabilities to the absorbing states.
- Merge all transition states in the same class Cj that have identical transition probabilities to common states. Entry probabilities are added. Exit probabilities remain the same.

Following the above reduction steps, the final reduction result is shown in Figure A.3. Detailed description of these reduction procedures can be found in reference [A.11].

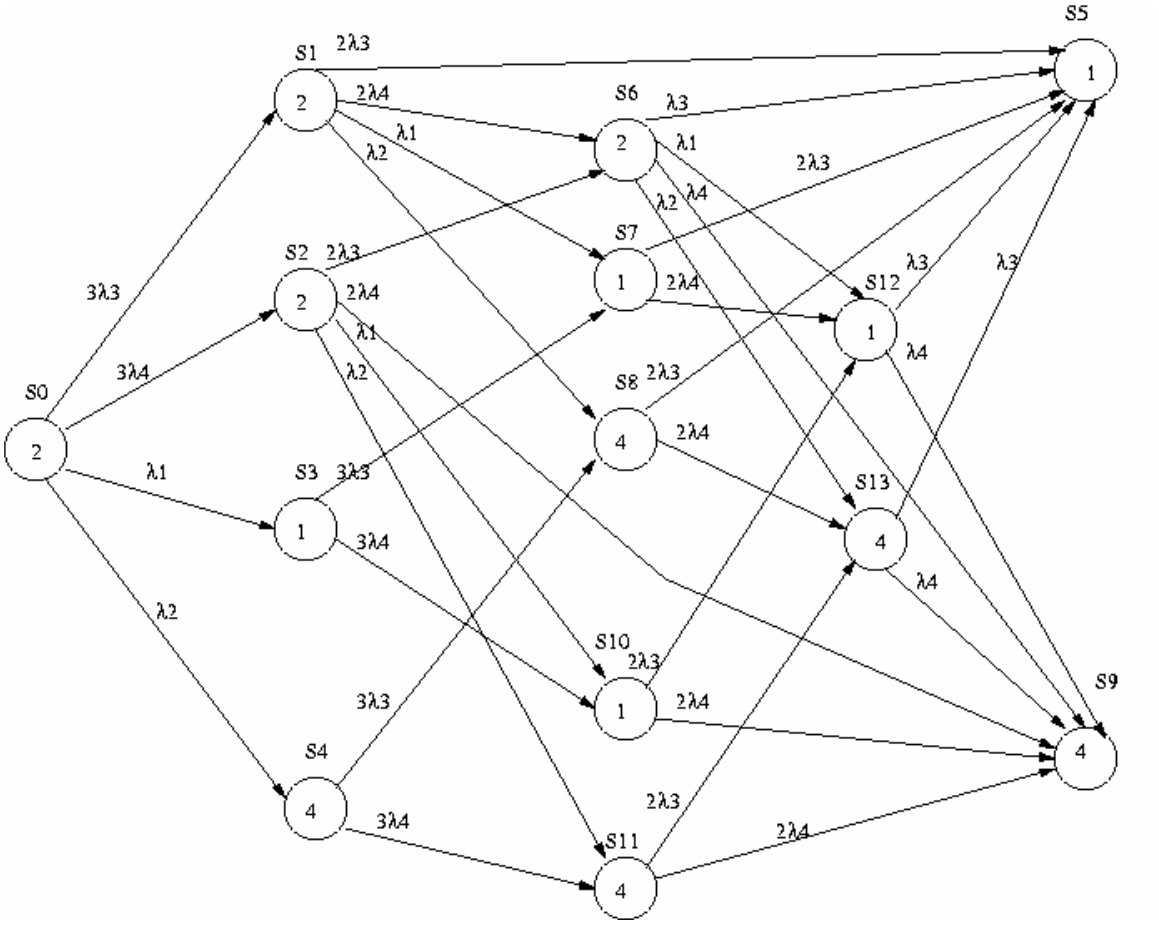

Figure A.3. The final simplification result.

5. Calculate the state probabilities:

Assume that the failure of the GRS components have exponential distributions. Therefore the pdf of component failure is $f(t) = \lambda e^{-\lambda t}$, where λ is the failure rate per unit time interval. Then the probability that the component fails before time t is

$$F(t) = \int_0^t \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t} \approx \lambda t \qquad \textbf{(eq. A.1)}$$

where the approximation improves as λt gets small. With this model, an *n+1* by *n+1* transition matrix A is obtained, where A_{ik} *(i=0,1,...,n, k=0,1,...,n)* indicates the probability that the system transfers from state *Si* to *Sk*, and n indicates the state number.

Assume the probability list at initial time $t=t_0$ is

$$\Pr^{(0)} = (\Pr(S0'(t_0)) \quad \Pr(S1'(t_0)) \quad \cdots \quad \Pr(Sn'(t_0)))$$

after *m* time intervals, the probability list is

$$\Pr^{(m)} = (\Pr(S0'(t_m)) \quad \Pr(S1'(t_m)) \quad \cdots \quad \Pr(Sn'(t_m))) = \Pr^{(0)} \times \underline{P}^m$$

The elements in the probability list $Pr^{(i)}$ provide the probability that the system is in state *Sj (j=1, ..., n)* after m time intervals. Then the following results are obtained

$$\begin{aligned}
\Pr(C_1) &= \sum \Pr(Si') \quad Si' \in C_1 \\
\Pr(C_2) &= \sum \Pr(Si') \quad Si' \in C_2 \\
\Pr(C_3) &= \sum \Pr(Si') \quad Si' \in C_3 \\
\Pr(C_4) &= \sum \Pr(Si') \quad Si' \in C_4
\end{aligned}$$

By defining the following terms,

$$\begin{aligned}
p_1 &= 1-\lambda_1-\lambda_2-3\lambda_3-3\lambda_4 \\
p_2 &= 1-\lambda_1-\lambda_2-2\lambda_3-2\lambda_4 \\
p_3 &= 1-3\lambda_3-3\lambda_4 \\
p_4 &= 1-\lambda_1-\lambda_2-\lambda_3-\lambda_4 \\
p_5 &= 1-2\lambda_3-2\lambda_4 \\
p_6 &= 1-\lambda_3-\lambda_4
\end{aligned}$$

the following state transition matrix is obtained (note that because we numbered the states 0,1,…,13, state k corresponds to the k+1 row in the below matrix).

$$\underline{P} = \begin{pmatrix}
p_1 & 3\lambda_3 & 3\lambda_4 & \lambda_1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p_2 & 0 & 0 & 0 & 2\lambda_3 & 2\lambda_4 & \lambda_1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p_2 & 0 & 0 & 0 & 2\lambda_3 & 0 & 0 & 2\lambda_4 & \lambda_1 & \lambda_2 & 0 & 0 \\
0 & 0 & 0 & p_3 & 0 & 0 & 0 & 3\lambda_3 & 0 & 0 & 3\lambda_4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_3 & 0 & 0 & 0 & 3\lambda_3 & 0 & 0 & 3\lambda_4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_3 & p_4 & 0 & 0 & \lambda_4 & 0 & 0 & \lambda_1 & \lambda_2 \\
0 & 0 & 0 & 0 & 0 & 2\lambda_3 & 0 & p_5 & 0 & 0 & 0 & 0 & 2\lambda_4 & 0 \\
0 & 0 & 0 & 0 & 0 & 2\lambda_3 & 0 & 0 & p_5 & 0 & 0 & 0 & 0 & 2\lambda_4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\lambda_4 & p_5 & 0 & 2\lambda_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\lambda_4 & 0 & p_5 & 0 & 2\lambda_3 \\
0 & 0 & 0 & 0 & 0 & \lambda_3 & 0 & 0 & 0 & \lambda_4 & 0 & 0 & p_6 & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_3 & 0 & 0 & 0 & \lambda_4 & 0 & 0 & 0 & p_6
\end{pmatrix}$$

Pr(Si(t0)) provides the probability that the system is in state j at time t=t0. It is assumed that at initial time $t=t_0$, every component is in perfect condition due to inspection or maintenance. Therefore

$$Pr^{(0)} = (\,Pr(S_0(t_0)) \quad Pr(S_1(t_0)) \quad \cdots \quad Pr(S_{13}(t_0))\,)$$
$$= (\,1 \quad 0 \quad \cdots \quad 0\,)$$

After *m* time intervals from initial time $t=t_0$, the probability list is

$$\Pr^{(m)} = (\Pr(S0(t_m)) \quad \Pr(S1(t_m)) \quad \cdots \quad \Pr(S13(t_m))) \qquad \textbf{(eq. A.2)}$$
$$= \Pr^{(0)} \times \underline{P}^m$$

For example, if the time interval is chosen as one day, the elements in the probability list $Pr^{(365)}$ provide the probability that the system is in state *Sj (j=1, ... , n)* after 365 time intervals, i.e., one year. Substituting the FMEA data in Table A.2 into (eq.A.2) gives

$$\begin{aligned}\Pr^{(365)} &= \Pr^{(0)} \times \underline{P}^{365} \\ &= (9.2200e-01 \quad 8.3845e-04 \quad 5.5896e-02 \quad 2.7948e-04 \\ &\quad 1.8632e-02 \quad 2.6384e-07 \quad 3.3792e-05 \quad 2.5344e-07 \\ &\quad 1.6896e-05 \quad 1.1574e-03 \quad 1.6896e-05 \quad 1.1264e-03 \\ &\quad 1.0186e-08 \quad 6.7907e-07)\end{aligned}$$

Since *S3*, *S5*, *S7*, *S10*, *S12* constitute C_1, *S0*, *S1*, *S2*, *S6* constitute category C_2, and *S4*, *S8*, *S9*, *S11*, *S13* constitute C_4, the following results are obtained

$$Pr(C_1) = Pr(S_3) + Pr(S_5) + Pr(S_7) + Pr(S_{10}) + Pr(S_{12}) = 2.9691e-04$$

$$Pr(C_2) = Pr(S_0) + Pr(S_1) + Pr(S_2) + Pr(S_6) = 9.7877e-01$$

$$Pr(C_3) = 0$$

$$Pr(C_4) = Pr(S_4) + Pr(S_8) + Pr(S_9) + Pr(S_{11}) + Pr(S_{13}) = 2.0933e-02$$

Figure A.4 shows how the GRS state probabilities change with time.

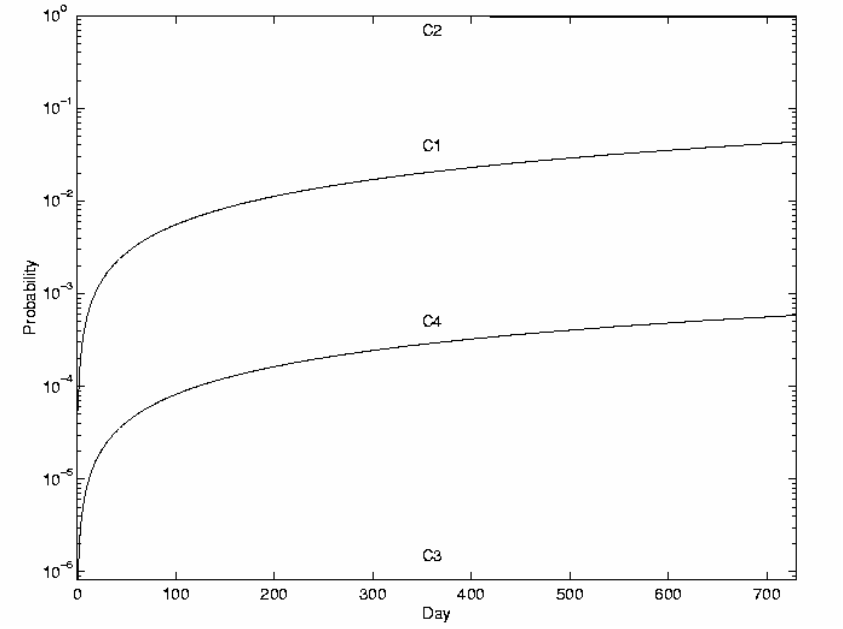

Figure A.4. GRS state probabilities.

Appendix References

[A.1] J. Findlay, G. Maria, and V. Wong, *Stability Limits for the Bruce LGR Scheme*, Proceedings of the Eighth Power Systems Computation Conference, Helsinki, August, 1984.

[A.2] J. Doudna, *Application and Implementation of Fast Valving and Generator Tripping Schemes at Gerald Gentleman Station*, IEEE Transactions on Power Systems, Vol. 3, No. 3, pp. 1155-1166, August 1988.

[A.3] C. Charman and J. Littman, *A Generation Shedding Scheme for the Jim Bridger Steam-Electric Plant*, presented at the 19th Annual Western Protective Relay Conference, Spokane, Washington, October, 1992.

[A.4] N. Bhatt, *Field Experience with Momentary Fast Turbine Valving and Other Special Stability Controls Employed at AEP's Rockport Plant*, IEEE Transactions on Power Systems, Vol. 11, No. 1, pp. 151-161, February, 1996.

[A.5] IEEE, *Tutorial on the Protection of Synchronous Generators*, IEEE PES Special Publication 95 TP 102.

[A.6] Y. Ohura, K. Matsuzawa, H. Ohtsuka, N. Nagai, T. Gouda, H. Oshida, S. Takeda and S. Nishida, *Development of a Generator Tripping System for Transient Stability Augmentation Based on the Energy Function Method*, IEEE Transactions on Power Delivery, Vol. PWRD-1, No. 3, pp. 68-77, July, 1986.

[A.7] M. Takahashi, K. Matsuzawa, M. Sato, K. Omata, R. Tsukui, T. Nakamura, and S. Mizuguchi, *Fast Generation Shedding Equipment Based On The Observation of Swings of Generators*, 87 WM 093-8, IEEE/PES 1987 winter meeting, New Orleans, Louisiana, February 1-6, 1987.

[A.8] R. Farmer, R. H. Hartley, M. H. Kent, and L. M. Wheeler, *Four Corners Project Stability Studies*, Presented at the ASME/IEEE Joint Power Generation Conference, San Francisco, California, September 15-19, 1968.

[A.9] D. Dodge, and B. Doel, *Power System Stability Control Using Fault Tolerant Technology*, Instrumentation in the Power Industry: Proceedings of the Thirty-third Power Instrumentation Symposium, pp. 217-234, May 21-23, 1990.

[A.10] IEEE Task Force Report, *The IEEE Reliability Test System - 1996*, 96 WM 326-9 PWRS.

[A.11] EPRI final report WO8604-01, *Risk-based Security Assessment*, December, 1998.