THE DOWNTIME DISTRIBUTION AFTER A FAILURE OF A SYSTEM WITH MULTISTATE INDEPENDENT COMPONENTS

ANTONIO PIEVATOLO,* IMATI-CNR

Abstract

Aven and Jensen (1999) proposed an approximation to the downtime distribution for arbitrary coherent systems with binary and independent repairable components, where each component follows an alternating renewal process. The approximation is based on a mixture of the duration distributions of the cut sets in the system. We extend this approach to coherent systems composed by multistate independent components, where every component follows a semi-Markov process and where we introduce the concept of a minimal cut set in the multistate setting. We test our proposal on a simple power demand–generation system, by comparing the analytic approximation with simulation results, and we find it is accurate when the system is highly available.

Keywords: Semi-Markov process; cut set; mixture distribution

2000 Mathematics Subject Classification: Primary 60K20
Secondary 62E17; 90B25

1. Introduction

Consider a system with any number of states, where every state is classified as acceptable (or functioning) or unacceptable (or failure). By system downtime we mean a period of time during which the system is in one of the failure states. We examine the probability distribution of the system downtime, or system downtime distribution (SDD), after a failure. The downtime is the time it takes to complete the repair of the system and bring it back to the functioning state. In particular, we consider a system that is started at time $-\infty$ and is monitored from time 0 (hence, it is found in the stationary state); the downtime is then computed directly after the first transition from the functioning state into the failure state, and we may refer to its distribution as the steady-state SDD.

The exact analytical form of the SDD can be obtained with difficulty even in relatively simple systems, because, while it is easy to calculate the probability that the system has gone down through a specific failure state, it is less easy to follow all the possible transitions among failure states the system may experience before returning to the functioning state.

Aven and Jensen [1] and Gåsemyr and Aven [3] discussed the case of a coherent system with binary components that evolve independently following alternating renewal processes, and they proposed an approximation to the steady-state SDD for highly available systems, based on the duration of the cut sets. We extend their approach to systems where every component evolves independently as a semi-Markov process, the simplest semi-Markov process being an...
Downtime distribution

alternating renewal process. Since the components are now possibly nonbinary, the concept of a cut set must also be extended. This operation was done in a similar way in [5] for a specific application. Here, we develop the theoretical side of it and specify the new cut set concept more precisely.

In Section 2 we recollect some useful results on semi-Markov processes and introduce the Weibull–Markov system (see [7]), which will be the particular system of semi-Markov processes we will use in our examples. Then, in Section 3 we provide and discuss the extension of the cut set concept, and in Section 4 we motivate the proposed approximation to the steady-state SDD, also with the help of an example. In Section 5 we reconsider the example and compare the approximation with the sample steady-state SDD obtained through a computer simulation.

2. Semi-Markov processes and the Weibull–Markov system

We denote the state of a system with \( n \) components by the vector \( x = (x_1, \ldots, x_n) \), where \( x_c \) is the state of the \( c \)th component taking values in \( 1, \ldots, m_c \). The state of the system at time \( t \) is indicated as \( x(t) = (x_1(t), \ldots, x_n(t)) \). For notational simplicity, we will use \( x, x(t), m, \) or other quantities without the index for the component when we focus on a single component only, without considering it as a part of the system.

Every component evolves according to a semi-Markov process, i.e. if the component has just entered state \( x = i \) then it will move to state \( j \) with probability \( p_{ij} \). Conditionally on the next state being \( j \), the sojourn time \( Y_{ij} \) is a random variable with cumulative density function \( F_{ij} \).

Therefore, the semi-Markov process is completely determined by the transition matrix \( P \) with elements \( p_{ij} \) and by the distributions \( F_{ij} \). We will assume throughout that these distributions are absolutely continuous, although it is not always strictly required. For this and other equivalent characterizations of semi-Markov processes, see [6].

Now, denote by \( N_t \) the process counting the number of transitions of the component up to time \( t \), denote by \( J_r \) the state of the Markov process at the \( r \)th transition, and denote by \( S_r \) the time of occurrence of the \( r \)th transition. In the following sections we will need Theorem 2.1, below; see [6, Theorem 5.17].

**Theorem 2.1.** If a Markov process is irreducible and nonlattice then

\[
\lim_{t \to \infty} P\{x(t) = j, J_{N_t+1} = i, S_{N_t+1} - t \leq z | J_0 = k\} = \frac{p_{ji}}{\mu_{jj}} \int_0^z (1 - F_{jj}(u)) \, du,
\]

where \( \mu_{jj} \) is the mean recurrence time of state \( j \).

Theorem 2.1 gives the asymptotic probability that a component is found in state \( j \), that the next state is \( i \), and that the residual life in state \( j \) is not greater than \( z \).

Another very useful result gives a way of calculating the limiting probability that the process is found in a given state, starting from any initial state. This result is stated correctly in [2, Theorem 5.22], whereas Theorem 5.16 of [6] assumes, unnecessarily, that the Markov chain \( J_r \) is aperiodic. Let \( \mu_i \) denote the mean sojourn time in state \( j \), given by

\[
\mu_i = \int_0^\infty t F_i(t) \, dt, \quad \text{with} \quad F_i(t) = \sum_{j \neq i} p_{ij} F_{jj}(t).
\]
Theorem 2.2. Suppose that $J_r$ is irreducible recurrent, and that $\pi$ is a solution to $\pi^\top P = \pi^\top$ (that is, the invariant distribution for $P$). Then, for any initial state $i$,

$$p(j) = \lim_{t \to \infty} P(x(t) = j \mid J_0 = i) = \frac{\pi_j \mu_j}{\sum_{k=1}^n \pi_k \mu_k},$$

(2.1)

provided that the distribution of the time between two successive occurrences of $j$ is nonlattice.

We also recall that $p(j) = \mu_j / \mu_{jj}$.

A particular semi-Markov process is the Weibull–Markov process, introduced in [7]. We can describe this process as follows. When the component has just entered state $x = i$, $m-1$ independent times are generated according to $m-1$ Weibull distributions with shape and rate parameters $(\beta_i, \lambda_{ij})$ for $j = 1, \ldots, m$ and $j \neq i$; the next state and the sojourn time in state $i$ are jointly identified by the minimum of these. It can be shown, by standard probability calculations, that this method yields

$$p_{ij} = \left(\frac{\lambda_{ij}}{\lambda_i}\right)^{\beta_i}, \quad \text{with} \quad \lambda_i = \left(\sum_{j \neq i} \beta_i \lambda_{ij}\right)^{1/\beta_i},$$

(2.2)

Therefore, after choosing the next state according to the transition matrix $P$, the conditional duration distribution depends on the current state only. As a special case, when $\beta_i = 1$, the Markov process (with exponentially distributed sojourn times) is obtained. Using Theorem 2.1, we can show that the transition rate from state $i$ to state $j$ is given by

$$\left(\frac{\lambda_{ij}}{\lambda_i}\right)^{\beta_i} \lambda_i / \Gamma(1 + 1/\beta_i),$$

which reduces to $\lambda_{ij}$ when $\beta_i = 1$.

A system of components which evolve independently as Weibull–Markov processes is called a Weibull–Markov system. We will consider this particular process in our examples below.

3. Cut sets for multistate coherent systems

For a system with $n$ components, let us consider a set of pairs

$$K_h = \{(c, x_c), \ c \in I_h\},$$

(3.1)

where $I_h$ denotes a set of indexes chosen from $\{1, \ldots, n\}$. The cardinality of $I_h$ can be less than $n$. Let us also assume that our system is coherent, so that there exists an ordering relationship ‘$\prec$’ of the states of the components such that the structure function of the system is nondecreasing.

Let us denote by $x'_c$ the smallest state $x'_c$ such that $x'_c \preceq x_c$.

Definition 3.1. We call $K_h$ in (3.1) a cut set if, when it occurs, it causes the system to fail. We call $K_h$ a minimal cut set if it is not a cut set anymore after changing $x_c$ to $x'_c$ or after removing $(c, x_c)$ from it for any $c \in I_h$.

In the binary case, we are back to the usual concept of a minimal cut set.

We notice that this definition does not coincide with the definition of a maximal lower vector for level 0 (that is, the failure state of the system in our case) given, for example, in [4]. (Note that a maximal lower vector for level 0 is a state $x$ of the system such that the system reaches level 1
A minimal cut set with \( n \) pairs in it, which corresponds to a particular system state, is a maximal lower vector for level 0. Now let us turn to minimal cut sets with less than \( n \) pairs; suppose, for example, that there are 3 components, and let our minimal cut set be \( K = \{(1, x_1^1), (2, x_2^2)\} \). It spans all the states of the form \((x_1^1, x_2^2, x_3)\), with \( x_3 = 1, \ldots, m_3 \). As \( x_3 \leq m_3 \), no state \((x_1^1, x_2^2, x_3)\) is a maximal lower vector of level 0, because otherwise the state \((x_1^1, x_2^2, x_3^+)\) would be a functioning state and \( K \) would no longer be a cut set. When \( x_3 = m_3 \), the state \((x_1^1, x_2^2, m_3)\) must be a maximal lower vector, because, if it were not, either the state \((x_1^+, x_2^2, m_3)\) or the state \((x_1^1, x_2^+, m_3)\) would be failure states, indicating that \( K \) is not minimal. Hence, there are as many minimal cut sets as maximal lower vectors for level 0 in the system.

The minimal cut set can be seen as a surrogate for the set of system states that share it, including one maximal lower vector for level 0 among these. Accordingly, we will surrogate the downtime with the lifetime of the cut set or the cut sets that have caused the failure. In fact, when any single component of the cut set improves its state, a transition to the system functioning state may take place. Hence, the potential number of such transitions is larger than that from the associated maximal lower vector of level 0, as we will show in the following example.

**Example 3.1.** (Electrical power demand and generation.) From [4] we also borrow an example of a small two-component system describing the joint evolution of generation and demand of electrical power. Component 1 is the power requested, and component 2 is the power generated, both discretized at three levels and forming a Weibull–Markov system. The demand level is \( x_1 = 1, x_1 = 2, \) and \( x_1 = 3 \), corresponding to 90 MW, 60 MW, and 5 MW, respectively; the generation level is \( x_2 = 1, x_2 = 2, \) and \( x_2 = 3 \), corresponding to 0 MW, 70 MW and 100 MW, respectively. The system is in a failure state if \( x_1 > x_2 \), which means that the requested power is greater than the generated power. In Figure 1 we show the transition diagrams for the demand, generation, and the system.
the generation components, and for the whole system. The demand component follows a Markov process, whereas the generation component follows a Weibull–Markov process; the parameter values will be specified in Section 5, below.

Minimal cut sets according to Definition 3.1 are \( K_1 = \{(1, 1), (2, 2)\} \) and \( K_2 = \{(2, 1)\} \). The elements of the cut sets are always two-dimensional vectors and must not be confused with the states of the system (which are also two-dimensional vectors in this specific case). The cut set \( K_1 \) corresponds to state (1, 2) in Figure 1, whereas cut set \( K_2 \) occurs whenever one of the states in the first column occurs. The two maximal lower vectors for level 0 are system states (3, 1) and (1, 2). We may observe that, from the set of states sharing cut set \( K_2 \), two transitions to the functioning state are possible, whereas there is only one direct link from (3, 1) (the maximal lower vector for level 0) to it.

4. The approximation of the steady-state SDD after a failure

Consider the collection of all the minimal cut sets of the system, \( K_h, h = 1, \ldots, k \). When a minimal cut set occurs at time \( t \), there is a system failure. Conversely, if the system fails at time \( t \), at least a minimal cut set must have occurred. We can write this equivalence of events as

\[
\{\Delta N_t = 1\} = \{\Delta K_{1,t} = 1\} \cup \cdots \cup \{\Delta K_{k,t} = 1\},
\]

where \( N_t \) and \( K_{h,t} \) are now the counting processes of system failures and of occurrences of \( K_h \) in the interval \([0, t]\), respectively, and \( \{\Delta N_t = 1\} \) and \( \{\Delta K_{h,t} = 1\} \) mean that such processes experience a jump at time \( t \). Let \( Y(t) \) denote the repair time of the system after its failure at \( t \). As the collection of events on the right-hand side of this equivalence is not a partition of \( \{\Delta N_t = 1\} \), we may write the following inequality for the SDD after a failure:

\[
P(Y(t) \leq y \mid \Delta N_t = 1) \leq \sum_{h=1}^{k} P(Y(t) \leq y \mid \Delta K_{h,t} = 1, \Delta N_t = 1) P(\Delta K_{h,t} = 1 \mid \Delta N_t = 1).
\] (4.1)

If the system is highly available, we may assume that two minimal cut sets will occur together with very low probability and we may expect to exit the minimal cut set that caused the failure before any other component moves to lower states. Hence, we approximate

\[
P(Y(t) \leq y \mid \Delta K_{h,t} = 1, \Delta N_t = 1) \approx P(Y(t) \leq y \mid \Delta K_{h,t} = 1, \Delta N_t = 1, K_{h,t} \text{ considered in isolation}).
\]

The weights appearing in the mixture distribution of (4.1) may be regarded as

\[
\lim_{\varepsilon \to 0^+} \frac{P(K_{h,t} + \varepsilon - K_{h,t} = 1, N_t + \varepsilon - N_t = 1)}{P(N_t + \varepsilon - N_t = 1)} = \lim_{\varepsilon \to 0^+} \frac{P(K_{h,t} + \varepsilon - K_{h,t} = 1)}{P(N_t + \varepsilon - N_t = 1)}.
\]

For the denominator, the following inequality holds:

\[
P(N_t + \varepsilon - N_t = 1) \leq \sum_{h=1}^{k} P(N_t + \varepsilon - N_t = 1 \mid K_{h,t} + \varepsilon - K_{h,t} = 1) P(K_{h,t} + \varepsilon - K_{h,t} = 1)
\]

\[
= \sum_{h=1}^{k} P(K_{h,t} + \varepsilon - K_{h,t} = 1).
\]
Finally,

\[ P(\Delta K_{h,t} = 1 \mid \Delta N_t = 1) \geq \lim_{\varepsilon \to 0^+} \frac{P(K_{h,t+\varepsilon} - K_{h,t} = 1) / \varepsilon}{\sum_{h=1}^{k} P(K_{h,t+\varepsilon} - K_{h,t} = 1) / \varepsilon} =: w_{h,t}, \]

which is the frequency of occurrences of the cut set \( K_h \) at time \( t \), normalized by the sum of such frequencies for all the minimal cut sets.

The approximation to \( P(Y(t) \leq y \mid \Delta N_t = 1) \) is therefore the following mixture of minimal cut set duration distributions:

\[ P(Y(t) \leq y \mid \Delta N_t = 1) \simeq \sum_{h=1}^{k} w_{h,t} G_{K_{h}}(y). \]

At this point we do not know whether we are approximating the left-hand side term from above or from below, because the cumulative distribution functions are approximations from above, and the weights are approximations from below. Taking the limit as \( t \) tends to \( \infty \), we obtain the following steady-state approximation:

\[ \lim_{t \to \infty} P(Y(t) \leq y \mid \Delta N_t = 1) \simeq \sum_{h=1}^{k} w_{h} G_{K_{h}}(y). \quad (4.2) \]

The form of \( G_{K_{h}}(y) \) is derived in a similar way, i.e. as a mixture of duration distributions where, for any given \( c \in I_h \), the \( c \)th distribution is the distribution of the duration of the cut set, given that it has been caused by component \( c \) taking state \( x_c \). The associated weight is the probability that component \( c \) has caused the entry into the cut set. In Section 2 we made the hypothesis that the \( F_{ij} \) distributions are absolutely continuous; hence, an entry into a cut set through the simultaneous change of state of two or more components cannot occur. This reasoning leads to identifying the \( c \)th component of the mixture as the steady-state distribution of the minimum among

1. the life of component \( c \) in state \( x_c \);
2. the residual lives of the other components, in the state they are found in, when component \( c \) moves to state \( x_c \).

Then we obtain

\[ G_{K_{h}}(y) = \sum_{c \in I_h} q_c \left\{ 1 - (1 - F_{x_c}^{(c)}(y)) \prod_{j \in I_h, j \neq c} \frac{1}{\mu_{x_j}} \int_{y}^{\infty} (1 - F_{x_j}^{(j)}(u)) \, du \right\}, \quad (4.3) \]

where \( q_c \) is the probability that component \( c \) has caused the entry into cut set \( h \), and between the braces we see the steady-state duration distribution of cut set \( h \) conditional on this event. The additional superscripts, \( (j) \) and \( (c) \), help identify the component. The distribution

\[ \frac{1}{\mu_{x_j}} \int_{y}^{\infty} (1 - F_{x_j}^{(j)}(u)) \, du \quad (4.4) \]

is derived with the help of Theorem 2.1 (where we must allow \( j =: x_j \)), by summing over all the values of \( i \neq j \) and using the fact that \( p(j) = \mu_{j} / \mu_{jj} \).
With regards to $q_c$, this is, in a similar way as $w_h$ above, but without approximation, the steady-state frequency of occurrences of state $x_c$ for component $c$, normalized by the sum of such frequencies for all the components of $K_h$. Using Theorem 2.1 again, where we let $z$ tend to $0$ and divide by $z$, we find that the steady-state frequency of state $x_c$ for component $c$ is

$$\frac{p_c(x_c)}{\mu^{(c)}_{x_c}} \prod_{j \in I_h, j \neq c} p_j(x_j),$$

so that

$$q_c = \frac{(p_c(x_c)/\mu^{(c)}_{x_c}) \prod_{j \in I_h, j \neq c} p_j(x_j)}{\sum_{r \in I_h} (p_r(x_r)/\mu^{(r)}_{x_r}) \prod_{j \in I_h, j \neq r} p_j(x_j)} = \frac{1/\mu^{(c)}_{x_c}}{\sum_{r \in I_h} (1/\mu^{(r)}_{x_r})}. \tag{4.5}$$

We also see that the denominator of the first nonsimplified expression for $q_h$ is the steady-state frequency of occurrence of the cut set $K_h$, so

$$w_h = \frac{\sum_{r \in I_h} (p_r(x_r)/\mu^{(r)}_{x_r}) \prod_{j \in I_h, j \neq r} p_j(x_j)}{\sum_{h=1}^k \sum_{r \in I_h} (p_r(x_r)/\mu^{(r)}_{x_r}) \prod_{j \in I_h, j \neq r} p_j(x_j)}. \tag{4.6}$$

The $p_r(x_r)$ terms are calculated by means of Theorem 2.2.

5. Testing the approximation with a generation–demand system

We consider the two-component Weibull–Markov system of the example introduced in Section 3. All the quantities needed to calculate the approximation to the steady-state SDD of (4.2), with its expansions in (4.3), (4.5), and (4.6), are derived from the expressions that identify the transition kernel of the Weibull–Markov process for each component in (2.2). For example, $F^{(c)}_{x_c}$ is a Weibull distribution with parameter $(\beta_{x_c}^{(c)}, \lambda_{x_c}^{(c)})$ and $p_c(x_c)$ follows from (2.1) specialized to the Weibull–Markov process. The only quantity that does not have a simple analytic expression is the integral in (4.4). Omitting the index of the component and letting it take state $i$, we have

$$\frac{1}{\mu_i} \int_0^\infty \exp(-\lambda_i u)^{\beta_i} \, du = \frac{1}{\Gamma(1/\beta_i)} \int_0^\infty \exp(-u)u^{1/\beta_i-1} \, du,$$

i.e. a normalized incomplete gamma function, remembering that $\mu_i = \Gamma(1 + 1/\beta_i)/\lambda_i$.

Lisnianski and Levitin [4] considered a Markov process for both components. We adapt the numerical values of their transition rates to our model. The transition matrix they gave for the demand is

$$P_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 7/2 & 5/2 & 0 \end{bmatrix},$$

and then they gave the expected duration in each state, being 8 hours for states 1 and 2 and 16 hours for state 3. For the demand, we keep the Markov model, so that the hazards for states 1 and 2 are

$$\lambda_1^{(1)} = \lambda_2^{(1)} = \frac{24 \cdot 370}{8} = 1110 \text{ year}^{-1},$$

where we have kept a year with 370 days as in [4]. Similarly,

$$\lambda_3^{(1)} = \frac{24 \cdot 370}{16} = 555 \text{ year}^{-1}.$$
For the generation component (that is, component 2) transition rates are given as follows:
\[
\begin{align*}
\lambda_{12}^{(2)} &= 300 \text{ year}^{-1}, \\
\lambda_{21}^{(2)} &= 1 \text{ year}^{-1}, \\
\lambda_{23}^{(2)} &= 1000 \text{ year}^{-1}, \\
\lambda_{32}^{(2)} &= 5 \text{ year}^{-1},
\end{align*}
\]
so that ‘repair rates’ are much larger than ‘failure rates’. From these rates, we have derived the following transition matrix, using (2.2) with \(\beta_2 = 1\):
\[
P_2 = \begin{bmatrix}
0 & 1 & 0 \\
\frac{1}{1000} & 0 & 1 \\
0 & 1 & 0
\end{bmatrix}.
\]

In order to exploit our Weibull–Markov process, we assign a Weibull distribution to the repair time, with \(\beta_1^{(2)} = 0.5\), \(\beta_2^{(2)} = 1.5\), and \(\beta_3^{(2)} = 1\). This implies new \(\lambda_i^{(2)}\) rates, obtained by equating the means \(\mu_i^{(2)}\) of the Weibull–Markov process with the corresponding means given for the original Markov process. In other words, we solve the following equations for \(\lambda_i^{(2)}\):
\[
\mu_i^{(2)} = \frac{1}{\lambda_i^{(2)}} \Gamma \left( 1 + \frac{1}{\beta_i^{(2)}} \right) = \frac{1}{\sum_{j \neq i} \lambda_{ij}^{(2)}} \text{ for all } i,
\]

obtaining \(\lambda_1^{(2)} = 600\), \(\lambda_2^{(2)} = 900\), and \(\lambda_3^{(2)} = 5\).

For a comparison with the analytic approximation, we simulated 10 000 independent down times, by starting 10 000 independent copies of the Weibull–Markov system in the steady state, and recording for each copy the first transition from the functioning state into the failed state and the subsequent time to repair.

The first comparison between the empirical survival function of the downtime and the analytical approximation of (4.2) is reported in Figure 2, where we can see that the curves are virtually indistinguishable. The 99.9% empirical quantile is 0.004, whereas from our approximation we obtain a close 0.003; the 99% quantiles coincide and are equal to 0.002. In Figure 3 we also show the two components of the mixture. The curve with the heavier tail is relative to the duration of \(K_2 = \{(2, 1)\}\), i.e. no power generation with Weibull-distributed repair time, but it has very little influence because the weights in the mixture are \(w_1 = 0.995\) and \(w_2 = 0.005\).

![Figure 2: Survival function of the steady-state SDD for the demand–generation system: empirical distribution from the simulation (dashed line) and approximation (solid line).](https://www.cambridge.org/core/terms).
Figure 3: Mixture components of the survival function of Figure 1.

Figure 4: Approximation to the survival function of the steady-state SDD with its two components (solid line) and the empirical counterpart (dashed line) for the generation–demand system when $p_{21} = \frac{1}{10}$ and $p_{23} = \frac{9}{10}$, with the other parameters unchanged.

$w_2 = 0.005$: indeed, in the simulation 99.4% of all the entries into the failure state occurred through cut set $K_1$, i.e. state $(1, 2)$. Moreover, because of the very high repair rates of the generation, state $(1, 1)$ is not easily reached from state $(1, 2)$.

Therefore, we tried to obtain a less reliable generation by modifying $P_2$ and letting $p_{21} = \frac{1}{10}$ and $p_{23} = \frac{9}{10}$. The result is shown in Figure 4, where, looking closely, we may see a little more distance between the approximation and the empirical distribution.

A further step towards a less reliable system can be obtained by trying to make the state of no power generation even more easily reachable and the sojourn in it longer. In this way cut set $K_1$ will be entered more often, and the exit from it will not correspond to the exit from the failure state whenever a visit to state $(1, 1)$ occurs. Therefore, we let $\lambda_1^{(2)} = 60$, which implies a lower repair rate for the generation, and we let $p_{21} = p_{23} = \frac{1}{2}$. The result is shown
in Figure 5, where we note that the empirical survival function has a higher upper tail than the approximation, a feature which can be ascertained with difficulty in Figure 4.

6. Conclusions

We have proposed a new concept for minimal cut sets in systems with multistate components, and we have found a useful approximation to the steady-state SDD after a failure using such a concept. A first experiment showed that the approximation behaves quite well, and we believe it would work for more complex systems as well, provided failure states, which are very deep into the set of system failure states, are reached with low probability and have relatively short duration.

Acknowledgement

I would like to thank Ivan Valadè for suggesting the idea of the new type of minimal cut sets.

References