Bounding Steady-state Availability Models with Phase Type Repair Distributions[*]

Juan A. Carrasco
Departament d'Enginyeria Electrònica
Universitat Politècnica de Catalunya
Diagonal 647, plta. 9, 08028 Barcelona, Spain
carrasco@eel.upc.es

Abstract

We propose a method to obtain bounds for the steady-state availability using Markov models in which only a small portion of the state space is generated. The method is applicable to models with phase type repair distributions and involves the solution of only 4 linear systems of the size of the generated state space, independently on the number of "return" states. A numerical example is presented to illustrate the method.

1 Introduction

A major drawback of continuous-time Markov chain (CTMC) models is that they usually have state space cardinalities which are far beyond the available computational resources. An approach which has been developed in the last few years is the use of bounding methods which require the generation of only a portion of the state space [2], [3], [4], [5], [7], [8], [9], [11], [15], [16]. Those methods perform well when, as in the case of availability models, the probability mass is concentrated in a small portion of the state space. The first of such methods was developed by Muntz, Souza e Silva and Goyal [11] using results from Courtois and Semal concerning bounds for conditional steady-state distributions in subsets of Markov chains [4], [5]. Let N be the number of components of the system. Denoting by C_k, $0 \leq k \leq N$, the subset of states with exactly k failed components, by G the subset of generated states, and by U the subset of non-generated states, the basic method proposed in [11] takes $G = \cup_{0 \leq k \leq K} C_k$ and bounds the behavior in U using a submodel with states c_k, $K < k \leq N$ associated to the subsets C_k. This basic method requires the solution of $|C_K|$ linear systems of size $|G| + N - K$, which is typically very costly. In order to reduce the computational cost of the method a state cloning technique is developed in [11] which introduces some looseness in the bounds but reduces the number of linear system to be solved to $|C_F|$, where

$0 \leq F < K$. Lui and Muntz [7] have proposed a refinement of the method for the particular case $F = 0$ including a reuse technique which, at the price of an additional looseness in the bounds, avoids a complete reapplication of the method each time K is incremented in the search for the desired accuracy. The additional looseness has been reduced in another paper from the same authors [8]. Souza e Silva and Ochoa [16] have developed state space exploration techniques in which G is generated incrementally following heuristics which try to obtain the tightest possible bounds for a given number of generated states. Semal has developed recently [15] a bounding method which refines iteratively the bounds using detailed knowledge about the model in U in the proximities of G. In [2] a bounding method is developed which exploits the failure distance concept to bound the behavior in U more tightly than in [11]. State space exploration techniques specifically targeted to that bounding method have also been developed [3]. Finally, the method described in [11] has been extended in [9] to models with infinite state spaces and subsets C_k, $k > K$ in which no every state has a transition to the left (subset C_{k-1}). Performance models were considered in [9] and the bounding part of the model was found using special developments for the models under consideration.

All previous methods to bound the steady-state availability assume exponential repair time distributions (the only exception being the machine repair model considered in [10], an extended version of [9], but the developments were specific for the considered model). In this paper we develop a new bounding method for a large class of models of repairable fault-tolerant systems with phase type repair time distributions. The method generates the subset of states $G = \cup_{0 \leq k \leq K} C_k$ and computes the bounds without using state cloning techniques by solving only 4 linear systems of size $|G|$. The rest of the paper is organized as follows. Section 2 describes the type of models considered. Section 3 describes the bounding method. Section 4 illustrates the method with a numerical example. Section 5 concludes the paper.

2 Type of models and assumed knowledge

We consider fault-tolerant systems made up of components which fail and are repaired. Failure processes have exponential distributions; repair processes have phase type distributions [12]. Components are grouped into classes, being indistinguishable the components of the same class, and thus collections of components will be bags (see, for instance [13] for a brief summary of bag theory). Any bag of components which can fail simultaneously will be called *failure bag*. Let E be the set of failure bags of the model. In general, failure bags may occur with rates which depend on the bag of failed components. We will assume known E and, for each $e \in E$, an upper bound $[\lambda(e)]_{ub}$ for its rate. Repair events are assumed to involve only one component. Each repair event has a repair time distribution taken from a set $\mathcal{P} = \{P_i, i = 1, \ldots, L\}$ of phase type distributions. Each phase type distribution P_i is defined by a transient CTMC $Z_i = \{Z_i(t), t \geq 0\}$ with finite state space $L_i \cup \{a\}$, where all states in L_i are transient, a is an absorbing state and $P[Z_i(0) \in L_i] = 1$: the repair time is the time to absorption of Z_i. We allow repair interruption. Thus, the failure of a component of higher repair priority may preempt an undergoing repair process; the repair process may be resumed later from the point it was stopped (preemptive-resume) or retaken as it had just started (preemptive-restart). The state of the system can then be completely described by giving the bag of failed components $\mathcal{F}$, the bag of failed components assigned to repair processes $\mathcal{R}$, the subbag of failed components which are under active repair $\mathcal{A} \subset \mathcal{R}$ (those for which repair is in progress), and, for each $s \in \mathcal{R}$, the phase type distribution $P_i \in \mathcal{P}$ associated with it and the state $a \in L_i$ of the corresponding transient CTMC Z_i.

We will denote by $X = \{X(t), t \geq 0\}$ the resulting CTMC model and by Ω its state space. Let N be the number of components of the system and let C_k be the subset of Ω including the states with k failed components. As in [11] we will take $G = \cup_{0 \leq k \leq K} C_k$ and, accordingly, $U = \cup_{K < k \leq N'} C_k$, where $K < \bar{N}' \leq N$. According to the assumed type of state description, we will have $|C_0| = 1$ and will denote by o the only state belonging to C_0. We will assume that some repair process is active in every state with failed components. Thus, o will be the only state without active repair processes and X will be irreducible.

3 Bounding method

3.1 Preliminaries

Although our bounding method is mainly addressed to the computation of bounds for the steady-state availability, it can, in fact, be used to bound any steady-state reward rate measure. Let r_i, $i \in \Omega$ be an arbitrary reward rate structure defined over X. We are interested in bounding the steady-state reward rate:

$$R = \lim_{t \to \infty} E[r_{X(t)}] = \sum_{i \in \Omega} r_i p_i,$$

where $p_i = \lim_{t \to \infty} P[X(t) = i]$. The steady-state availability is a particular case of R in which $r_i = 1$ for the up (operational) states and $r_i = 0$ for the down (non-operational) states. Let $\mathbf{p} = (p_i)_{i \in \Omega}$ be the steady-state probability vector of X. Let $S = C_K$ be the subset of states in G which may have some transition from U (the so-called "return" subset), and for each $s \in S$ consider the CTMC $X_s = \{X_s(t), t \geq 0\}$ obtained from X by redirecting to s all transitions from U to S. From the fact that state o is reachable from s it is easy to prove that X_s is irreducible. Let $\mathbf{p}_s = ((p_s)_i)_{i \in \Omega}$ be the steady-state probability vector of X_s. Using general results from Courtois and Semal [4, Theorem 3], [5, Section III] concerning bounds for left eigenvectors of nonnegative irreducible matrices we have:

Theorem 1 *There exist* β_s, $s \in S$ *with* $\beta_s \geq 0$, $\sum_{s \in S} \beta_s = 1$ *such that* $\mathbf{p} = \sum_{s \in S} \beta_s \mathbf{p}_s$.

As an immediate consequence, if we call R_s the steady-state reward rate of X_s, i.e.

$$R_s = \lim_{t \to \infty} E[r_{X_s(t)}] = \sum_{i \in \Omega} r_i (p_s)_i,$$

we have:

Corollary 1 $\min_{s \in S} \{R_s\} \leq R \leq \max_{s \in S} \{R_s\}$.

Proof Using Theorem 1:

$$
\begin{aligned}
R &= \sum_{i \in \Omega} r_i p_i = \sum_{i \in \Omega} r_i \sum_{s \in S} \beta_s (p_s)_i \\
&= \sum_{s \in S} \beta_s \sum_{i \in \Omega} r_i (p_s)_i = \sum_{s \in S} \beta_s R_s.
\end{aligned}
$$

Then, since $\beta_s \geq 0$ and $\sum_{s \in S} \beta_s = 1$, $R \leq \sum_{s \in S} \beta_s \max_{s \in S} \{R_s\} = \max_{s \in S} \{R_s\}$. Similarly, we have $R \geq \min_{s \in S} \{R_s\}$. $\square$

Corollary 1 allows us to compute bounds for R from bounds for R_s, $s \in S$:

$$[R]_{lb} = \min_{s \in S} \{[R_s]_{lb}\}, \qquad (1)$$

$$[R]_{ub} = \max_{s \in S} \{[R_s]_{ub}\}. \qquad (2)$$

Consider the regenerative behavior of X_s with $X_s(0) = s$ defined by the times at which X_s hits s from U. Let T_s and C_s be, respectively, the mean time and mean reward of X_s between regenerations. From regenerative process theory (see, for instance, [14]) we have $R_s = C_s / T_s$. Let $T_{G,s}$ and $T_{U,s}$ ($C_{G,s}$ and $C_{U,s}$) be the contributions of, respectively, the states in G and U to T_s (C_s). We have:

$$R_s = \frac{C_{G,s} + C_{U,s}}{T_{G,s} + T_{U,s}}.$$

Assume that $C_{G,s}$, $T_{G,s}$, an upper bound $[T_{U,s}]_{\text{ub}}$ for $T_{U,s}$, and lower and upper bounds $[r]_{\text{lb}}$ and $[r]_{\text{ub}}$ for r_i, $i \in \Omega$ are known (for the steady-state availability we would take $[r]_{\text{lb}} = 0$ and $[r]_{\text{ub}} = 1$). We have:

Theorem 2

$$[R_s]_{\text{lb}} = \frac{C_{G,s} + [r]_{\text{lb}}[T_{U,s}]_{\text{ub}}}{T_{G,s} + [T_{U,s}]_{\text{ub}}} \leq R_s$$

$$\leq \frac{C_{G,s} + [r]_{\text{ub}}[T_{U,s}]_{\text{ub}}}{T_{G,s} + [T_{U,s}]_{\text{ub}}} = [R_s]_{\text{ub}}. \qquad (3)$$

Proof Consider the function $f_1(x) = (C_{G,s} + [r]_{\text{ub}}x)/(T_{G,s} + x)$. Since $[r]_{\text{ub}}$ upper bounds the reward rate from any state of X_s, we have $C_{G,s} \leq [r]_{\text{ub}}T_{G,s}$ and $df_1/dx = ([r]_{\text{ub}}T_{G,s} - C_{G,s})/(T_{G,s} + x)^2 \geq 0$. Also, $C_{U,s} \leq [r]_{\text{ub}}T_{U,s}$. Then:

$$\begin{aligned} R_s &\leq \frac{C_{G,s} + [r]_{\text{ub}}T_{U,s}}{T_{G,s} + T_{U,s}} = f_1(T_{U,s}) \leq f_1([T_{U,s}]_{\text{ub}}) \\ &= \frac{C_{G,s} + [r]_{\text{ub}}[T_{U,s}]_{\text{ub}}}{T_{G,s} + [T_{U,s}]_{\text{ub}}}. \end{aligned}$$

Similarly, consider the function $f_2(x) = (C_{G,s} + [r]_{\text{lb}}x)/(T_{G,s} + x)$. Since $[r]_{\text{lb}}$ lower bounds the reward rate from any state of X_s, we have $C_{G,s} \geq [r]_{\text{lb}}T_{G,s}$ and $df_2/dx = ([r]_{\text{lb}}T_{G,s} - C_{G,s})/(T_{G,s} + x)^2 \leq 0$. Also, $C_{U,s} \geq [r]_{\text{lb}}T_{U,s}$. Then:

$$\begin{aligned} R_s &\geq \frac{C_{G,s} + [r]_{\text{lb}}T_{U,s}}{T_{G,s} + T_{U,s}} = f_2(T_{U,s}) \geq f_2([T_{U,s}]_{\text{ub}}) \\ &= \frac{C_{G,s} + [r]_{\text{lb}}[T_{U,s}]_{\text{ub}}}{T_{G,s} + [T_{U,s}]_{\text{ub}}}. \quad \square \end{aligned}$$

3.2 Derivation of $[T_{U,s}]_{\text{ub}}$

In the remainder of the paper we will denote by λ_{ij}, $i, j \in \Omega$ the transition rate from state i to state j, by $\lambda_i = \sum_{j \in \Omega} \lambda_{ij}$, $i \in \Omega$ the output rate of state i, and by $\lambda_{iC} = \sum_{j \in C} \lambda_{ij}$, $i \in \Omega$, $C \subset \Omega$ the transition rate from i to subset C, all of them referred to X, unless otherwise stated. We will also consider a number of transient CTMCs Y. Each CTMC Y has a state space of the form $B \cup \{a\}$, where all states in B are transient and a is an absorbing state, and has a well-defined initial probability distribution with $P[Y(0) \in B] = 1$. We will denote by $\tau(i, Y)$, $i \in B$ the mean time spent by Y in i before absorption ($\tau(i, Y) = \int_0^\infty P[Y(t) = i]dt$). We will also use the notation $\tau(C, Y) = \sum_{i \in C} \tau(i, Y)$. It is well-known (see, for instance, [1]) that the mean times to absorption vector $\boldsymbol{\tau} = (\tau(i, Y))_{i \in B}$ is the solution of the linear system $\boldsymbol{\tau}^T \mathbf{A}_B = -\mathbf{q}^T$, where $\mathbf{A}_B$ is the restriction of the transition rate matrix of Y to B and $\mathbf{q} = (P[Y(0) = i])_{i \in B}$. The expected number of times that a transition (i, j) with rate λ_{ij}

is followed is $\mu_{ij} = \tau(i, Y)\lambda_{ij}$. The result follows easily: $\mu_{ij} = \int_0^\infty P[Y(t) = i]\lambda_{ij}dt = \lambda_{ij}\int_0^\infty P[Y(t) = i]dt = \lambda_{ij}\tau(i, Y)$. It can be similarly shown that, given a partition $B \cup B^c$ of the state space of X and assuming $X(0) \in B$, the probability that X enters B^c through a state $j \in B^c$ is $\sum_{i \in B} \tau(i, Y_B)\lambda_{ij}$, where Y_B is the transient CTMC tracking X till exit of B (Y_B has state space $B' \cup \{a\}$, where a is an absorbing state and B' is the subset of B including the states reachable before exiting B from the states with non-null initial probability, same initial probability distribution in B' and transition rates among states in B' as X, and transition rates $\lambda'_{i,a} = \lambda_{i,B^c}$, $i \in B'$, so that Y_B enters a whenever X exits B). Note that $\tau(i, Y_B) > 0$ for $i \in B'$.

Let Y_U^m, $m \in U$ be the transient CTMC with initial state m tracking X from m till exit from U and let T_U^m be the mean time to absorption of Y_U^m. Let $\alpha_{s,m}$ be the probability that X with initial state $s \in S$ will enter U through state m. We have:

$$T_{U,s} = \sum_{m \in U} \alpha_{s,m} T_U^m. \qquad (4)$$

Invoking (4), we can easily upper bound $T_{U,s}$ from upper bounds for T_U^m, $m \in U$. To obtain these bounds we will invoke the exact aggregation theorem for transient CTMCs and a lemma, which generalizes the mean holding time lemma proved in [11]. Exact aggregation results for irreducible CTMCs are given in [4]. These results extend easily to transient CTMCs. We have:

Theorem 3 (Exact aggregation for transient CTMCs)
Let $Y = \{Y(t); t \geq 0\}$ be a transient CTMC with state space $B \cup \{a\}$, where all states in B are transient and a is an absorbing state, transition rates λ_{ij}, $i \in B$, $j \in B \cup \{a\}$, $i \neq j$, and initial probability distribution $P[Y(0) = i] = \pi_i$, $i \in B$, $\sum_{i \in B} \pi_i = 1$. Assume $\tau(i, Y) > 0$ for all $i \in B$. Let $B_1 \cup B_2 \cup \ldots \cup B_n$ be a partition of B. Then, there exists a transient CTMC $Y' = \{Y'(t); t \geq 0\}$ (the exact aggregation of Y) with state space $\{b_1, b_2 \ldots, b_n\} \cup \{a\}$, transition rates $\lambda'_{b_k, b_l} = \sum_{i \in B_k} w_i^k \lambda_{i, B_l}$, $1 \leq k, l \leq n$, $k \neq l$ and $\lambda'_{b_k, a} = \sum_{i \in B_k} w_i^k \lambda_{i, a}$, $1 \leq k \leq n$, with $w_i^k > 0$, $\sum_{i \in B_k} w_i^k = 1$, and initial probability distribution $P[Y'(0) = b_k] = \pi'_k = \sum_{i \in B_k} \pi_i$, such that $\tau(b_k, Y') = \tau(B_k, Y)$.

Proof See appendix. $\square$

Consider the exact aggregation, $Y_U^{m\prime}$ of Y_U^m, $m \in C_k$, $K < k \leq N'$ under the partition $\cup_{k=K+1}^{N'_m} C_k^m$, where C_k^m is the subset of C_k including the states reachable from m before exiting U and $K + 1 \leq N'_m \leq N'$. $Y_U^{m\prime}$ has a transition state diagram like the one given in Figure 1, (a) with N' substituted by N'_m. The following lemma shows how the times to absorption vector of $Y_U^{m\prime}$ can be upper bounded.

Lemma 1 *Assume $N' \leq N$. Let $Y' = \{Y'(t); t \geq 0\}$ be a transient CTMC with the state transition diagram of Figure 1, (a) and initial probability distribution $P[Y'(0) =$*

$c_i] = \pi_i$, $K + 1 \le i \le N'$, $\sum_{i=K+1}^{N'} \pi_i = 1$. *Let $Y = \{Y(t); t \ge 0\}$ be the transient CTMC with the state transition diagram of Figure 1, (b) and initial probability distribution $P[Y(0) = c_i] = \pi_i$, $K + 1 \le i \le N'$, $P[Y(0) = c_i] = 0$, $N' < i \le N$. Assume $f_{i,j} \le f_{i,j}^+$ and $g_i \ge g_i^- > 0$, $K + 1 \le i \le N'$. Then, $\tau(c_i, Y) \ge \tau(c_i, Y')$, $K + 1 \le i \le N'$.*

Proof For notational conciseness let $\tau_i = \tau(c_i, Y)$, $\tau_i' = \tau(c_i, Y')$. We will use as a basic tool the balance equation for a subset of states of a transient CTMC, which establishes that the initial probability of the subset plus the expected number of entries must be equal to the final probability of the subset plus the expected number of exits. The states c_i of Y and Y' are transient and, therefore, have final probabilities equal to 0.

The balance equation applied to Y' and the subset $\{c_{K+1}, c_{K+2}, \ldots, c_{N'}\}$ gives:

$$1 = \tau_{K+1}' g_{K+1}, \tag{5}$$

$$\tau_{K+1}' = \frac{1}{g_{K+1}}. \tag{6}$$

The balance equation applied to Y' and the subset $\{c_{K+1}, c_{K+2}, \ldots, c_{k-1}\}$, $K + 1 < k \le N'$ gives:

$$\sum_{i=K+1}^{k-1} \pi_i + \tau_k' g_k = \tau_{K+1}' g_{K+1} + \sum_{i=K+1}^{k-1} \tau_i' \sum_{k-i \le j \le N'-i} f_{i,j},$$

which, using (5) and $1 - \sum_{i=K+1}^{k-1} \pi_i = \sum_{i=k}^{N'} \pi_i$ gives:

$$\tau_k' = \frac{\sum_{i=k}^{N'} \pi_i + \sum_{i=K+1}^{k-1} \tau_i' \sum_{k-i \le j \le N'-i} f_{i,j}}{g_k}. \tag{7}$$

Equations (6), (7) define a recursive solution for τ_k', $k = K + 1, \ldots, N'$. Analysis of Y gives similar equations for τ_k (it has been used $P[Y(0) = c_i] = 0$ for $N' < i \le N$):

$$\tau_{K+1} = \frac{1}{g_{K+1}^-}, \tag{8}$$

$$\tau_k = \frac{\sum_{i=k}^{N} \pi_i + \sum_{i=K+1}^{k-1} \tau_i \sum_{k-i \le j \le N-i} f_{i,j}^+}{g_k^-}$$

$$= \frac{\sum_{i=k}^{N'} \pi_i + \sum_{i=K+1}^{k-1} \tau_i \sum_{k-i \le j \le N-i} f_{i,j}^+}{g_k^-}. \tag{9}$$

The result is proved inductively for $k = K + 1, \ldots, N'$. Since $g_{K+1}^- \le g_{K+1}$, using (8) and (6):

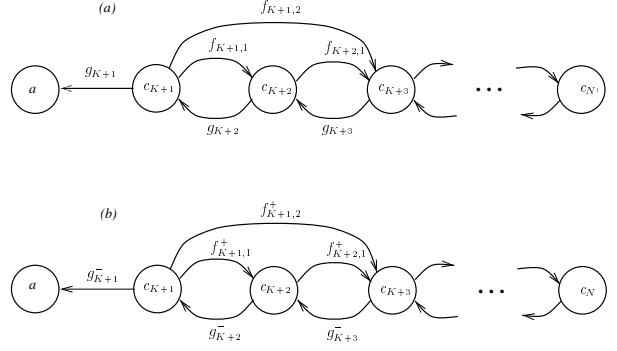$$\tau_{K+1} = \frac{1}{g_{K+1}^-} \ge \frac{1}{g_{K+1}} = \tau_{K+1}'.$$

Figure 1. State transition diagrams of CTMCs of Lemma 1.

Assume $\tau_l \ge \tau_l'$, $K + 1 \le l < k$. Using (9), $g_k^- \le g_k$, $N' \le N$, $f_{i,j}^+ \ge f_{i,j}$, and (7):

$$\tau_k = \frac{\sum_{i=k}^{N'} \pi_i + \sum_{i=K+1}^{k-1} \tau_i \sum_{k-i \le j \le N-i} f_{i,j}^+}{g_k^-}$$

$$\ge \frac{\sum_{i=k}^{N'} \pi_i + \sum_{i=K+1}^{k-1} \tau_i' \sum_{k-i \le j \le N'-i} f_{i,j}}{g_k} = \tau_k'. \square$$

Let $\alpha_s(k) = \sum_{m \in C_k} \alpha_{s,m}$ be the probability that X with initial state $s \in S$ will enter U through subset C_k. Let $f_{i,j}^+$ (g_i^-) be upper (lower) bounds for the transition rates $f_{i,j}^{m'}$ from c_i to c_{i+j} ($g_i^{m'}$ from c_i to c_{i-1}, $i > K + 1$ or from c_{K+1} to a) of any $Y_U^{m'}$. Let $T(k)$ be the mean time to absorption of the transient CTMC Y of Figure 1, (b) with initial state c_k. We have:

Theorem 4 $T_{U,s} \le \sum_{k=K+1}^{N'} \alpha_s(k) T(k) = [T_{U,s}]_{\text{ub}}$.

Proof Let Y^k be the transient CTMC with the state transition diagram of Figure 1, (b) and initial state c_k. $Y_U^{m'}$, $m \in C_k$ and Y^k satisfy the conditions of Lemma 1 and, therefore, $\tau(c_i, Y_U^{m'}) \le \tau(c_i, Y^k)$. By Theorem 3 we have $T_U^m = \sum_{i=K+1}^{N_m'} \tau(c_i, Y_U^{m'})$. Then, $T_U^m \le \sum_{i=K+1}^{N'} \tau(c_i, Y^k) = T(k)$. It follows (4):

$$T_{U,s} = \sum_{m \in U} \alpha_{s,m} T_U^m = \sum_{k=K+1}^{N'} \sum_{m \in C_k} \alpha_{s,m} T_U^m$$

$$\le \sum_{k=K+1}^{N'} \sum_{m \in C_k} \alpha_{s,m} T(k) = \sum_{k=K+1}^{N'} \alpha_s(k) T(k). \square$$

Upper bounds $f_{i,j}^+$ for the transition rates $f_{i,j}^{m'}$ can be easily derived. Let E_j be the subset of E including the

failure bags of cardinality j. It is clear that $\lambda_{n,C_i+j}, n \in C_i$ is upper bounded by $\sum_{e \in E_j}[\lambda(e)]_{\mathrm{ub}}$. Using Theorem 3:

$$f_{i,j}^{m'} = \sum_{n \in C_i^m} w_n^i \lambda_{n,C_i+j},$$

with $w_n^i > 0, \sum_{n \in C_i^m} w_n^i = 1$. Then, it follows:

$$f_{i,j}^{m'} \leq \sum_{n \in C_i^m} w_n^i \sum_{e \in E_j}[\lambda(e)]_{\mathrm{ub}} = \sum_{e \in E_j}[\lambda(e)]_{\mathrm{ub}} = f_{i,j}^+.$$
$$(10)$$

In [11] the lowest repair rate of the model is used as lower bounds g_i^-. Unfortunately, a similar approach cannot be taken for the models considered in this paper, since depending on the characteristics of the phase type distributions the lowest rate to the left from the states of a subset C_i may be 0. A more sophisticated approach is needed. That approach is developed in the next section.

3.3 Computation of g_i^-

To derive the bounds g_i^- we need a result from [9] which is obtained for irreducible CTMCs. To establish a link with this result we define irreducible CTMCs $X_U^m, m \in U$ as follows: X_U^m has state space $U_m' \cup \{a\}$, where U_m' is the subset of U including the states reachable from m before exiting U, transition rates from U_m' to $U_m' \cup \{a\}$ as Y_U^m and a transition rate 1 from a to m. Let $X_U^{m'}$ be the exact aggregate of X_U^m under the partition $(\cup_{k=K+1}^{N_m'} C_k^m) \cup \{a\}$. Given the connection between Theorem 3 and the exact aggregation theorem for irreducible CTMCs and the relationships between the involved CTMCs, it is not hard to see that the transition rates of $X_U^{m'}$ from $\{c_{K+1}, \ldots, c_{N_m'}\}$ to $\{c_{K+1}, \ldots, c_{N_m'}, a\}$ are equal to the corresponding transition rates of $Y_U^{m'}$.

Let $q_{k,i}, i \in C_k^m$ denote the probability that X_U^m will jump from $C_k^m, k > K+1$ (C_{K+1}^m) to C_{k-1}^m (a) given entry in C_k^m (C_{K+1}^m) through state i. Let $h_{k,i}, i \in C_k^m$ be the mean holding time of X_U^m in C_k^m conditioned to entry in the subset through state i. Then, it has been shown in [9] that:

$$g_k^{m'} \geq \min_{i \in C_k^m} \frac{q_{k,i}}{h_{k,i}}. \qquad (11)$$

Assume that a lower bound, q^-, for $q_{k,i}, K < k \leq N_m'$, $i \in C_k^m$ and an upper bound, h^+, for $h_{k,i}, K < k \leq N_m'$, $i \in C_k^m$ are available. Using (11) we have:

$$g_k^{m'} \geq \frac{q^-}{h^+} = g_k^-. \qquad (12)$$

In the following we show how q^- and h^+ can be derived. To that end we first introduce some notation. Let a state $i \in C_k^m, 0 \leq k < N_m'$. We will denote by $\lambda_{i,C_{>k}}$ the transition rate from i to $\cup_{k'=k+1}^{N_m'} C_{k'}^m$ (note that $\lambda_{j,C_{>k}}$ is the same for all states j which are visited in C_k^m from a given

entry state i, since all these states have the same bag of failed components), by A_i the number of active repair processes in i, by $a_j(i), 1 \leq j \leq A_i$ the phase type distribution of the jth active repair process in state i ($1 \leq a_j(i) \leq L$), and by $s_j(i)$ the state of the phase type distribution $Z_{a_j(i)}$ in state i. Let W_j^s be the random variable time to absorption of Z_j with initial state s. Let $\lambda_{\mathrm{ub}} = \sum_{e \in E}[\lambda(e)]_{\mathrm{ub}}$ and let $\mathrm{EXP}(\lambda)$ denote an exponential random variable with parameter λ. Since $\lambda_{i,C_{>k}} \leq \lambda_{\mathrm{ub}}$ and the random variables W_j^s, $\mathrm{EXP}(\lambda_{i,C_{>k}})$ and $\mathrm{EXP}(\lambda_{\mathrm{ub}})$ are independent we have:

$$
\begin{aligned}
q_{k,i} &= P\left[\min_{1 \leq j \leq A_i} W_{a_j(i)}^{s_j(i)} < \mathrm{EXP}(\lambda_{i,C_{>k}})\right] \\
&\geq P\left[\min_{1 \leq j \leq A_i} W_{a_j(i)}^{s_j(i)} < \mathrm{EXP}(\lambda_{\mathrm{ub}})\right] \\
&\geq \min_{1 \leq j \leq A_i} P\left[W_{a_j(i)}^{s_j(i)} < \mathrm{EXP}(\lambda_{\mathrm{ub}})\right] \\
&\geq \min_{1 \leq j \leq L} \min_{s \in L_j} P[W_j^s < \mathrm{EXP}(\lambda_{\mathrm{ub}})].
\end{aligned}
$$

Also:

$$
\begin{aligned}
h_{k,i} &= E\left[\min\left\{\mathrm{EXP}(\lambda_{i,C_{>k}}), \min_{1 \leq j \leq A_i} W_{a_j(i)}^{s_j(i)}\right\}\right] \\
&< E\left[\min_{1 \leq j \leq A_i} W_{a_j(i)}^{s_j(i)}\right] \\
&\leq \max_{1 \leq j \leq A_i} E\left[W_{a_j(i)}^{s_j(i)}\right] \\
&\leq \max_{1 \leq j \leq L} \max_{s \in L_j} E[W_j^s].
\end{aligned}
$$

Let us denote $P[W_j^s < \mathrm{EXP}(\lambda_{\mathrm{ub}})]$ by Q_j^s and $E[W_j^s]$ by H_j^s. We use:

$$q^- = \min_{1 \leq j \leq L} \min_{s \in L_j} Q_j^s, \qquad (13)$$

$$h^+ = \max_{1 \leq j \leq L} \max_{s \in L_j} H_j^s. \qquad (14)$$

Let $\mathbf{B}_j$ be the transition rate matrix of Z_j restricted to the transient states L_j and let $\mathbf{b}_j$ be the vector whose entries are the transition rates of Z_j from L_j to the absorbing state a. Let $\mathbf{Q}_j$ and $\mathbf{H}_j$ be the vectors with entries Q_j^s and H_j^s, respectively, $s \in L_j$. $\mathbf{Q}_j$ and $\mathbf{H}_j$ can be obtained as:

$$\mathbf{Q}_j = -(\mathbf{B}_j - \lambda_{\mathrm{ub}}\mathbf{I})^{-1}\mathbf{b}_j, \qquad (15)$$

$$\mathbf{H}_j = -\mathbf{B}_j^{-1}\mathbf{1}. \qquad (16)$$

Equation (16) is trivial since the component in row s and column i of $-\mathbf{B}_j^{-1}$ is the mean time to absorption spent by Z_j in state i given that the initial state is s. Equation (15) follows considering transient CTMCs Z_j' with state space $L_j \cup \{a, b\}$, where a and b are absorbing states, same transition rates from L_j to a as Z_j and transition rates $\lambda_{s,b} = \lambda_{\mathrm{ub}}, s \in L_j$. The transition rate matrix of Z_j' restricted to L_j is $\mathbf{B}_j - \lambda_{\mathrm{ub}}\mathbf{I}$ and Q_j^s is the probability of being absorbed in state a given that the initial state is s. These comments justify the equation.

3.4 Computation of $T(k)$

Let M be the set of indices k associated to the subsets C_k, $K < k \leq N$ with $\lambda_{i,C_k} \neq 0$ for some $i \in G$. Remember that $T(k)$ is the mean time to absorption of the transient CTMC Y of Figure 1, (b) with initial state c_k. In order to obtain the bounds $[T_{U,s}]_{\mathrm{ub}}$ given by Theorem 4 we have to compute $T(k)$, $k \in M$. A direct computation of each $T(k)$ solving Y with initial state c_k would require the solution of $|M|$ linear systems. A more efficient procedure, specially for large $|M|$, can be developed exploiting the following equations, where $\phi_k = g_k^- + \sum_i f_{k,i}^+$ denotes the output rate of Y from c_k (see Figure 1, (b)):

$$T(k) = \frac{1}{\phi_k} + \frac{g_k^-}{\phi_k} T(k-1) + \sum_i \frac{f_{k,i}^+}{\phi_k} T(k+i),$$
$$K + 2 \leq k < N, \quad (17)$$

$$T(N) = \frac{1}{g_N^-} + T(N-1). \quad (18)$$

These equations are obtained as follows. First, consider (17). $T(k)$, mean time to absorption of Y with initial state c_k, is equal to the mean sojourn time in c_k, $1/\phi_k$, plus the mean time to absorption from the next visited state, which is c_{k-1} with probability g_k^-/ϕ_k and c_{k+i} with probability $f_{k,i}^+/\phi_k$. Equation (18) is obtained similarly; in this case, $\phi_N = g_N^-$ and state c_{N-1} is the next visited state with probability 1. Equations (17), (18) can be solved recursively in terms of $T(N)$, yielding:

$$T(N-1) = T(N) - \frac{1}{g_N^-}, \quad (19)$$

$$T(k) = \frac{1}{g_{k+1}^-} \Big[\phi_{k+1} T(k+1) - 1$$
$$- \sum_i f_{k+1,i}^+ T(k+1+i) \Big],$$
$$k = N - 2, \ldots, K + 1. \quad (20)$$

It remains to discuss the computation of $T(N)$. Let τ_i denote the mean time to absorption in state c_i of Y with initial state c_N. Then:

$$T(N) = \sum_{i=K+1}^{N} \tau_i. \quad (21)$$

The row vector $\boldsymbol{\tau}^T = (\tau_{K+1} \ldots \tau_N)$ is the solution of the linear system:

$$\boldsymbol{\tau}^T \mathbf{A} = -(0 \ldots 01), \quad (22)$$

where $\mathbf{A}$ is the restriction of the transition rate matrix of Y to the transient states. A direct solution of (22) is possible exploiting the upper Hessenberg structure of $\mathbf{A}$ and the fact that all components except the last one of the right-hand vector of (22) are null. Defining $\nu_i = \tau_i/\tau_{K+1}$ ($\nu_{K+1} =$

1), all the equations except the last one give a triangular linear system on ν_i, $K + 2 \leq i \leq N$ which can be solved easily. Substituting then τ_i by $\nu_i \tau_{K+1}$, $K + 2 \leq i \leq N$, in the last equation of (22) and using the solution for ν_i, $K + 2 \leq i \leq N$ found in the previous step gives an equation on τ_{K+1}. Solving that equation and using $\tau_i = \nu_i \tau_{K+1}$, $K + 2 \leq i \leq N$ we obtain τ_i, $K + 2 \leq i \leq N$. The solution procedure can be described as follows:

$$\nu_{K+1} = 1,$$
$$\nu_i = \frac{1}{g_i^-} \Big[\phi_{i-1} \nu_{i-1} - \sum_{j=K+1}^{i-2} f_{j,i-j-1}^+ \nu_j \Big], \quad (23)$$
$$i = K + 2, \ldots, N,$$

$$\tau_{K+1} = \frac{1}{\phi_N \nu_N - \sum_{i=K+1}^{N-1} f_{i,N-i}^+ \nu_i}, \quad (24)$$
$$\tau_i = \nu_i \tau_{K+1}, \; i = K + 2, \ldots, N.$$

3.5 Computation of the bounds

$T_{G,s}$, $C_{G,s}$ and $\alpha_s(k)$, $k \in M$ could be computed from the mean times to absorption vector of the transient CTMCs Y_G^s tracking X from state s till exit from G. This however would involve the solution of $|S|$ linear systems of size $|G|$, which is very expensive. To avoid that requirement, a state cloning technique is developed in [11] which reduces the number of linear systems which have to be solved but introduces looseness in the bounds. In this section we develop a new computational procedure which obtains the bounds $[R]_{\mathrm{lb}}$ and $[R]_{\mathrm{ub}}$ solving only 4 linear systems.

Let:

$$T_s' = T_{G,s} + [T_{U,s}]_{\mathrm{ub}},$$
$$C_s' = C_{G,s} + [r]_{\mathrm{lb}}[T_{U,s}]_{\mathrm{ub}},$$
$$C_s'' = C_{G,s} + [r]_{\mathrm{ub}}[T_{U,s}]_{\mathrm{ub}}.$$

Using (3), the bounds (1), (2) for R can be expressed in terms of T_s', C_s' and C_s'', $s \in S$ as:

$$[R]_{\mathrm{lb}} = \min_{s \in S} \Big\{ \frac{C_s'}{T_s'} \Big\}, \quad (25)$$

$$[R]_{\mathrm{ub}} = \max_{s \in S} \Big\{ \frac{C_s''}{T_s'} \Big\}. \quad (26)$$

The key of the new method is the derivation of forward equations for T_i', C_i' and C_i'', $i \in G$. To that end we first write these variables in terms of $\alpha_i(k)$ and $T(k)$, using Theorem 4:

$$T_i' = T_{G,i} + \sum_{k \in M} \alpha_i(k) T(k), \; i \in G,$$

$$C_i' = C_{G,i} + [r]_{\mathrm{lb}} \sum_{k \in M} \alpha_i(k) T(k), \; i \in G,$$

$$C_i'' = C_{G,i} + [r]_{\text{ub}} \sum_{k \in M} \alpha_i(k) T(k), \ i \in G.$$

Each of these variables can be expressed as the sum of a contribution associated with the visit to state i plus the corresponding variable for the next visited state. This gives the forward equations:

$$T_i' = \frac{1}{\lambda_i} + \sum_{k \in M} \frac{\lambda_{i,C_k}}{\lambda_i} T(k) + \sum_{\substack{j \in G \\ j \neq i}} \frac{\lambda_{ij}}{\lambda_i} T_j', \ i \in G, \quad (27)$$

$$C_i' = \frac{r_i}{\lambda_i} + \sum_{k \in M} \frac{\lambda_{i,C_k}}{\lambda_i} [r]_{\text{lb}} T(k) + \sum_{\substack{j \in G \\ j \neq i}} \frac{\lambda_{ij}}{\lambda_i} C_j', \ i \in G, \quad (28)$$

$$C_i'' = \frac{r_i}{\lambda_i} + \sum_{k \in M} \frac{\lambda_{i,C_k}}{\lambda_i} [r]_{\text{ub}} T(k) + \sum_{\substack{j \in G \\ j \neq i}} \frac{\lambda_{ij}}{\lambda_i} C_j'', \ i \in G. \quad (29)$$

Let $\rho_{ij} = \lambda_{ij}/\lambda_i$. The sets of equations (27), (28) and (29) can be formulated as linear systems introducing the matrix $\mathbf{B} = \mathbf{I} - (\rho_{ij})_{i,j \in G, i \neq j}$ and the vectors $\mathbf{T}' = (T_i')_{i \in G}$, $\mathbf{C}' = (C_i')_{i \in G}$, $\mathbf{C}'' = (C_i'')_{i \in G}$, $\boldsymbol{\mu}' = ((1/\lambda_i) + \sum_{k \in M} (\lambda_{i,C_k}/\lambda_i) T(k))_{i \in G}$, $\mathbf{c}' = ((r_i/\lambda_i) + \sum_{k \in M} (\lambda_{i,C_k}/\lambda_i)[r]_{\text{lb}} T(k))_{i \in G}$, and $\mathbf{c}'' = ((r_i/\lambda_i) + \sum_{k \in M} (\lambda_{i,C_k}/\lambda_i)[r]_{\text{ub}} T(k))_{i \in G}$:

$$\mathbf{B}\mathbf{T}' = \boldsymbol{\mu}', \quad (30)$$

$$\mathbf{B}\mathbf{C}' = \mathbf{c}', \quad (31)$$

$$\mathbf{B}\mathbf{C}'' = \mathbf{c}''. \quad (32)$$

Matrix $\mathbf{B}$ can be large and iterative methods should be used to solve the linear systems (30), (31), (32). From the properties of $\mathbf{B}$ it is easy to prove [17] that Gauss-Seidel will converge. We found that the convergence under Gauss-Seidel was typically extremely slow. However, a decomposition technique can be used to speed up the convergence. The price is to solve one more linear system, but we have found that then Gauss-Seidel converges very fast. See [6] for an analysis of the convergence properties of the linear systems obtained with the decomposition technique.

To describe the decomposition technique let us consider the generic problem of computing for $i \in G$ the expected accumulated reward up to absorption V_i of the transient CTMC Y_G^i with initial state i tracking X till exit from G for the generic reward rate structure $v_i, i \in G$. Note that T_i', C_i' and C_i'' can be formulated as V_i with v_i equal to, respectively, $1 + \sum_{k \in M} \lambda_{i,C_k} T(k)$, $r_i + \sum_{k \in M} \lambda_{i,C_k} [r]_{\text{lb}} T(k)$ and $r_i + \sum_{k \in M} \lambda_{i,C_k} [r]_{\text{ub}} T(k)$. Let the vectors $\mathbf{V} = (V_i)_{i \in G}$ and $\mathbf{b} = (v_i/\lambda_i)_{i \in G}$. Then, $\mathbf{V}$ is the solution of the linear system:

$$\mathbf{B}\mathbf{V} = \mathbf{b}.$$

Without loss of generality let us assume that the state o in which all components are up has index 1. Let $\widetilde{V}_i$ denote the expected accumulated reward to absorption or hit of state 1. Let γ_i denote the probability that Y_G^i will exit G without

hitting state 1. Decomposing V_i in its two contributions delimited by the time at which Y_G^i gets absorbed or hits state 1, we obtain:

$$V_i = \widetilde{V}_i + (1 - \gamma_i)V_1, \ i \in G, \quad (33)$$

The set of equations (33) can be solved in $V_i, i \in G$, yielding:

$$V_i = \widetilde{V}_i + \frac{1 - \gamma_i}{\gamma_1} \widetilde{V}_1, \ i \in G.$$

Note that $\widetilde{V}_i$ is the expected accumulated reward to absorption of the transient CTMC $\widetilde{Y}_G^i$ obtained from Y_G^i by directing to the absorbing state the entries in state 1. Then, $\widetilde{V}_i$, $i \in G$ can be computed as $V_i, i \in G$, using the matrix $\widetilde{\mathbf{B}}$:

$$\widetilde{\mathbf{B}} = \begin{pmatrix} 1 & -\rho_{12} & \cdots & -\rho_{1,|G|} \\ 0 & 1 & \cdots & -\rho_{2,|G|} \\ & & \ddots & \\ 0 & -\rho_{|G|,2} & \cdots & 1 \end{pmatrix}$$

instead of $\mathbf{B}$. Let the vectors $\widetilde{\mathbf{T}}' = (\widetilde{T}_i')_{i \in G}$, $\widetilde{\mathbf{C}}' = (\widetilde{C}_i')_{i \in G}$, $\widetilde{\mathbf{C}}'' = (\widetilde{C}_i'')_{i \in G}$. Applying the previous result we have that these vectors are the solutions of the linear systems:

$$\widetilde{\mathbf{B}}\widetilde{\mathbf{T}}' = \boldsymbol{\mu}', \quad (34)$$

$$\widetilde{\mathbf{B}}\widetilde{\mathbf{C}}' = \mathbf{c}', \quad (35)$$

$$\widetilde{\mathbf{B}}\widetilde{\mathbf{C}}'' = \mathbf{c}''. \quad (36)$$

The probabilities γ_i can be formulated as the expected accumulated reward up to absorption of $\widetilde{Y}_G^i$ with reward rate $\lambda_{i,U}$. Then, letting the vectors $\boldsymbol{\gamma} = (\gamma_i)_{i \in G}$ and $\boldsymbol{\omega} = (\lambda_{i,U}/\lambda_i)_{i \in G}$, $\boldsymbol{\gamma}$ is the solution of the linear system:

$$\widetilde{\mathbf{B}}\boldsymbol{\gamma} = \boldsymbol{\omega}. \quad (37)$$

Finally, T_s', C_s' and C_s'', $s \in S$ can be computed using:

$$T_s' = \widetilde{T}_s' + \frac{1 - \gamma_s}{\gamma_1} \widetilde{T}_1', \quad (38)$$

$$C_s' = \widetilde{C}_s' + \frac{1 - \gamma_s}{\gamma_1} \widetilde{C}_1', \quad (39)$$

$$C_s'' = \widetilde{C}_s'' + \frac{1 - \gamma_s}{\gamma_1} \widetilde{C}_1''. \quad (40)$$

The complete method to compute the bounds can be summarized as follows:

Algorithm
1. Compute $f_{i,j}^+$ using (10) and g_i^- using (15), (16), (13), (14) and (12).
2. Compute $T(N)$ using (23), (24) and (21).
3. Compute $T(k)$, $K + 1 \leq k < N$ using (19), (20).
4. Generate the transient CTMC Y_G.
5. Solve the linear systems (34), (35), (36) and (37).
6. Compute T_s', C_s' and C_s'', $s \in S$ using (38), (39) and (40).
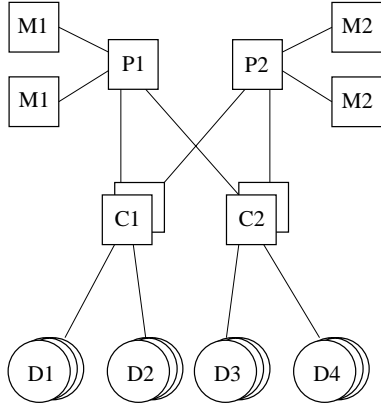7. Compute $[R]_{\text{lb}}$, $[R]_{\text{ub}}$ using (25), (26).

Figure 2. Block diagram of the example.

4 Numerical results

In this section we illustrate the bounding method using an example having a number of states precluding its exact numerical solution. The example is the fault-tolerant database system whose block diagram is given in Figure 2. The system is made up of two processing subsystems, each including one processor P and two memories M, two sets of controllers C1 and C2, each with two controllers, and four sets of disks D1, D2, D3 and D4, each with three disks. The system is up if at least one processor and one memory connected to it are operational, one controller of each set is operational, and two disks of each set are operational. Processors fail with rate $\lambda_P = 10^{-5}$; a processor failure is soft with probability $S_P = 0.8$ and hard with probability $1 - S_P = 0.2$. In addition, either being soft or hard, a processor failure contaminates (fails) the operational memories to which it is connected with probability $1 - C_P = 0.05$. Memories fail with rate $\lambda_M = 10^{-4}$, controllers fail with rate $\lambda_C = 2 \times 10^{-5}$. Disks fail with rates $\lambda_{D1} = 10^{-6}$, $\lambda_{D2} = 1.5 \times 10^{-6}$, $\lambda_{D3} = 2 \times 10^{-6}$, and $\lambda_{D4} = 3 \times 10^{-6}$. There are two repairmen. One performs restarts of processors in soft failure and the other performs all the other maintenance actions with first priority given to disks, next to controllers, next to processors and next to memories. Components with the same repair priority are chosen at random. The policy is preemptive-resume. Figure 3 gives the phase type distributions for all repair actions, with the initial probabilities shown inside the circles denoting the states of Z_i. The state of the system can be described by giving the number of components of each class which are operational and the number of components of each class in each state of the phase type distributions. The complete model has about 4.9×10^9 states, clearly outside of current computing capabilities.

Next, we illustrate the bounding method using the example. The example has 10 component classes and $N = 22$ components. Table 1 gives the failure bags of the model and for each failure bag e the upper bound $[\lambda(e)]_{\mathrm{ub}}$ for its rate.
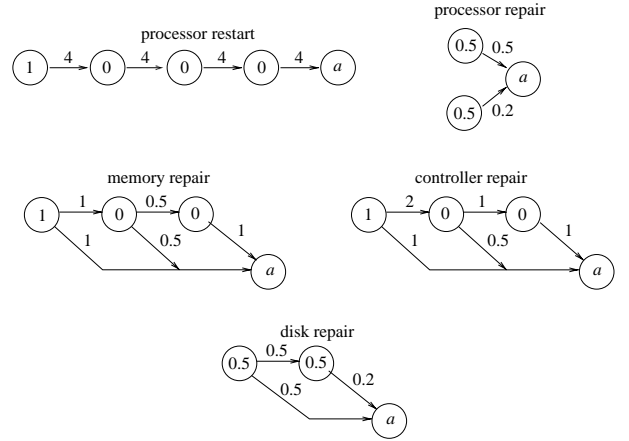

Figure 3. Phase type repair distributions for the repair actions of the example.

Table 1. Failure bags e and $[\lambda(e)]_{\mathrm{ub}}$ for the example.

e	$[\lambda(e)]_{\mathrm{ub}}$
P1[1]	10^{-5}
P1[1] M1[1]	5×10^{-7}
P1[1] M1[2]	5×10^{-7}
M1[1]	2×10^{-4}
P2[1]	10^{-5}
P2[1] M2[1]	5×10^{-7}
P2[1] M1[2]	5×10^{-7}
M2[1]	2×10^{-4}
C1[1]	4×10^{-5}
C2[1]	4×10^{-5}
D1[1]	3×10^{-6}
D2[1]	4.5×10^{-6}
D3[1]	6×10^{-6}
D4[1]	9×10^{-6}

We use the notation $c[n]$ to indicate n instances of component class c. The upper bounds $f_{i,j}^+$ are $f_{i,1}^+ = 5.225 \times 10^{-4}$, $f_{i,2}^+ = 10^{-6}$ and $f_{i,3}^+ = 10^{-6}$. The upper bound λ_{ub} is $\lambda_{\mathrm{ub}} = 5.245 \times 10^{-4}$. We also have $h^+ = 5$, $q^- = 0.997384$ and $g_i^- = 0.199477$.

Table 2 gives the bounds for the steady-state unavailability (R with $r_i = 1$ for down states and $r_i = 0$ for up states and, therefore, $[r]_{\mathrm{lb}} = 0$ and $[r]_{\mathrm{ub}} = 1$) obtained for $K = 2$, 3, 4 and 5. We also give the number of detailed states ($|G|$). By profiling the code we have found out that about 75% of the CPU time is devoted to the generation of the detailed models, while the solution of the four linear systems accounts for the remaining 25%. The CPU time for $K = 5$ was about 10 minutes in a SPARC10 workstation.

The 4-Erlang phase type distribution used for processor

Table 2. Results for the example and increasing K.

K	$|G|$	lower bound	upper bound
2	519	6.62938×10^{-9}	1.50585×10^{-5}
3	5,259	6.69364×10^{-9}	3.26380×10^{-8}
4	38,914	6.75319×10^{-9}	6.82341×10^{-9}
5	224,950	6.75578×10^{-9}	6.78371×10^{-9}

Table 3. Results for the example with $K = 5$ and increasing number of stages of the k-Erlang distribution of processor restart time.

k	$|G|$	lower bound	upper bound
1	119,257	6.75662×10^{-9}	6.78303×10^{-9}
2	151,026	6.75607×10^{-9}	6.78343×10^{-9}
3	186,257	6.75588×10^{-9}	6.78361×10^{-9}
4	224,950	6.75578×10^{-9}	6.78371×10^{-9}

restarts can be imagined as an approximation to a deterministic restart time of value 1. The goodness of the approximation improves with the number of exponential stages k. We explored that issue for the model under consideration and obtained results with increasing k for $K = 5$. Table 3 gives the results. We can note that the steady-state unavailability is quite insensitive to the shape of the restart time distribution and a small value of k is enough to obtain an accurate approximation.

5 Conclusions

A method to bound the steady-state availability with phase type repair distributions has been developed. Previous bounding methods assumed a lower bound repair rate to the left greater than zero and are not applicable in general to models having phase type repair, since the smallest repair rate to the left may be zero. In addition previous bounding methods either had to solve many linear systems to obtain the tightest possible bounds or introduced looseness if state cloning techniques were used to reduce the number of linear systems to be solved, whereas our method does not clone states and requires the solution of only 4 linear systems of the size of the generated state space, being the time devoted to the solution of these linear systems small compared with the time to generate the detailed model. As we have pointed out our method per se is not confined to compute bounds for the steady-state availability: it can be used to compute bounds for the steady-state reward rate of any model exhibiting a similar structure.

Appendix: Proof of Theorem 3

Without loss of generality, assume that the transient states of Y are sorted following the subset ordering $B_1, B_2, \ldots, B_n$. For notational conciseness let $\tau_i = \tau(i, Y)$ and $\tau'_k = \tau(B_k, Y)$. Note that $\tau'_k > 0$. Let the vectors $\boldsymbol{\tau} = (\tau_i)_{i \in B}$, $\boldsymbol{\pi} = (\pi_i)_{i \in B}$ and let $\mathbf{A}$ be the transition rate matrix of Y restricted to B. $\boldsymbol{\tau}$ satisfies the linear system:

$$\boldsymbol{\tau}^T \mathbf{A} = -\boldsymbol{\pi}^T. \tag{41}$$

Let $w_i^k = \tau_i / \tau'_k$, $i \in B_k$, $1 \leq k \leq n$. Note that $w_i^k > 0$ and $\sum_{i \in B_k} w_i^k = 1$. Defining the column vectors $\mathbf{w}(k) = (w_i^k)_{i \in B_k}$, $\boldsymbol{\pi}(k) = (\pi_i)_{i \in B_k}$, we can rewrite (41) as:

$$\left(\tau'_1 \mathbf{w}(1)^T \cdots \tau'_n \mathbf{w}(n)^T \right) \begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1n} \\ & \ddots & \\ \mathbf{A}_{n1} & \cdots & \mathbf{A}_{nn} \end{pmatrix}$$
$$= -\left(\boldsymbol{\pi}(1)^T \cdots \boldsymbol{\pi}(n)^T \right),$$

where $\mathbf{A}_{kl}$ are the blocks of $\mathbf{A}$ induced by the partition of B. This block decomposition gives the set of equations:

$$\sum_{k=1}^{n} \tau'_k \mathbf{w}(k)^T \mathbf{A}_{kl} = -\boldsymbol{\pi}(l)^T, \quad 1 \leq l \leq n.$$

Postmultiplying by $\mathbf{1}$, a column vector of all ones with appropriate dimension:

$$\sum_{k=1}^{n} \tau'_k \mathbf{w}(k)^T \mathbf{A}_{kl} \mathbf{1} = -\boldsymbol{\pi}(l)^T \mathbf{1}, \quad 1 \leq l \leq n.$$

Defining $\pi'_k = \boldsymbol{\pi}(k)^T \mathbf{1} = \sum_{i \in B_k} \pi_i$, $\lambda'_{b_k, b_l} = \mathbf{w}(k)^T \mathbf{A}_{kl} \mathbf{1} = \sum_{i \in B_k} w_i^k \lambda_{i, B_l}$, $k \neq l$, and $\lambda'_{b_k} = -\mathbf{w}(k)^T \mathbf{A}_{kk} \mathbf{1}$, we get:

$$\sum_{\substack{k=1 \\ k \neq l}}^{n} \tau'_k \lambda'_{b_k, b_l} - \tau'_l \lambda'_{b_l} = -\pi'_l, \quad 1 \leq l \leq n.$$

Thus, $\boldsymbol{\tau}' = (\tau'_k)_{1 \leq k \leq n}$ satisfies the linear system:

$$\boldsymbol{\tau}'^T \mathbf{A}' = -\boldsymbol{\pi}'^T,$$

with $\boldsymbol{\pi}' = (\pi'_k)_{1 \leq k \leq n}$ and

$$\mathbf{A}' = \begin{pmatrix} -\lambda'_{b_1} & \lambda'_{b_1, b_2} & \cdots & \lambda'_{b_1, b_n} \\ \lambda'_{b_2, b_1} & -\lambda'_{b_2} & \cdots & \lambda'_{b_2, b_n} \\ & & \cdots & \\ \lambda'_{b_n, b_1} & \lambda'_{b_n, b_2} & \cdots & -\lambda'_{b_n} \end{pmatrix}. \tag{42}$$

In summary, under the condition $\lambda'_{b_k, a} = \lambda'_{b_k} - \sum_{\substack{l=1 \\ l \neq k}}^{n} \lambda'_{b_k, b_l} \geq 0$, $1 \leq k \leq N$, $\tau'_k = \tau(B_k, Y)$ ($< \infty$ since all states in B of Y are transient) is the mean time to

absorption in state b_k of the transient CTMC Y' with state space $\{b_1, b_2, \ldots, b_N\} \cup \{a\}$, transition rate matrix (42), and initial probability distribution $P[Y'(0) = b_k] = \pi'_k$, $1 \le k \le N$. The transition rates λ'_{b_k, b_l} satisfy the conditions of the theorem. It remains to show that the transition rates to the absorbing state $\lambda'_{b_k, a}$ also satisfy those conditions and are ≥ 0. First, note that the output rates of Y' can be written as:

$$\lambda'_{b_k} = -\mathbf{w}(k)^T \mathbf{A}_{kk} \mathbf{1} = \sum_{i \in B_k} w_i^k \lambda_i - \sum_{i \in B_k} w_i^k \lambda_{i, B_k}.$$

Then, using $\lambda'_{b_k, a} = \lambda'_{b_k} - \sum_{\substack{l=1 \\ l \neq k}}^{n} \lambda'_{b_k, b_l}$ and $\lambda_{ia} = \lambda_i - \sum_{l=1}^{n} \lambda_{i, B_l}$:

$$
\begin{aligned}
\lambda'_{b_k, a} &= \lambda'_{b_k} - \sum_{\substack{l=1 \\ l \neq k}}^{n} \lambda'_{b_k, b_l} \\
&= \sum_{i \in B_k} w_i^k \lambda_i - \sum_{i \in B_k} w_i^k \lambda_{i, B_k} - \sum_{\substack{l=1 \\ l \neq k}}^{n} \sum_{i \in B_k} w_i^k \lambda_{i, B_l} \\
&= \sum_{i \in B_k} w_i^k \left(\lambda_i - \sum_{l=1}^{n} \lambda_{i, B_l} \right) = \sum_{i \in B_k} w_i^k \lambda_{ia} \ge 0. \square
\end{aligned}
$$

References

[1] U. N. Bhat, *Elements of Applied Stochastic Processes*, 2nd edition, John Wiley and Sons, 1984.

[2] J. A. Carrasco, "Improving Availability Bounds using the Failure Distance Concept," in *Dependable Computing and Fault-Tolerant Systems*, vol. 5, Springer-Verlag, 1995, pp. 479–497.

[3] J.A. Carrasco, J. Escribá and A. Calderón, "Efficient Exploration of Availability Models Guided by Failure distances", *Performance Evaluation Review*, vol. 24, no. 1, May 1996.

[4] P.J. Courtois and P. Semal, "Bounds for the positive eigenvectors of nonnegative matrices and for their approximations," *Journal of the ACM*, vol. 31, no. 4, pp. 804–825, October 1984.

[5] P.J. Courtois and P. Semal, "Computable bounds for conditional steady-state probabilities in large Markov chains and queueing models," *IEEE J. Select. Areas Commun,*, vol. SAC-4, no. 6, pp. 926–937, September 1986.

[6] P. Heidelberger, J. K. Muppala and K. S. Trivedi, "Accelerating Mean Time to Failure Computations," IBM Research Report RC-20415, March 1996, to appear in *Performance Evaluation*.

[7] J.C.S. Lui and R. Muntz, "Evaluating Bounds on Steady-State Availability of Repairable Systems from Markov Models," in *Numerical Solution of Markov chains*, Marcel Dekker, New York, pp. 435–454, 1991.

[8] J.C.S. Lui and R.R. Muntz, "Computing Bounds on Steady State Availability of Repairable Computer Systems," *Journal of the ACM*, vol. 41, no. 4, July 1994, pp. 676–707.

[9] S. Mahévas and G. Rubino, "Bounding asymptotic dependability and performance measures," in *Proc. 2nd IEEE Int. Performance and Dependability Symp.*, Urbana-Champaign, USA, September 1996, pp. 176–186.

[10] S. Mahévas and G. Rubino, "Bound Computation of Dependability and Performance Measures," INRIA Technical Report no. 3135, March 1997, available from ftp.inria.fr.

[11] R.R. Muntz, E. de Souza e Silva and A. Goyal, "Bounding Availability of Repairable Computer Systems," *IEEE Trans. on Computers*, vol. 38, no. 12, pp. 1714–1723, December 1989.

[12] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach*, Dover Publications Inc., New York, 1994, chapter 2.

[13] J. L. Peterson, *Petri Net Theory and the Modeling of Systems*, Appendix, Prentice-Hall, 1981.

[14] S.M. Ross, *Stochastic Processes*, John Wiley & Sons, New York, 1983.

[15] P. Semal, "Refinable Bounds for Large Markov Chains," *IEEE Trans. on Computers*, vol. 44, no. 10, October 1995, pp. 1216–1222.

[16] E. de Souza e Silva and P.M. Ochoa, "State Space Exploration in Markov Models," *Performance Evaluation Review*, vol. 20, no. 1, June 1992, pp. 152–166.

[17] W. J. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, 1994.