Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

A Generalized Method for the Transient Analysis of Markov Models of Fault-Tolerant Systems with Deferred Repair

Jamal Temsamani & Juan A. Carrasco

Departament d'Enginyeria Electrònica, Universitat Politècnica de Catalunya, Barcelona, Spain
Published online: 15 Feb 2007.


To link to this article: http://dx.doi.org/10.1081/SAC-200068517

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Reliability

A Generalized Method for the Transient Analysis of Markov Models of Fault-Tolerant Systems with Deferred Repair

JAMAL TEMSAMANI AND JUAN A. CARRASCO

Departament d’Enginyeria Electrònica, Universitat Politècnica de Catalunya, Barcelona, Spain

Randomization is an attractive alternative for the transient analysis of continuous time Markov models. The main advantages of the method are numerical stability, well-controlled computation error, and ability to specify the computation error in advance. However, the fact that the method can be computationally expensive limits its applicability. Recently, a variant of the (standard) randomization method, called split regenerative randomization has been proposed for the efficient analysis of reliability-like models of fault-tolerant systems with deferred repair. In this article, we generalize that method so that it covers more general reward measures: the expected transient reward rate and the expected averaged reward rate. The generalized method has the same good properties as the standard randomization method and, for large models and large values of the time \( t \) at which the measure has to be computed, can be significantly less expensive. The method requires the selection of a subset of states and a regenerative state satisfying some conditions. For a class of continuous time Markov models, class \( C'_2 \), including typical failure/repair reliability models with exponential failure and repair time distributions and deferred repair, natural selections for the subset of states and the regenerative state exist and results are available assessing approximately the computational cost of the method in terms of “visible” model characteristics. Using a large model class \( C'_2 \) example, we illustrate the performance of the method and show that it can be significantly faster than previously proposed randomization-based methods.

**Keywords** Continuous-time Markov chains; Deferred repair; Fault-tolerant systems; Randomization; Transient analysis.

**Mathematics Subject Classification** 60J22.
1. Introduction

Repair deferment is an interesting approach in fault-tolerant systems in which actions of replacement of failed components are expensive, for instance, because the system is located at a remote site. Clearly, there are several tradeoffs that can be analyzed in fault-tolerant systems with deferred repair. One of them could be an appropriate repair-deferment policy: a policy allowing many faults to happen before starting repair could result in too small a system’s reliability. These and other tradeoffs can be studied with the aid of models. Homogeneous continuous time Markov chain (CTMC) models are frequently used to analyze the reliability and performability of fault-tolerant systems. To illustrate such models, Fig. 1 depicts a small reliability CTMC model of a fault-tolerant system with deferred repair using the pair-and-spare technique (Johnson, 1989), in which active modules have failure rate $\lambda_M$, the spare module does not fail, the failure of an active module is “soft” with probability $S_M$ and “hard” with probability $1 - S_M$, and whether soft or hard, the failure of an active module is covered with probability $C_M$. Modules in soft failure mode are independently recovered at rate $\mu_S$ and modules in hard failure mode are repaired by a single repairman at rate $\mu_H$. Repair is deferred till two modules are failed and, when that condition is reached, repair proceeds till reaching the state 1 without failed components, unless the system fails before. The states with deferred repair are states 2 and 3.

Rewarded CTMC models have emerged in the last years as a useful modeling paradigm. Let $X = \{X(t); t \geq 0\}$ be a CTMC with state space $\Omega$ modeling the system under study. In this article, we will consider rewarded CTMC models obtained by defining a reward rate structure $r_i \geq 0$, $i \in \Omega$. The quantity $r_i$ has the meaning of “rate” at which reward is earned while $X$ is in state $i$. In that context, two useful measures to consider are the expected transient reward rate $ETRR(i) = E[r_{X(t)}]$ and

![Figure 1. CTMC reliability model of a repairable fault-tolerant system with deferred repair using the pair-and-spare technique.](image-url)
the expected averaged reward rate $EARR(t) = E[(1/t) \int_0^t r_X(\tau) d\tau]$. As examples of instances of those generic measures, consider a CTMC modeling a fault-tolerant system with deferred repair that can be up or down, and assume that a reward rate 0 is assigned to the states in which the system is up and a reward rate 1 is assigned to the states in which the system is down. Then, $ETRR(t)$ would be the unavailability of the system at time $t$ and $EARR(t)$ would be the expected interval unavailability at time $t$ (i.e., the expected value of the fraction of time that the system is down in the interval $[0, t]$). The reward rates could also represent the “performance” rate of the system and, then, the $ETRR(t)$ measure would be the expected performance rate of the system at time $t$ and the $EARR(t)$ measure would be the expected averaged performance rate of the system during the time interval $[0, t]$.

Computation of the $ETRR(t)$ and $EARR(t)$ measures involves the transient analysis of $X$. Randomization (also called uniformization) is a well-known method for performing such analysis. The randomization method is attractive because it is numerically stable and, unlike ODE solvers (Malhotra et al., 1994; Malhotra, 1996; Reibman and Trivedi, 1988) the computation error is well controlled and can be specified in advance. It was first proposed by Grassman (1977) and has been further developed by Gross and Miller (1984). The randomization method is based on the following result (Kijima, 1997, Theorem 4.19). Let $\hat{\lambda}_{i,j}, i, j \in \Omega, j \neq i,$ be the transition rate of $X$ from state $i$ to state $j$ and let $\hat{\lambda}_i = \sum_{j \in \Omega - \{i\}} \hat{\lambda}_{i,j}, i \in \Omega,$ be the output rate of $X$ from state $i$. Consider any $\Lambda \geq \max_{i \in \Omega} \hat{\lambda}_i$ and define the homogeneous discrete time Markov chain (DTMC) $\hat{X} = \{\hat{X}_n; n = 0, 1, 2, \ldots\}$ with same state space and initial probability distribution as $X$ and transition probabilities $P[\hat{X}_{n+1} = j | \hat{X}_n = i] = P_{i,j} \hat{\lambda}_{i,j}/\Lambda, i \in \Omega, j \neq i,$ $P[\hat{X}_{n+1} = i | \hat{X}_n = i] = P_{i,i} = 1 - \hat{\lambda}_i/\Lambda, i \in \Omega.$ Let $Q = \{Q(t); t \geq 0\}$ be a Poisson process with arrival rate $\Lambda$ independent of $\hat{X}$ ($P[Q(t) = n] = e^{-\Lambda t}(\Lambda t)^n/n!$). Then, $X = \{X(t); t \geq 0\}$ is probabilistically identical to $\{\hat{X}_{Q(t)}; t \geq 0\}$. We call this the randomization result. We will review next typical implementations of the randomization method for the computation of the $ETRR(t)$ and $EARR(t)$ measures.

Using the randomization result, we can express $ETRR(t)$ as

$$ETRR(t) = \sum_{n=0}^{\infty} d(n) e^{-\Lambda t} \frac{(\Lambda t)^n}{n!},$$

with $d(n) = \sum_{i \in \Omega} r_i P[\hat{X}_n = i]$, and, using $EARR(t) = (1/t) \int_0^t ETRR(\tau) d\tau$ and $\int_0^t e^{-\Lambda \tau} (\Lambda \tau)^n/n! d\tau = (1/\Lambda) \sum_{l=n+1}^{\infty} e^{-\Lambda t} (\Lambda t)^l/l!$, we can express $EARR(t)$ as

$$EARR(t) = \frac{1}{\Lambda t} \sum_{n=0}^{\infty} d(n) \sum_{l=n+1}^{\infty} e^{-\Lambda t} (\Lambda t)^l/l!.$$
Taking into account \( 0 \leq d(n) \leq r_{\max} = \max_{i \in \Omega} r_i \), it can be easily shown that both \( \text{ETRR}(t) - \text{ETRR}^n_N(t) \) and \( \text{EARR}(t) - \text{EARR}^n_N(t) \) are \( \geq 0 \) and are upper bounded by \( r_{\max} \sum_{n=N+1}^{\infty} e^{-\lambda t} (\Lambda t)^n / n! \). Then, being \( \epsilon \) an error control parameter, \( N \) is chosen as

\[
N = \min \left\{ m \geq 0 : r_{\max} \sum_{n=m+1}^{\infty} e^{-\lambda t} (\Lambda t)^n / n! \leq \epsilon \right\}.
\]

guaranteeing an absolute error \( \leq \epsilon \) in both \( \text{ETRR}(t) \) and \( \text{EARR}(t) \). Let \( q(n) \) be the row vector \( (P[\hat{X}_n = i])_{i \in \Omega} \) and let \( P = (p_{i,j})_{i,j \in \Omega} \) be the transition probability matrix of \( \hat{X} \). Computation of \( \text{ETRR}^n_N(t) \) and \( \text{EARR}^n_N(t) \) requires the knowledge of \( q(n) \), \( 0 \leq n \leq N \). Vector \( q(0) \) is known, since it is the initial probability row vector of \( X \). Vectors \( q(n), 0 < n \leq N \) can be computed from \( q(0) \) using

\[
q(n + 1) = q(n)P.
\]

(1.1)

Stable and efficient computation of the Poisson probabilities \( e^{-\lambda t} (\Lambda t)^n / n! \) avoiding overflows and intermediate underflows is a delicate issue and several alternatives have been proposed (see Bowerman et al., 1990; Fox and Glynn, 1988; Knüsel, 1986; Moorsel and Sanders, 1997). Our implementation of all randomization-based methods will use the approach described in (Knüsel, 1986, pp. 1028–1029) (see also Abramowitz and Stegun, 1964), which has good numerical stability.

For large models, the computational cost of the randomization method is roughly due to the \( N \) vector-matrix multiplications (1.1). The truncation parameter \( N \) increases with \( \Lambda t \) and, for that reason, \( \Lambda \) is usually taken equal to \( \max_{i \in \Omega} \hat{\lambda}_i \). Using the well-known result (Ross, 1983, Theorem 3.3.5) that \( Q(t) \) has for \( \Lambda t \to \infty \) an asymptotic normal distribution with mean and variance \( \Lambda t \), it is easy to realize that, for large \( \Lambda t \) and \( \epsilon \ll 1 \), the required \( N \) will be \( \approx \Lambda t \). Then, if the model is large and has to be solved for values of \( t \) for which \( \Lambda t \) is large, the randomization method will be expensive.

Several variants of the (standard) randomization method have been proposed to improve its efficiency. Miller (1983) has used selective randomization to solve reliability models with detailed representation of error handling activities. The idea behind selective randomization (Melamed and Yadin, 1984) is to randomize the model only in a subset of the state space. Reibman and Trivedi (1988) have proposed an approach based on the multistep concept. The idea is to compute \( P^M \) explicitly, where \( M \) is the length of the multistep, and use the recurrence \( q(n + M) = q(n)P^M \) to advance \( \hat{X} \) faster for steps which have negligible contributions to the transient solution of \( X \) at time \( t \). Since, for large \( \Lambda t \), the number of \( q(n) \)'s with significant contributions is of the order of \( \sqrt{\Lambda t} \), the multistep concept allows a significant reduction of the required number of vector-matrix multiplications when \( \Lambda t \) is large. However, when \( P \) is sparse, significant fill-in can occur when computing \( P^M \). Adaptive uniformization (Moorsel and Sanders, 1994) is a method in which the randomization rate is adapted depending on the states in which the randomized DTMC can be at a given step. Numerical experiments have shown that adaptive uniformization can be faster than standard randomization for short to medium mission times. In addition, it can be used to solve models with infinite state spaces and not uniformly bounded output rates. Recently, it has been proposed to combine adaptive uniformization and standard randomization to obtain a method which outperforms both adaptive uniformization and standard randomization for most
A Generalized Method for the Transient Analysis 635

models (Moorsel and Sanders, 1997). Steady-state detection (Malhotra et al., 1994) is another proposal to speed up the standard randomization method. A method based on steady-state detection with error bounds has been developed (Sericola, 1999). Steady-state detection is useful for models which reach their steady-state before the largest time at which the measure has to be computed. Another recently proposed randomization-based method is regenerative randomization (Carrasco, 2002b, 2003). That method covers rewarded CTMC models $X$ with finite state space $\Omega = S \cup \{f_1, f_2, \ldots, f_A\}$, $A \geq 0$, satisfying some conditions. In the method, a truncated transformed model is obtained having the same measure as the original model with some arbitrarily small error and the truncated transformed model is, then, solved by standard randomization. The method requires the selection of a regenerative state $r \in S$ and its performance depends on that selection. The truncated transformed model is constructed by characterizing with enough accuracy the behavior of the original model from $S' = S - \{r\}$ up to state $r$ or a state $f_i$ and from $r$ until next hit of $r$ or a state $f_i$, and its size depends on how fast the randomized DTMC $\hat{X}$ of $X$ with a randomization rate slightly larger than $\max_{i \in \Omega} \hat{\lambda}_i$ hits with high probability $r$ or a state $f_i$ starting at a state in $S'$. For large enough models and large enough $t$, regenerative randomization will be significantly more efficient than standard randomization. Furthermore, for a class of models, class $C''$, including typical failure/repair models with exponential failure and repair time distributions and repair in every state with failed components, a natural selection for the regenerative state exists and theoretical results are available assessing approximately the performance of the method for that natural selection in terms of “visible” model characteristics. The bounding regenerative randomization method (Carrasco, 2002a) allows to compute inexpensively tight bounds for a certain class of models, class $C'''$, including typical failure/repair reliability-like models with exponential failure and repair time distributions and repair in every state with failed components. Randomization with quasistationarity detection (Carrasco, 2004) is another recently proposed randomization-based method. The method is applicable to CTMC models with state space $S \cup \{f_1, \ldots, f_A\}$, where the states $f_i$, $1 \leq i \leq A$, are absorbing and all states in $S$ are transient and reachable from each other, and is based on the existence of a quasistationary distribution in the subset of transient states of DTMCs with a certain structure. For those models and large $t$ the method can be significantly more efficient than the standard randomization method.

Recently, Temsamani and Carrasco (2004) have proposed a method called split regenerative randomization that is specifically targeted to the transient analysis of CTMC models of fault-tolerant systems with deferred repair. The method covers CTMCs $X$ with finite state space $\Omega = S \cup \{f_1, f_2, \ldots, f_A\}$, $|S| \geq 3$, $A \geq 1$, where $f_i$ are absorbing states and $S$ has to satisfy some conditions, and allows to compute the measure $m(t) = \sum_{i=1}^{A} r_{f_i} P[X(t) = f_i]$, where all $r_{f_i}$ are different and $\geq 0$. The method requires the selection of a subset $E$ of states and a regenerative state $r$. For a class of CTMC models, model class $C_2$, including typical failure/repair models of fault-tolerant systems with exponential failure and repair time distributions and deferred repair, natural selections for $E$ and $r$ exist and, for those natural selections, theoretical results are available predicting approximately the computational cost of the method. Numerical experiments have shown that, for models in that class, the method can be significantly faster than all other randomization-based methods.

In this article we generalize the split regenerative randomization method. The generalized method considers the same class of CTMCs as the previously proposed
split regenerative randomization method with $A \geq 0^1$ and allows to compute the $ETRR(t)$ and $EARR(t)$ measures with an arbitrary reward rate structure $r_i \geq 0$, $i \in \Omega$. The method has the same good properties as standard randomization (numerical stability, well-controlled computation error, and ability to specify the computation error in advance) and can be much faster than that method. In fact, it can be proved that the computational cost of the method increases smoothly with $t$. That property is called “benign” behavior. For a class of rewarded CTMC models, class $C_2$, generalizing model class $C_2$, the computational cost of the generalized method can be predicted approximately. The rest of the article is organized as follows. Section 2 develops the generalized method. Section 3 states the benign behavior of the method, discusses qualitatively the efficiency of the method compared with that of standard randomization, defines randomization, defines the model class $C$, and discusses how the computational cost of the method for those models can be predicted approximately. Using a large class $C_2$, model, Sec. 4 analyzes the performance of the method and compares it with that of standard randomization, regenerative randomization, randomization with quasistationarity detection and, for $ETRR(t)$, adaptive uniformization, which has been shown (Moorsel and Sanders, 1994) to improve the performance of standard randomization for failure/repair models with deferred repair for short to medium mission times. Finally, Sec. 5 concludes the article.

2. The Generalized Method

The method covers rewarded CTMCs $X$ with finite state space $\Omega$ and selections of the subset of states $E$ and the regenerative state $r$ such that, letting $E' = E - \{r\}$ and $\overline{E} = S - E$, the following conditions are satisfied:

C1. $\Omega = S \cup \{f_1, \ldots, f_A\}$, $|S| \geq 3$, $A \geq 0$, where the states $f_i$, $1 \leq i \leq A$, are absorbing and either all states in $S$ are transient or $S$ includes a single recurrent class of states $C \subset S$.

C2. All states are reachable (from some state with nonnull initial probability).

C3. $r_i \geq 0$, $i \in \Omega$, and all $r_i\neq r_j$ are different.

C4. $E \subset S$.

C5. $r \in E$ and if $X$ includes a single recurrent class of states $C \subset S$, $r \in C$.

C6. $|E| \geq 2$.

C7. $|\overline{E}| \geq 1$.

C8. $r$ can only be entered from $\overline{E}$ ($\lambda_{i,r} = 0$, $i \in E'$).

C9. $r$ is the only entry point in $E$ ($\lambda_{i,j} = 0$, $i \in \overline{E}$, $j \in E'$).

C10. $\lambda_{r,j} > 0$ for some $j \in E'$.

Condition C10 can be easily circumvented in practice by adding, in case $\lambda_{r,j} = 0$ for all $j \in E'$, a tiny transition rate $\lambda \leq 10^{-10} \epsilon/(2r_{\text{max}} t_{\text{max}})$ from $r$ to some state in $E'$, where $\epsilon$ is the allowed error, $r_{\text{max}} = \max_{i \in \Omega} r_i$, and $t_{\text{max}}$ is the largest time at which the measure has to be computed, introducing an error $\leq 10^{-10} \epsilon$ in both $ETRR(t)$ and $EARR(t)$, $t \leq t_{\text{max}}$ (see Carrasco, 2003). Also, if $X$ has a single recurrent class of states $C \subset S$, by conditions C5 and C10, $|C| \geq 2$, since $|C| = 1$ would imply through condition C5 that $r$ would be absorbing, in contradiction with condition C10.

$^1$The case $A = 0$ was not previously considered because in that case the $m(t)$ measure is identical to 0. The developments made in Temsamani and Carrasco (2004) for the case $A \geq 1$ carry immediately to the more general case $A \geq 0$ considered here.
Therefore, when the method is applicable, \( f_1, f_2, \ldots, f_A \) have to be the only absorbing states. This makes it easy to check whether the method is applicable to a given finite CTMC with given selections for \( E \) and \( r \). The part \( r_i \geq 0, i \in \Omega \), from condition C3 can be circumvented by shifting the reward rates by a positive quantity \( d \) so that all new reward rates \( r'_i = r_i + d \) are \( \geq 0 \). The \( ETRR(t) \) and \( EARR(t) \) measures of the original rewarded CTMC are related to the corresponding measures, \( ETRR'(t) \) and \( EARR'(t) \), of the rewarded CTMC with shifted reward rates by \( ETRR(t) = ETRR'(t) - d \) and \( EARR(t) = EARR'(t) - d \). The part that all reward rates of states \( f_i \) are different from condition C3 can be obviated by merging absorbing states with same reward rate. Finally, condition C2 can be obviated by deleting non-reachable states.

In the following, we will let \( \alpha = P[X(0) = i], \alpha_C = \sum_{i \in C} \alpha_i, C \subset \Omega \), and \( \lambda_i = \sum_{j \in \epsilon} \lambda_{ij}; C \subset \Omega - \{i\} \). Also, given a DTMC \( Y = \{Y_n; n = 0, 1, 2, \ldots\} \), we will use the notation \( Y_{l,m,c} \) for the predicate which is true when \( Y_n \) satisfies condition \( c \) for all \( n, l \leq n \leq m \) (by convention, the predicate will be true for \( l > m \) and \#(\( Y_{l,m,c} \)) for the number of indices \( n, l \leq n \leq m \), for which \( Y_n \) satisfies condition \( c \).

In the generalized method, a truncated transformed rewarded CTMC model is built with having error \( \leq \varepsilon/2 \) the same \( ETRR(t) \) and \( EARR(t) \) measures as the original rewarded CTMC model \( X \) and the \( ETRR(t) \) \( (EARR(t)) \) measure of the truncated transformed rewarded CTMC model is computed with error \( \leq \varepsilon/2 \) using the standard randomization method.

Let \( \tilde{X} \) be the DTMC obtained by randomizing \( X \) with rate \( \Lambda_E \) in \( E \) and rate \( \Lambda_T \) in \( E \cup \{f_1, f_2, \ldots, f_A\} \), where \( \Lambda_T \) is slightly larger than \( \max_{j \in \epsilon} \lambda_j \) and \( \Lambda_T \) is slightly larger than \( \max_{j \in \epsilon} \lambda_j \), e.g., \( \Lambda_E = (1 + \theta) \max_{j \in \epsilon} \lambda_j \), \( \Lambda_T = (1 + \theta) \max_{j \in \epsilon} \lambda_j \), where \( \theta \) is a small quantity, say, \( 10^{-4} \). The DTMC \( \tilde{X} \) has the same state space and initial probability distribution as \( X \) and transition probabilities \( P_{i,j} = \lambda_{ij}/\Lambda_E, i \in E, j \neq i \); \( P_{i,i} = 1 - \lambda_i/\Lambda_E, i \in E \); \( P_{i,j} = \lambda_{ij}/\Lambda_T, i \in E \cup \{f_1, f_2, \ldots, f_A\}, j \neq i \); \( P_{i,i} = 1 - \lambda_i/\Lambda_T, i \in E \cup \{f_1, f_2, \ldots, f_A\} \). Note that \( P_{i,i} > 0, i \in \Omega \). We will say that \( \tilde{X} \) is the randomized DTMC of \( X \) with randomization rate \( \Lambda_E \) in \( E \) and \( \Lambda_T \) in \( E \cup \{f_1, f_2, \ldots, f_A\} \) and that \( X \) is the derandomized CTMC of \( \tilde{X} \) with randomization rate \( \Lambda_E \) in \( E \) and \( \Lambda_T \) in \( E \cup \{f_1, f_2, \ldots, f_A\} \). In the following we will let \( P_{i,c} = \sum_{j \in \epsilon} P_{i,j}, C \subset \Omega \).

As in Temsami and Carrasco (2004), to develop the generalized method we will find it convenient to consider three DTMCs. The first one, \( Z = \{Z_n; n = 0, 1, 2, \ldots\} \), follows \( \tilde{X} \) from \( r \) till re-entry in \( r \). Formally, \( Z \) can be defined from a version, \( \tilde{X}' \), of \( \tilde{X} \) with initial state \( r \) as

\[
Z_0 = r,
\]

\[
Z_n = \begin{cases} i & \text{if } \tilde{X}_{i,n}^r \neq r \wedge \tilde{X}_{i,n}^r = i, \quad i \in S' \cup \{f_1, f_2, \ldots, f_A\}, \\ a & \text{if } \#(\tilde{X}_{i,n}^r = r) > 0. \end{cases}
\]

The DTMC \( Z \) has state space \( S' \cup \{f_1, f_2, \ldots, f_A, a\} \), where \( f_i, 1 \leq i \leq A \), and \( a \) are absorbing states and all states in \( S \) are transient (Proposition 5 in Temsami and Carrasco, 2004), and its (possibly) nonnull transition probabilities are:

\[
P[Z_{n+1} = j \mid Z_n = i] = P_{i,j}, \quad i \in S, \ j \in S' \cup \{f_1, f_2, \ldots, f_A\},
\]

\[
P[Z_{n+1} = a \mid Z_n = i] = P_{i,r}, \quad i \in S,
\]

\[
P[Z_{n+1} = f_i \mid Z_n = f_j] = P[Z_{n+1} = a \mid Z_n = a] = 1, \quad 1 \leq i \leq A.
\]
The second DTMC, \( Z' = \{ Z'_n; n = 0, 1, 2, \ldots \} \), follows \( \hat{X} \) from \( E' \) till its first visit to \( r \). Formally, \( Z' \) can be defined from \( \hat{X} \) as

\[
Z'_n = \begin{cases} 
  i & \text{if } \hat{X}_0 \in E' \cap \hat{X}_{1:n} \neq r \cap \hat{X}_n = i, \quad i \in S' \cup \{ f_1, f_2, \ldots, f_A \}, \\
  a & \text{otherwise.} 
\end{cases}
\]

The DTMC \( Z' \) has state space \( S' \cup \{ f_1, f_2, \ldots, f_A, a \} \), where \( f_i, 1 \leq i \leq A, \) and \( a \) are absorbing states and all states in \( S' \) are transient (Proposition 6 in Temsamani and Carrasco, 2004). The initial probability distribution of \( Z' \) is \( P[Z'_0 = i] = \pi_i, \quad i \in E', \) \( P[Z'_0 = \epsilon] = 0, \quad i \in E \cup \{ f_1, f_2, \ldots, f_A \} \), \( P[Z'_0 = a] = \pi_{\epsilon, E \cup \{ f_1, f_2, \ldots, f_A \}} \), and its (possibly) nonnull transition probabilities are:

\[
P[Z'_{n+1} = j | Z'_n = i] = P_{i,j}, \quad i \in S', \quad j \in S' \cup \{ f_1, f_2, \ldots, f_A \},
\]

\[
P[Z'_{n+1} = a | Z'_n = i] = P_{i,a}, \quad i \in S',
\]

\[
P[Z'_{n+1} = f_i | Z'_n = j] = P[Z'_n = a] = 1, \quad 1 \leq i \leq A.
\]

The third DTMC, \( Z'' = \{ Z''_n; n = 0, 1, 2, \ldots \} \), follows \( \hat{X} \) from \( E \) till its first visit to state \( r. \) \( Z'' \) can be defined from \( \hat{X} \) as (note that, by condition C9, the only entry point of \( \hat{X} \) in \( E \) is state \( r \))

\[
Z''_n = \begin{cases} 
  i & \text{if } \hat{X}_0 \in E \cap \hat{X}_{1:n} \neq r \cap \hat{X}_n = i, \quad i \in E \cup \{ f_1, f_2, \ldots, f_A \}, \\
  a & \text{otherwise.} 
\end{cases}
\]

The DTMC \( Z'' \) has state space \( E \cup \{ f_1, f_2, \ldots, f_A, a \} \), where \( f_i, 1 \leq i \leq A, \) and \( a \) are absorbing states and all states in \( E \) are transient (Proposition 7 in Temsamani and Carrasco, 2004). The initial probability distribution of \( Z'' \) is \( P[Z''_0 = i] = \pi_i, \quad i \in E, \) \( P[Z''_0 = \epsilon] = 0, \quad 1 \leq i \leq A, \) \( P[Z''_0 = a] = \pi_{E \cup \{ f_1, f_2, \ldots, f_A \}} \), and its (possibly) nonnull transition probabilities are:

\[
P[Z''_{n+1} = j | Z''_n = i] = P_{i,j}, \quad i \in E, \quad j \in E \cup \{ f_1, f_2, \ldots, f_A \},
\]

\[
P[Z''_{n+1} = a | Z''_n = i] = P_{i,a}, \quad i \in E,
\]

\[
P[Z''_{n+1} = f_i | Z''_n = j] = P[Z''_n = a] = 1, \quad 1 \leq i \leq A.
\]

Let \( P = (P_{i,j})_{i,j \in \Omega} \) be the transition probability matrix of \( \hat{X} \). Denoting by \( P_{C',C''}, C', C'' \subset \Omega \), the subblock of \( P \) collecting the transition probabilities from states in \( C' \) to states in \( C'' \) and letting \( P_{E,E} \) the matrix identical to \( P_{E,E} \) except that the elements of the column corresponding to state \( r \) are 0, the transition probability matrix of \( Z \) restricted to its subset of transient states, \( S \), has, with the ordering of states \( E, \bar{E} \), the form:

\[
P_Z = \begin{pmatrix} P_{E,E} & P_{E,E} \\ 0 & P_{E,E} \end{pmatrix},
\]

where \( 0 \) is a matrix of all zeroes of appropriate dimensions. The restriction of the transition probability matrix of \( Z' \) to its subset of transient states, \( S' \), has with the
ordering of states $E', \overline{E}$ the form:

$$P_{Z'} = \begin{pmatrix} P_{E',E'} & P_{E',\overline{E}} \\ 0 & P_{\overline{E},\overline{E}} \end{pmatrix}. $$

The transition probability matrix of $Z''$ restricted to its subset of transient states, $\overline{E}$, is

$$P_{Z''} = P_{\overline{E},\overline{E}}. $$

Let $\pi_i(n) = P[Z_n = i], i \in E, \pi_i(n, l) = P[Z_n \in E \land Z_{n+l} \in \overline{E} \land Z_{n+l} = i], i \in \overline{E}, \pi'_i(n) = P[Z'_n = i], i \in E', \pi'(n, l) = P[Z'_n \in E' \land Z'_{n+l} \in \overline{E} \land Z'_{n+l} = i], i \in \overline{E}$, and $\pi''_i(n) = P[Z''_n = i], i \in \overline{E}$, and consider the row vectors $\pi(n) = (\pi_i(n))_{i \in E}, \pi(n, l) = (\pi_i(n, l))_{i \in \overline{E}}, \pi'(n) = (\pi'_i(n))_{i \in E'}, \pi'(n, l) = (\pi'_i(n, l))_{i \in \overline{E}}$, and $\pi''(n) = (\pi''_i(n))_{i \in \overline{E}}$.

Assuming that, within $\overline{E}$, state $r$ is numbered first, those vectors, can be computed for $n \geq 0$, $l \geq 1$ using:

\begin{align*}
\pi(0) &= (100 \cdots 0), \\
\pi(n + 1) &= \pi(n)P_{E',E'}, \quad n \geq 0, \\
\pi(n, 1) &= \pi(n)P_{\overline{E},\overline{E}}, \quad n \geq 0, \\
\pi(n, l + 1) &= \pi(n, l)P_{\overline{E},\overline{E}}, \quad l \geq 1, \\
\pi'(0) &= (x_i)_{i \in E'}, \\
\pi'(n + 1) &= \pi'(n)P_{E',E'}, \quad n \geq 0, \\
\pi'(n, 1) &= \pi'(n)P_{\overline{E},\overline{E}}, \quad n \geq 0, \\
\pi''(n, l + 1) &= \pi''(n, l)P_{\overline{E},\overline{E}}, \quad l \geq 1, \\
\pi''(0) &= (x_{s_i})_{i \in \overline{E}}, \\
\pi''(n + 1) &= \pi''(n)P_{\overline{E},\overline{E}}, \quad n \geq 0.
\end{align*}

To define the truncated transformed model we will consider a discrete-time stochastic process $\widehat{V} = \{\widehat{V}_n; n = 0, 1, 2, \ldots\}$ defined from $\widehat{X}$ as:

\begin{align*}
\widehat{V}_n = \begin{cases}
s_k & \text{if } 0 \leq k \leq n \land \widehat{X}_{n-k} = r \land \widehat{X}_{n-k+1:n} \in E', \\
s_{k,l} & \text{if } 0 \leq k \leq n-1 \land 1 \leq l \leq n-k \land \widehat{X}_{n-k-l} = r \land \widehat{X}_{n-k-l+1:n} \in \overline{E}, \\
s'_n & \text{if } \widehat{X}_{0:n} \in E', \\
\widehat{s}'_{k,n-k} & \text{if } 0 \leq k \leq n-1 \land \widehat{X}_{0:k} \in E' \land \widehat{X}_{k+1:n} \in \overline{E}, \\
s''_n & \text{if } \widehat{X}_{0:n} \in \overline{E}, \\
f_i & \text{if } \widehat{X}_{n} = f_i.
\end{cases}
\end{align*}

In words, $\widehat{V}_n = s_k$ if, by step $n$, $\widehat{X}$ has not left $S$, has visited $r$, the last time it visited $r$ was $k$ steps before, and has not left $E$ since then; $\widehat{V}_n = s_{k,l}$ if $\widehat{X}$ has not left $S$, has visited $r$, the last time it visited $r$ was $k + l$ steps before and, since then, has been first $k + 1$ steps in $E$ and, after that, $l$ steps in $\overline{E}$; $\widehat{V}_n = s'_n$ if, by step $n$, $\widehat{X}$ has not left $E'$; $\widehat{V}_n = s'_{k,n-k}$ if, by step $n$, $\widehat{X}$ has been in $E'$ the first $k + 1$ steps and, after that,
has been in \( E \) \( n - k \) steps; \( \hat{V}_n = s'_n \) if, by step \( n \), \( \hat{X} \) has not left \( E \); and \( \hat{V}_n = f_i \) if, by step \( n \), \( \hat{X} \) has been absorbed into \( f_i \). Note that \( \hat{V}_n = s_0 \) if and only if \( \hat{X}_n = r \) and that \( \hat{V}_n = f_i \) if and only if \( \hat{X}_n = f_i \). Let

\[
a(k) = \sum_{i \in E} \pi_i(k), \tag{2.1}
\]

\[
a(k, l) = \sum_{i \in E} \pi_i(k, l), \tag{2.2}
\]

\[
a'(k) = \sum_{i \in E'} \pi'_i(k), \tag{2.3}
\]

\[
a'(k, l) = \sum_{i \in E'} \pi'_i(k, l), \tag{2.4}
\]

\[
a''(k) = \sum_{i \in E} \pi''_i(k), \tag{2.5}
\]

\[
u_k = \frac{\sum_{i \in E} \pi_i(k) P_{i,E'}}{a(k)}, \tag{2.6}
\]

\[
v'_k = \frac{\sum_{j \in E} \pi_j(k) P_{j,f_i}}{a(k)}, \tag{2.7}
\]

\[
h_k = \frac{\sum_{i \in E} \pi_i(k) P_{i,E}}{a(k)}, \tag{2.8}
\]

\[
w_{k, l} = \frac{\sum_{i \in E} \pi_i(k, l) P_{i,E}}{a(k, l)}, \tag{2.9}
\]

\[
q_{k, l} = \frac{\sum_{i \in E} \pi_i(k, l) P_{i,E}}{a(k, l)}, \tag{2.10}
\]

\[
v'_{k, l} = \frac{\sum_{j \in E} \pi_j(k, l) P_{j,f_i}}{a(k, l)}, \tag{2.11}
\]

\[
w'_k = \frac{\sum_{i \in E} \pi_i'(k) P_{i,E'}}{a'(k)}, \tag{2.12}
\]

\[
v''_k = \frac{\sum_{j \in E} \pi_j'(k) P_{j,f_i}}{a'(k)}, \tag{2.13}
\]

\[
h'_k = \frac{\sum_{i \in E} \pi_i'(k) P_{i,E}}{a'(k)}, \tag{2.14}
\]

\[
w'_{k, l} = \frac{\sum_{i \in E} \pi_i'(k, l) P_{i,E}}{a'(k, l)}, \tag{2.15}
\]

\[
q'_{k, l} = \frac{\sum_{i \in E} \pi_i'(k, l) P_{i,E}}{a'(k, l)}, \tag{2.16}
\]

\[
v''_{k, l} = \frac{\sum_{j \in E} \pi_j'(k, l) P_{j,f_i}}{a'(k, l)}, \tag{2.17}
\]
Note that, being $P_{r,E} > 0$ (by condition C10) and $P_{r,i} > 0$, $i \in E'$, there will exist $i \in E$ with $\pi_i(k) > 0$ for all $k \geq 0$, implying $a(k) > 0$ for all $k \geq 0$. Also, for $k$ such that $a(k, 1) > 0$, we have $\pi_i(k, 1) > 0$ for some $i \in E$ and, since $P_{r,i} > 0$, $i \in E$, there will exist $i \in E$ with $\pi_i(k, l) > 0$ for all $l \geq 1$, implying $a(k, l) > 0$ for all $l \geq 1$.

In addition, assuming $\pi_E > 0$, $\pi'_i(0) > 0$ for some $i \in E'$ and, since $P_{r,i} > 0$, $i \in E'$, there will exist $i \in E'$ with $\pi'_i(k) > 0$ for all $k \geq 0$, implying $a'(k) > 0$ for all $k \geq 0$.

Assuming $\pi_E > 0$, for $k$ such that $a'(k, 1) > 0$, $\pi'_i(k, 1) > 0$ for some $i \in E$ and, since $P_{r,i} > 0$, $i \in E$, there will exist $i \in E$ with $\pi'_i(k, l) > 0$, implying $a'(k, l) > 0$ for all $l \geq 1$. Finally, assuming $\pi_E > 0$, $\pi'_i(0) > 0$ for some $i \in E$ and, since $P_{r,i} > 0$, $i \in E$, there will exist $i \in E$ with $\pi'_i(k) > 0$ for all $k \geq 0$, implying $a'(k) > 0$ for all $k \geq 0$.

Assume $x_E > 0$ and $x_{E'} > 0$. Then, it has been shown in Temsamani and Carrasco (2004) that $\hat{V}$ is a DTMC with reachable state space $E_V \cup \hat{E}_V \cup \{f_1, f_2, \ldots, f_A\}$, $E_V = \{s_k, k \geq 0\} \cup \{s_k', k \geq 0\}$, $\hat{E}_V = \{s_{k,l}: k \geq 0 \land a(k, 1) > 0 \land l \geq 1\} \cup \{s_{k,l}', k \geq 0\}$, initial probability distribution $P[\hat{V}_0 = s_0] = x_r$, $P[\hat{V}_0 = s'_0] = x_{E'}$, $P[\hat{V}_0 = f_i] = x_{f_i}$, $1 \leq i \leq A$, $P[\hat{V}_0 = i] = 0$, $i \not\in \{s_0, s_0', s_{0,l}, f_1, f_2, \ldots, f_A\}$, and (possibly) nonnull transition probabilities $P[\hat{V}_{n+1} = s_0 | \hat{V}_n = s_0] = P_{r,s_0}$, $P[\hat{V}_{n+1} = s_{k+1} | \hat{V}_n = s_k] = w_k$, $P[\hat{V}_{n+1} = s_{k,l} | \hat{V}_n = s_{k,l}] = q_{k,l}$, $P[\hat{V}_{n+1} = f_i | \hat{V}_n = s_{k,l}] = h_{k,l}$. The state transition diagram of $\hat{V}$ has, for the case $x_{E'} > 0$ and $x_E > 0$, two combs and a string of states as illustrated in Fig. 2 for the case $A = 1$. The first comb has as a back the states $s_k$ and as teeth the strings of states $s_{k,l}$ with $k$ fixed. The second comb has as a back the states $s'_k$ and as teeth the strings of states $s'_{k,l}$ with $k$ fixed. The string includes the states $s'_k$. When $x_E = 0$, $\hat{V}$ loses the second comb. When $x_E > 0$, $\hat{V}$ loses the string of states $s'_k$. Formally, the state space of $\hat{V}$ can be defined in the general case as $E_V \cup \hat{E}_V \cup \{f_1, f_2, \ldots, f_A\}$, where, when $x_E = 0$, $E_V$ does not include the states $s'_k$ and $\hat{E}_V$ does not include the states $s'_{k,l}$ and, when $x_E = 0$, $E_V$ does not include the states $s'_k$.

Let $V = \{V(t); t \geq 0\}$ be the CTMC obtained by derandomizing $\hat{V}$ with rate $\Lambda_E$ in $E_V$ and rate $\Lambda_{E'}$ in $\hat{E}_V \cup \{f_1, f_2, \ldots, f_A\}$. The CTMC $V$ has same state space and initial probability distribution as $\hat{V}$. Figure 3 illustrates the state transition diagram of $V$ for the case $x_{E'} > 0$, $x_E > 0$ and $A = 1$.

All developments up to now (with the generalization to the case $A \geq 0$) are taken from Temsamani and Carrasco (2004). Let $I_e$ denote the indicator function returning the value 1 if condition $c$ is satisfied and the value 0 otherwise and let,

\begin{align*}
    u''_k &= \frac{\sum_{i \in E} \pi_i'(k) P_{i,k}}{a''(k)}, \\
    q''_i &= \frac{\sum_{i \in E} \pi_i'(k) P_{i,r}}{a''(k)}, \\
    v''_k &= \sum_{i \in E} \pi_i'(k) P_{j,k}. \tag{2.20}
\end{align*}
Figure 2. State transition diagram of the DTMC $\hat{V}$ for the case $x_{\alpha} > 0$, $x_{\tau} > 0$, and $A = 1$. (There can exist transitions to $f_1$ from any state and transitions to $s_0$ from any state $s_{k,j}$, $s_{k,j}'$, and $s_{k,j}''$.)

$\text{Figure 2.}$ $\text{State transition diagram of the DTMC } \hat{V} \text{ for the case } x_{\alpha} > 0, x_{\tau} > 0, \text{ and } A = 1. \text{ (There can exist transitions to } f_1 \text{ from any state and transitions to } s_0 \text{ from any state } s_{k,j}, s_{k,j}', \text{ and } s_{k,j}''.)$
Figure 3. State transition diagram of the CTMC $V$ for the case $z_G > 0$, $z_F > 0$, and $A = 1$. (There can exist transitions to $f_1$ from any state and transitions to $s_0$ from any state $s_k,i$, $s_k,i'$, and $s_k''$.)
conventionally, the product of 0 by a non-defined quantity be equal to 0. The key
to generalize the method is the following result:

**Proposition 2.1.** For \( i \in S \),

\[
P[X(t) = i] = \{ \sum_{k=0}^{\infty} \frac{\pi_i(k)}{a(k)} P[V(t) = s_k] + \sum_{k=0}^{\infty} \frac{\pi_i(k, l)}{a(k, l)} P[V(t) = s_{k,l}] \\
+ \sum_{k=0}^{\infty} \frac{\pi_i'(k, l)}{a'(k, l)} P[V(t) = s_{k,l}'] \} + \sum_{k=0}^{\infty} \frac{\pi_i''(k, l)}{a''(k, l)} P[V(t) = s_{k,l}'']
\]

**Proof.** See Temsamani and Carrasco (2005). \( \square \)

Let \( \text{ETRR}^V(t) \) and \( \text{EARR}^V(t) \) be, respectively, the expected transient reward rate
and the expected averaged reward rate of \( V \) with the reward rate structure:

\[
r_{f_i}' = r_f, \quad \quad \quad (2.21)
\]

\[
r_{s_k}' = b(k) = \sum_{i \in E} r_i \pi_i(k), \quad \quad \quad (2.22)
\]

\[
r_{s_{k,l}}' = b(k, l) = \sum_{i \in E} r_i \pi_i(k, l), \quad \quad \quad (2.23)
\]

\[
r_{s_k}'' = b'(k) = \sum_{i \in E} r_i \pi_i'(k), \quad \quad \quad (2.24)
\]

\[
r_{s_{k,l}}'' = b'(k, l) = \sum_{i \in E} r_i \pi_i'(k, l), \quad \quad \quad (2.25)
\]

\[
r_{s_k}''' = b''(k) = \sum_{i \in E} r_i \pi_i''(k), \quad \quad \quad (2.26)
\]

Then:

**Theorem 2.1.** \( \text{ETRR}^V(t) = \text{ETRR}(t) \) and \( \text{EARR}^V(t) = \text{EARR}(t) \).

**Proof.** Using (proof of Theorem 1 of Temsamani and Carrasco, 2004) \( P[V(t) = f_i] = P[X(t) = f_i], \ 1 \leq i \leq A \), Proposition 2.1, and (2.21)--(2.26):

\[
\text{ETRR}(t) = \sum_{i \in \Omega} r_i P[X(t) = i] = \sum_{i \in S} r_i P[X(t) = i] + \sum_{i=1}^{A} r_{f_i} P[X(t) = f_i]
\]

\[
= \sum_{k=0}^{\infty} \frac{\sum_{i \in E} r_i \pi_i(k)}{a(k)} P[V(t) = s_k] + \sum_{k=0}^{\infty} \frac{\sum_{i \in E} r_i \pi_i'(k, l)}{a(k, l)} P[V(t) = s_{k,l}']
\]
\[ + I_{x_E > 0} \left( \sum_{k=0}^{\infty} \frac{\sum_{r \in E'} r_i \pi'_i(k)}{a'(k)} P[V(t) = s'_k] \right) \]
\[ + \sum_{k=0}^{\infty} I_{a(k,1) > 0} \sum_{l=1}^{\infty} \frac{\sum_{r \in E} r_i \pi'_i(k, l)}{a'(k, l)} P[V(t) = s'_k,l] \]
\[ + I_{x_E > 0} \sum_{k=0}^{\infty} \frac{\sum_{r \in E} r_i \pi''_i(k)}{a''(k)} P[V(t) = s''_k] \]
\[ + \sum_{k=0}^{\infty} I_{a(k,1) > 0} \sum_{l=1}^{\infty} \frac{\sum_{r \in E} r_i \pi''_i(k, l)}{a''(k, l)} P[V(t) = s''_k,l] \]
\[ = \sum_{k=0}^{\infty} b(k) P[V(t) = s_k] + \sum_{k=0}^{\infty} I_{a(k,1) > 0} \sum_{l=1}^{\infty} b(k, l) P[V(t) = s_k,l] \]
\[ + I_{x_E > 0} \sum_{k=0}^{\infty} b'(k) P[V(t) = s'_k] + \sum_{k=0}^{\infty} I_{a(k,1) > 0} \sum_{l=1}^{\infty} b'(k, l) P[V(t) = s'_k,l] \]
\[ + I_{x_E > 0} \sum_{k=0}^{\infty} b''(k) P[V(t) = s''_k] + \sum_{i=1}^{A} r_i P[V(t) = f_i] \]
\[ = \text{ETRR}^V(t). \]

Finally, using \( EARR(t) = (1/t) \int_0^t \text{ETRR}(\tau)d\tau \) and \( EARR^V(t) = (1/t) \int_0^t \text{ETRR}^V(\tau)d\tau \),
\[ EARR(t) = \frac{1}{t} \int_0^t \text{ETRR}(\tau)d\tau = \frac{1}{t} \int_0^t \text{ETRR}^V(\tau)d\tau = \text{EARR}^V(t). \]

The truncated transformed rewarded CTMC, \( V_T \), is obtained from \( V \) by introducing an absorbing state \( a \) with null reward rate capturing the truncated behavior and:

1. keeping the states \( s_k \) up to \( s_K, K \geq 1 \), and directing to \( a \) the transition rates from \( s_k \);
2. for each \( k, 0 \leq k \leq K - 1 \), for which \( a(k, 1) > 0 \), keeping the states \( s_k,l \) up to \( l = K_k \geq 1 \) and directing the transition rates from \( s_k,k \) to \( a \);
3. if \( x_E > 0 \), keeping the states \( s'_k \) up to \( s'_L, L \geq 1 \), and directing to \( a \) the transition rates from \( s'_k \);
4. if \( x_E > 0 \), for each \( k, 0 \leq k \leq L - 1 \), for which \( a'(k, 1) > 0 \), keeping the states \( s'_k,l \) up to \( l = L_k \geq 1 \) and directing the transitions rates from \( s_k,k \) to \( a \); and,
5. if \( x_E > 0 \), keeping the states \( s''_k \) up to \( s''_M, M \geq 1 \), and directing to \( a \) the transition rates from \( s''_k \). The CTMC \( V_T \) can be defined from \( V \) as:

\[ V_T(t) = \begin{cases} 
V(t) & \text{if, by time } t, V \text{ has not exited state } s_K, \\
\text{a state } s_k,k, \text{ a state } s'_k,l, \text{ or state } s''_k & \text{otherwise.}
\end{cases} \tag{2.27} \]

The initial probability distribution of \( V_T \) is the same as that of \( V \), i.e., \( P[V(0) = s_0] = \alpha, P[V(0) = s'_0] = \alpha_E, \ P[V(0) = s''_0] = \alpha_T, \ P[V(0) = f_i] = \alpha_f, \ 1 \leq i \leq A, \ P[V(0) = i] = 0, i \not\in \{s_0, s'_0, f_1, f_2, \ldots, f_A\} \). Let \( E^T_V \) denote the set of states in \( E_V \) kept in \( V_T \) and let \( \overline{E}^T_V \) denote the set of states in \( \overline{E} \) kept in \( V_T \). Note that the state space of \( V_T \) is \( E^T_V \cup \overline{E}^T_V \cup \{f_1, f_2, \ldots, f_A, a\} \).
The truncated transformed rewarded CTMC model \( V_T \) yields approximate values \( \text{ETRR}^a(t) \) and \( \text{EARR}^a(t) \), for, respectively, \( \text{ETRR}(t) \) and \( \text{EARR}(t) \). Formally, \( \text{ETRR}^a(t) \) and \( \text{EARR}^a(t) \) are, respectively, the expected transient reward rate and expected averaged reward rate of \( V_T \). Let \( r_{\max} = \max_{i \in \Omega} r_i \). The following two theorems upper bound the model truncation error for, respectively, the measure \( \text{ETRR}(t) \) and the measure \( \text{EARR}(t) \).

**Theorem 2.2.** \( 0 \leq \text{ETRR}(t) - \text{ETRR}^a(t) \leq r_{\max} P[V_T(t) = a] = \text{ETRR}^c(t) \).

**Proof.** We can write:

\[
\text{ETRR}(t) - \text{ETRR}^a(t) = \sum_{i \in E_Y \cup \overline{E_Y}} r_i P[V(t) = i] + \sum_{i=1}^A r_i P[V(t) = f_i] \\
- \left( \sum_{i \in E_Y \cup \overline{E_Y}} r_i P[V_T(t) = i] + \sum_{i=1}^A r_i P[V_T(t) = f_i] \right) \\
= \sum_{i \in (E_Y \cup \overline{E_Y}) \cup (E_Y \cup \overline{E_Y})} r_i P[V(t) = i] + \sum_{i \in E_Y \cup \overline{E_Y}} r_i P[V(t) = i] \\
- P[V_T(t) = i] + \sum_{i=1}^A r_i (P[V(t) = f_i] - P[V_T(t) = f_i]).
\]

According to (2.27), \( P[V_T(t) = i] \leq P[V(t) = i], \quad i \in E_Y \cup \overline{E_Y} \) and \( P[V_T(t) = f_i] \leq P[V(t) = f_i], \quad 1 \leq i \leq A \). implying \( \text{ETRR}(t) - \text{ETRR}^a(t) \geq 0 \). Also, since \( \sum_{i \in E_Y \cup \overline{E_Y}} P[V(t) = i] + \sum_{i=1}^A P[V(t) = f_i] = 1 \) and \( \sum_{i \in E_Y \cup \overline{E_Y}} P[V_T(t) = i] + \sum_{i=1}^A P[V_T(t) = f_i] = 1 \):

\[
\text{ETRR}(t) - \text{ETRR}^a(t) \\
\leq r_{\max} \left( \sum_{i \in (E_Y \cup \overline{E_Y}) \cup (E_Y \cup \overline{E_Y})} P[V(t) = i] + \sum_{i \in E_Y \cup \overline{E_Y}} (P[V(t) = i] - P[V_T(t) = i]) \\
+ \sum_{i=1}^A (P[V(t) = f_i] - P[V_T(t) = f_i]) \right) \\
= r_{\max} \left( \sum_{i \in E_Y \cup \overline{E_Y}} P[V(t) = i] + \sum_{i=1}^A P[V(t) = f_i] \\
- \sum_{i \in E_Y \cup \overline{E_Y}} P[V_T(t) = i] - \sum_{i=1}^A P[V_T(t) = f_i] \right) \\
= r_{\max} \left( 1 - \sum_{i \in E_Y \cup \overline{E_Y}} P[V_T(t) = i] - \sum_{i=1}^A P[V_T(t) = f_i] \right) \\
= r_{\max} P[V_T(t) = a] = \text{ETRR}^c(t). \]

**Theorem 2.3.** \( 0 \leq \text{EARR}(t) - \text{EARR}^a(t) \leq (r_{\max}/t) \int_0^t P[V_T(\tau) = a]d\tau = \text{EARR}^c(t) \).
Proof. Using \( \text{EARR}(t) = (1/t) \int_0^t \text{ETRR}(\tau) \, d\tau \), \( \text{EARR}^a(t) = (1/t) \int_0^t \text{ETRR}^a(\tau) \, d\tau \), and Theorem 2.2,
\[
\text{EARR}(t) - \text{EARR}^a(t) = \frac{1}{t} \int_0^t \text{ETRR}(\tau) \, d\tau - \frac{1}{t} \int_0^t \text{ETRR}^a(\tau) \, d\tau \\
= \frac{1}{t} \int_0^t (\text{ETRR}(\tau) - \text{ETRR}^a(\tau)) \, d\tau,
\]

\[0 \leq \text{EARR}(t) - \text{EARR}^a(t) \leq \frac{r_{\text{max}}}{t} \int_0^t P[V_T(\tau) = a] \, d\tau. \]

The upper bound for the model truncation error for the \( \text{ETRR}(t) \) measure given by Theorem 2.2 is formally identical to the model truncation error upper bound for the less general measure considered in Temsamani and Carrasco (2004). Then, letting \( \gamma_K = \{k : 0 \leq k \leq K - 1 \wedge a(k, 1) > 0\} \) and \( \gamma_L = \{k : 0 \leq k \leq L - 1 \wedge a'(k, 1) > 0\} \), we can state the following result:

Theorem 2.4.
\[
\text{ETRR}^c(t) \leq I_{\tau_T > 0} r_{\text{max}} a''(M) \sum_{k=M+1}^{\infty} e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^k}{k!}} \\
+ I_{\tau_T > 0} \left( r_{\text{max}} a'(L) \sum_{k=L+1}^{\infty} e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^k}{k!}} \\
+ \sum_{k \in \gamma_L} r_{\text{max}} a'(k, L_k) \sum_{l=k+1}^{\infty} e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^l}{l!}} \right) \\
+ r_{\text{max}}(\alpha_S - a''(M)) a(K) \sum_{k=K+1}^{\infty} (k - K) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^k}{k!}} \\
+ \sum_{k \in \gamma_k} r_{\text{max}}(\alpha_S - a''(M)) a(k, K_k) \sum_{l=k+1}^{\infty} (l-k) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^l}{l!}}.
\]

The following theorem gives an upper bound for the model truncation error for the \( \text{EARR}(t) \) measure.

Theorem 2.5.
\[
\text{EARR}^c(t) \leq I_{\tau_T > 0} r_{\text{max}} a''(M) \sum_{k=M+1}^{\infty} (k - M - 1) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^k}{k!}} \\
+ I_{\tau_T > 0} \left( r_{\text{max}} a'(L) \frac{\Lambda_{E} t}{\Lambda_{E} t} \sum_{k=L+2}^{\infty} (k - L - 1) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^k}{k!}} \\
+ \sum_{k \in \gamma_L} r_{\text{max}} a'(k, L_k) \frac{\Lambda_{E} t}{\Lambda_{E} t} \sum_{l=k+2}^{\infty} (l - k - 1) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^l}{l!}} \right) \\
+ r_{\text{max}}(\alpha_S - a''(M)) a(K) \frac{\Lambda_{E} t}{\Lambda_{E} t} \sum_{k=K+2}^{\infty} (k - K)(k - K - 1) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^k}{k!}} \\
+ \sum_{k \in \gamma_k} r_{\text{max}}(\alpha_S - a''(M)) a(k, K_k) \frac{\Lambda_{E} t}{\Lambda_{E} t} \sum_{l=k+2}^{\infty} (l-k)(l-k-1) e^{-\Lambda_{E} t \frac{(\Lambda_{E} t)^l}{l!}}.
\]
Proof. From Theorems 2.2–2.4.

\[ EARR^c(t) = \frac{r_{\max}}{t} \int_0^t P[V_\tau(\tau) = a] d\tau = \frac{1}{t} \int_0^t ETRR^c(\tau) d\tau \]

\[ \leq I_{\alpha > 0} \left( \frac{r_{\max} a''(M)}{t} \sum_{k=M+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} (\Lambda_{\alpha \tau}^c)^k \frac{k!}{k!} d\tau \right. \]

\[ + I_{\alpha > 0} \left( \frac{r_{\max} a'(L)}{t} \sum_{k=L+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} (\Lambda_{\alpha \tau}^c)^k \frac{k!}{k!} d\tau \right. \]

\[ + \left. \sum_{k=K}^{\infty} \frac{r_{\max} a(1)}{t} \left. \sum_{l=k+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} (\Lambda_{\alpha \tau}^c)^k \frac{k!}{k!} d\tau \right) \right) \]

\[ + \frac{r_{\max} (a - a''(M)) a(K)}{t} \sum_{k=K}^{\infty} \sum_{l=k+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} (\Lambda_{\alpha \tau}^c)^k \frac{k!}{k!} d\tau \]

Using \( \int_0^t e^{-\Lambda t} (\Lambda t)^k / k! \) \( d\tau = (1/\Lambda) \sum_{l=k+1}^{\infty} e^{-\Lambda t}(\Lambda t)^l / l! \):

\[ \sum_{k=M+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!} d\tau = \frac{1}{\Lambda_{E}} \sum_{k=M+2}^{\infty} (k-M-1)e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!}, \]

\[ \sum_{k=L+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!} d\tau = \frac{1}{\Lambda_{E}} \sum_{k=L+2}^{\infty} (k-L-1)e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!}, \]

\[ \sum_{l=k+1}^{\infty} \int_0^t e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!} d\tau = \frac{1}{\Lambda_{E}} \sum_{l=k+2}^{\infty} (l-k-1)e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!}, \]

\[ \sum_{k=K}^{\infty} (k-K) \int_0^t e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!} d\tau = \frac{1}{\Lambda_{E}} \sum_{k=K+2}^{\infty} \left( \sum_{l=K+1}^{l=l+k-1} \right) e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!} \]

\[ = \frac{1}{\Lambda_{E}} \sum_{k=K+2}^{\infty} (k-K)(k-K-1) \frac{e^{-\Lambda_{\alpha \tau}^c} (\Lambda_{\alpha \tau}^c)^k}{k!}, \]

\[ \sum_{l=k+1}^{\infty} (l-k) \int_0^t e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{l!} d\tau = \frac{1}{\Lambda_{E}} \sum_{l=k+2}^{\infty} (l-k)(l-k-1) \frac{e^{-\Lambda_{\alpha \tau}^c} (\Lambda_{\alpha \tau}^c)^l}{l!}, \]

and the result follows. \( \square \)

The truncation parameters \( K, L, M, K_1, k \in \gamma_K, \) and \( L_1, k \in \gamma_L, \) have to be selected so that the upper bound for the model truncation error given by Theorem 2.4 for the measure \( ETTR(t) \) and by Theorem 2.5 for the measure \( EARR(t) \) is \( \leq \varepsilon_2 / 2. \) For the \( ETTR(t) \) measure the truncation parameters are selected as follows. First, for the case \( x_{\alpha} > 0, M \) is selected using:

\[ M = \min \left\{ m \geq 1 : r_{\max} a''(m) \sum_{k=m+1}^{\infty} e^{-\Lambda_{\alpha \tau}^c} \frac{(\Lambda_{\alpha \tau}^c)^k}{k!} \leq \varepsilon_1 \right\}. \]
where $\varepsilon_1 = \varepsilon/6$ if $x_E > 0$ and $\varepsilon_1 = \varepsilon/4$ if $x_E = 0$. The truncation parameter $K$ is then chosen using:

$$K = \min \left\{ m \geq 1 : r_{\max}(x_S - a''(M))a(m) \sum_{k=m+1}^{\infty} (k-m)e^{-\Lambda_E t} \frac{(\Lambda_E t)^k}{k!} \leq \varepsilon_2 \right\},$$

where $\varepsilon_2 = \varepsilon/12$ if $x_E > 0$ and $\varepsilon_2 = \varepsilon/8$ if $x_E = 0$. $K$ is chosen using:

$$K = \min \left\{ m \geq 1 : r_{\max}(x_S - a''(M))a(m) \sum_{l=k+1}^{\infty} (l-k)e^{-\Lambda_E t} \frac{(\Lambda_E t)^l}{l!} \leq \varepsilon_2 \right\}.$$ 

Finally, for the case $x_E > 0$, the truncation parameter $L$ is chosen using:

$$L = \min \left\{ m \geq 1 : r_{\max}a'(m) \sum_{k=m+1}^{\infty} e^{-\Lambda_E t} \frac{(\Lambda_E t)^k}{k!} \leq \varepsilon_3 \right\},$$

where $\varepsilon_3 = \varepsilon/12$ if $x_E > 0$ and $\varepsilon_3 = \varepsilon/8$ if $x_E = 0$, and the truncation parameters $L_k$, $k \in \gamma_L$, are chosen using:

$$L_k = \min \left\{ m \geq 1 : r_{\max}a'(k, m) \sum_{l=k+1}^{\infty} e^{-\Lambda_E t} \frac{(\Lambda_E t)^l}{l!} \leq \varepsilon_3 \right\}.$$ 

For the measure $EARR(t)$, for the case $x_E > 0$, $M$ is selected using:

$$M = \min \left\{ m \geq 1 : \frac{r_{\max}a''(m)}{\Lambda_E t} \sum_{k=m+2}^{\infty} (k-m-1)e^{-\Lambda_E t} \frac{(\Lambda_E t)^k}{k!} \leq \varepsilon_1 \right\}.$$ 

The truncation parameter $K$ is then chosen using:

$$K = \min \left\{ m \geq 1 : \frac{r_{\max}(x_S - a''(M))a(m)}{\Lambda_E t} \sum_{k=m+2}^{\infty} \frac{(k-m)(k-m-1)}{2} e^{-\Lambda_E t} \frac{(\Lambda_E t)^k}{k!} \leq \varepsilon_2 \right\}.$$ 

The truncation parameters $K_k$, $k \in \gamma_K$, are chosen using:

$$K_k = \min \left\{ m \geq 1 : \frac{r_{\max}(x_S - a''(M))a(k, m)}{\Lambda_E t} \right\} \times \sum_{l=k+2}^{\infty} \frac{(l-k)(l-k-1)}{2} e^{-\Lambda_E t} \frac{(\Lambda_E t)^l}{l!} \leq \varepsilon_2 \right\}.$$ 

Finally, for the case $x_E > 0$, the truncation parameter $L$ is chosen using:

$$L = \min \left\{ m \geq 1 : \frac{r_{\max}a'(m)}{\Lambda_E t} \sum_{k=m+2}^{\infty} (k-m-1)e^{-\Lambda_E t} \frac{(\Lambda_E t)^k}{k!} \leq \varepsilon_3 \right\}.$$
and the truncation parameters \( L_k, k \in \gamma'_L \), are chosen using:

\[
L_k = \min \left\{ m \geq 1 : \frac{r_{\max}d'(k, m)}{\Lambda_E t} \sum_{l=k+2}^{\infty} \frac{(l - k - 1) e^{-\Lambda_E t} (\Lambda_E t)^l}{l!} \leq \frac{\varepsilon_3}{|\gamma'_L|} \right\}
\]

It has been proved in Temsamani and Carrasco (2004) that the upper bound for the model truncation error for the \( ETTR(\tau) \) measure given by Theorem 2.4 is increasing with \( t \). Since the upper bound for the model truncation error for the \( EARR(t) \) measure given by Theorem 2.5 is the averaged value in the interval \([0, t] \) of the upper bound given by Theorem 2.4, it follows that the upper bound given by Theorem 2.5 is also increasing with \( t \). Then, if either \( ETTR(\tau) \) or \( EARR(t) \) has to be computed for several values of \( t \), the truncation parameters can be selected using the largest \( t \).

To clarify, Figs. 4 and 5 give a C-like algorithmic description of the method for the \( ETTR(\tau) \) measure. The algorithm has as inputs the CTMC \( X \), the number of absorbing states \( A \), the reward rates \( r_i, i \in \Omega \), an initial probability distribution row vector \( s = (s_i)_{i \in \Omega} \), the subset \( E \subseteq S \), the regenerative state \( r \in E \), the allowed error \( \varepsilon \), the number of time points \( n \) at which which estimates for the measure have to be computed, and the time points \( t_1, t_2, \ldots, t_n \). The algorithm has as outputs the estimates for the measure at the time points \( t_1, E\overline{TRR}(t_1), E\overline{TRR}(t_2), \ldots, E\overline{TRR}(t_n) \). It is assumed that conditions C1–C10 regarding the structure of \( X \) and the selection of the subset \( E \) and the regenerative state \( r \in E \) are satisfied. The truncated transformed CTMC model, called \( V \) in the algorithmic description, is built using the functions \( add\_state(V, s, p) \) and \( add\_transition(V, s, s', \lambda) \). The first function adds to \( V \) the state \( s \) with initial probability \( p \); the second function adds to \( V \) a transition rate \( \lambda \) from state \( s \) to state \( s' \). The model truncation error is controlled for \( t_{\max} = \max\{t_1, t_2, \ldots, t_n\} \). The algorithm makes two traversals of the backs of the combs: the first one to determine \( K \) and \( |\gamma'_K| \) (called \( n_k \) in the algorithm), and, if \( s_{2'} > 0 \), \( L \) and \( |\gamma'_L| \) (also called \( n_k \) in the algorithm), and the second one to build the teeth. The method for \( EARR(t) \) can be described similarly, with the obvious changes.

The method requires the computation of the summatories

\[
S(m) = \sum_{k=m+1}^{\infty} \frac{e^{-\Lambda t}(\Lambda t)^k}{k!},
\]

\[
S'(m) = \sum_{k=m+1}^{\infty} (k - m) \frac{e^{-\Lambda t}(\Lambda t)^k}{k!},
\]

\[
S''(m) = \sum_{k=m+2}^{\infty} (k - m - 1) \frac{e^{-\Lambda t}(\Lambda t)^k}{k!},
\]

\[
S'''(m) = \sum_{k=m+2}^{\infty} \frac{(k - m)(k - m - 1)}{2} \frac{e^{-\Lambda t}(\Lambda t)^k}{k!},
\]

for \( \Lambda = \Lambda_E \) or \( \Lambda = \Lambda_E' \), \( t = t_{\max} \), and increasing values of \( m \). Efficient and numerically stable procedures for computing \( S(m) \), \( S'(m) \), and \( S''(m) \) are described in Carrasco (2002b) and Carrasco (2003). Since \( S''(m) = S(m + 1) \), an efficient and numerically stable procedure for computing \( S''(m) \) can be obtained easily by adapting the procedure for computing \( S'(m) \).

We note that, once \( P \) has been computed, the transition rates of the truncated transformed model are obtained without subtractions. Thus, the method has
A Generalized Method for the Transient Analysis

Figure 4. Algorithmic description of split regenerative randomization for the ETTR$^\prime$/$\alpha_\text{E}$ measure.
652 Temsamani and Carrasco

Figure 5. Algorithmic description of split regenerative randomization for the ETTR(t) measure (continuation).

the same excellent numerical stability as the standard randomization method. In addition, the computation error is well-controlled and can be specified in advance.

3. Theoretical Properties

The model truncation error bound for the ETTR(t) measure is formally identical to the model truncation error bound for the less general measure considered in Temsamani and Carrasco (2004). Then, letting $K_T = \sum_{t \in T} K_k$ and $L_T = \sum_{t \in T} L_k$, we have the following result:

**Theorem 3.1.** The number of steps, $K$, $L$, $M$, $K_T$, and $L_T$, required in the split regenerative randomization method for the ETTR(t) measure are, respectively, $O(\log(\Lambda T t / \varepsilon))$, $O(\log(1 / \varepsilon))$, $O((\log(\Lambda_T t / \varepsilon))^2)$, and $O((\log(1 / \varepsilon))^2)$.
A similar result is available regarding the $EARR(t)$ measure:

**Theorem 3.2.** The number of steps, $K$, $L$, $M$, $K_{\overline{\pi}}$, and $L_{\overline{\pi}}$, required in the split regenerative randomization method for the $EARR(t)$ measure are, respectively, $O(\log(\Lambda_{ET}t/\varepsilon))$, $O(\log(1/\varepsilon))$, $O(\log(1/\varepsilon))$, $O(\log(\Lambda_{ET}t/\varepsilon)^2)$, and $O((\log(1/\varepsilon))^2)$.

**Proof.** The terms of the model truncation error bound used in the split regenerative randomization method for the $EARR(t)$ measure are the averaged values in the interval $[0, t]$ of the corresponding terms of the model truncation error bound for the $ETRR(t)$ measure. Furthermore, the terms of the model truncation error bound for the $ETRR(t)$ measure increase with $t$. Then, the terms of the model truncation error bound for $EARR(t)$ are not greater than the corresponding terms of the model truncation error bound for $ETRR(t)$ and the result follows from Theorem 3.1. □

Theorems 3.1 and 3.2 tell that $K$, $L$, $M$, $K_{\overline{\pi}}$, and $L_{\overline{\pi}}$ are all smooth functions of $t$ and $\varepsilon$ for both $ETRR(t)$ and $EARR(t)$. That property is called benign behavior and implies that, for large enough $X$ and large enough $t$, the proposed method will be significantly less costly than standard randomization. This is because (1) the cost of the first phase of the method (generation of the truncated transformed model) is made up of components approximately proportional to, respectively, $K$, $L$, $M$, $K_{\overline{\pi}}$ and $L_{\overline{\pi}}$, while the cost of standard randomization is, for large $t$, approximately proportional to $\max_{i\in\Omega} A_{i} t$, and (2) being the maximum output rate of the truncated transformed model at most $(1 + \theta)$ times the maximum output rate of the original model, the cost of the second phase of the method (solution of the truncated transformed model by standard randomization) will scale with the cost of standard randomization at most as the size of the truncated transformed model scales with the size of the original model, $X$.

The performance of the method depends, of course, on the selections for the subset $E$ and the regenerative state $r$, since those selections influence the behavior of $a(k)$, $a'(k)$, $a''(k)$, $a(k, l)$, and $a'(k, l)$, and, then, the required values for the truncation parameters $K$, $L$, $M$, $K_{k}$, $k \in \gamma_{K}$, and $L_{k}$, $k \in \gamma_{L}$. Ideally, $E$ and $r$ should be chosen so that $a(k)$, $a'(k)$, $a''(k)$, $a(k, l)$, and $a'(k, l)$ decrease as fast as possible. For general models, automatic selection of $E$ and $r$ does not seem to be easy in general.

A model class, class $C_{2}$, can, however, be defined for which natural selections for $E$ and $r$ exist, and for models in that class and those natural selections, theoretical results are available assessing approximately the performance of the method in terms of “visible” model characteristics.

The model class $C_{2}$ includes all CTMCs $X$ with finite state space $\Omega$ satisfying the following conditions:

C11. $\Omega = S \cup \{f_{1}, f_{2}, \ldots, f_{A}\}$, $|S| \geq 3$, $A \geq 0$, where the states $f_{i}$, $1 \leq i \leq A$, are absorbing and either all states in $S$ are transient or $X$ has a single recurrent class of states $C \subset S$.

C12. All states are reachable (from some state with nonnull initial probability).

C13. $r_{i} \geq 0$, $i \in \Omega$ and all $r_{f_{i}}$ are different.

C14. There exists a partition $S_{o} \cup S_{1} \cup \cdots \cup S_{N_{c}} \cup \overline{S}_{1} \cup \overline{S}_{2} \cup \cdots \cup \overline{S}_{N_{c}}$ for $S$ satisfying the following properties:

P1. $|S_{o}| = \{o\}$ (i.e., $|S_{o}| = 1$).

P2. If $X$ has a single recurrent class of states $C \subset S$, then $o \in C$. 

\[\text{A Generalized Method for the Transient Analysis 653}\]
P3. $|S_0 \cup S_1 \cup \cdots \cup S_{N_C}| \geq 2$, and $|\bar{S}_1 \cup \bar{S}_2 \cup \cdots \cup \bar{S}_{N_C}| \geq 1$.

P4. $\lambda_{o,S_1 \cup \cdots \cup S_{N_C}} > 0$.

P5. For each $i \in S_k$, $0 < k \leq N_c$, $\lambda_{i,S_i \cup \cdots \cup S_k} = 0$.

P6. For each $i \in \bar{S}_k$, $1 \leq k \leq N_c$, $\lambda_{i,S_i \cup \cdots \cup S_k} = 0$.

P7. $\max_{1 \leq k \leq N_c} \max_{i \in S_k} \lambda_{i,\bar{S}_i \cup \cdots \cup \bar{S}_{k-1} \cup \{f_1, f_2, \ldots, f_a\}}$ is significantly smaller than $\min_{1 \leq k \leq N_c} \min_{i \in \bar{S}_k} \lambda_{i,\bar{S}_i \cup \cdots \cup \bar{S}_{k-1} \cup \{f_1, f_2, \ldots, f_a\}} > 0$.

The class includes failure/repair models with exponential failure and repair time distributions in which repair is deferred until some condition on the subset of failed components is fulfilled, and then, proceeds until the state in which no component is failed is reached, when failure rates are significantly smaller than repair rates. For those models, a partition for $S$ for which properties P1–P7 would be satisfied is the partition in which $S_k$ includes the states without repair and the same number of failed components, with the subsets $S_k$ ordered following increasing number of failed components, and $\bar{S}_k$ includes the states with repair and the same number of failed components, with the subsets $\bar{S}_k$ similarly ordered following increasing number of failed components. Similar failure/repair models with exponential failure time distributions and repair times with acyclic phase-type distributions (Neuts, 1994) (which can be used to fit distributions of non-exponential positive random variables Bobbio and Telek, 1994), are also covered by model class C2, provided that failure rates are significantly smaller than the transition rates of the transient CTMCs defining the phase-type distributions.

With the selection $E = S_0 \cup S_1 \cup \cdots \cup S_{N_c}$ and $r = o$, models in $r = o$, models in class C2 satisfy the conditions applicable. Furthermore, with those selections, the models move “fast” from states in $\bar{E}$ to either state $o$ or a state $f_i$, making those selections natural ones. Let

$$R_E = \frac{\max_{0 \leq k \leq N_c} \max_{i \in S_k} \lambda_i}{\min_{0 \leq k \leq N_c} \min_{i \in S_k} \lambda_i},$$

$$R_\bar{E} = \frac{\max_{1 \leq k \leq N_c} \max_{i \in \bar{S}_k} \lambda_i}{\min_{1 \leq k \leq N_c} \min_{i \in \bar{S}_k} \lambda_i}.$$

Note that once $E$ and $r$ have been identified, both $R_E$ and $R_\bar{E}$ are model characteristics that can be easily estimated. Let

$$\delta = \frac{\max_{1 \leq k \leq N_c} \max_{i \in S_k} \lambda_{i,\bar{S}_i \cup \cdots \cup \bar{S}_{k-1} \cup \{f_1, f_2, \ldots, f_a\}}}{\min_{1 \leq k \leq N_c} \min_{i \in \bar{S}_k} \lambda_{i,\bar{S}_i \cup \cdots \cup \bar{S}_{k-1} \cup \{f_1, f_2, \ldots, f_a\}}}.$$

The $\delta$ can be regarded as a “rarity” parameter measuring how strongly property P7 is satisfied. Then, it has been shown in Temsamani and Carrasco (2004) that with the natural selections for $E$ and $r$, (1) both $a(k)$ and $a'(k)$ are, for $k \to \infty$, upper bounded by functions of the form $C(\binom{k}{p-1})q^k$, $C > 0$, $p$ integer $\geq 1$, where $q_E \approx 1 - 1/R_E$, and (2) $a(k, l)$, $a'(k, l)$, and $a''(l)$ are, for $l \to \infty$, upper bounded by functions of the form $C(\delta)^{\binom{l}{p-1}} \rho(\delta)$, $C(\delta) > 0$, $p(\delta)$ integer $\geq 1$, with $\lim_{\delta \to 0} \rho(\delta) = q_\bar{E} \approx 1 - 1/R_\bar{E}$.

Then, for $R_E$ close to 1, the required $K$ and $L$ should be small and, as $R_E$ gets apart from 1, the required $K$ and $L$ should increase. A similar behavior exhibit $M$, $K^*$, and $L^*$ with respect to $R_\bar{E}$. Moreover, for small $\epsilon$, the required $M$, $K^*$, and $L^*$ will be mainly determined by the decay rate of, respectively, $a(k, l)$, $a'(k, l)$, and $a''(l)$.
and, following the discussion done in Temsamani and Carrasco (2004), for $R_\tau \gg 1$, the required $M$, $K$, and $L$ can be roughly upper bounded by $30R_\tau$. Regarding the truncation parameters $K$ and $L$, for small $\varepsilon$, they can be upper bounded roughly using $a(k) = a'(k) = q_k \approx (1 - 1/R_\epsilon)^k$. Then, for class $C_2$ models with the natural selections for $E$ and $r$, the computational cost of split regenerative randomization can be estimated roughly.

### 4. A Large Example

In this section we analyze the performance of the method and will compare it with that of standard randomization, regenerative randomization, randomization with quasistationarity detection, and, for the $ETRR(t)$ measure, adaptive uniformization using a class $C_1$ performability model of a fault-tolerant multiprocessor including 16 processors interconnected by a 8-node hypercube, as shown in Fig. 6. Processors fail with rate $\lambda_p$; nodes of the hypercube fail with rate $\lambda_N$; links of the hypercube fail with rate $\lambda_L$. A fault of a processor is covered with probability $C_p$; a fault of a node of the hypercube is covered with probability $C_N$. Coverage to link faults is assumed perfect. There is an unlimited number of repairmen. Repair starts when the number of failed components gets $\geq 2$. The repair rate is $\mu_p$ for processors, $\mu_N$ for nodes, and $\mu_L$ for links. A completely down system because there was an uncovered fault is brought to a fully operational state without failed components at rate $\mu_G$. It is assumed the availability of diagnosis and reconfiguration procedures to both determine a subset of interconnected unfailed processors of maximal size and to reconfigure the multiprocessor so that it works using that maximal healthy subset. As reward rates, we take the speedup function of the number of processors in the system at time $t$ and $EARR(t)$ will be the expected speedup of the system averaged over the time interval $[0, t]$. As model parameters we use $\lambda_p = 2 \times 10^{-5}$ h$^{-1}$, $\lambda_N = 10^{-5}$ h$^{-1}$, $\lambda_L = 5 \times 10^{-6}$ h$^{-1}$, $C_p = 0.99$, $C_N = 0.995$, $\mu_p = 0.1$ h$^{-1}$, $\mu_N = 0.05$ h$^{-1}$, $\mu_L = 0.05$ h$^{-1}$, and $\mu_G = 0.2$ h$^{-1}$. Regarding the initial probability distribution, we will consider two cases: (1) the initial state of the system is the state without failed components; and (2) with probability 0.5 the initial state is the state with deferred repair in which processor $P_0$ is the only failed component, and with probability 0.25 the initial state is the state in which processor $P_0$ is the only failed component and repair is underway.

An exact model of the multiprocessor system has an unmanageable size and we will consider instead bounding models with state space $S \cup \{f_1\}$, where $S$ includes the states with up to $N_f$ covered faults and the state in which the system is down due to an uncovered fault and entry into the absorbing state $f_1$ occurs when the exact model enters a state with more than $N_f$ covered faults. A lower (upper) bound for $ETRR(t)$ and $EARR(t)$ is obtained by assigning to the absorbing state $f_1$ a reward rate equal to 0 (12). The bounding models belong to model class $C_1$. Taking $N_f = 4$ is enough to get very tight bounds. Thus, for Case 1 and $t = 100,000$ h, the lower and upper bounds thus obtained for $ETRR(t)$ are $11.760559$ h$^{-1}$ and $11.760562$ h$^{-1}$ and the lower and upper bounds for $EARR(t)$ are $11.762899$ h$^{-1}$ and $11.762901$ h$^{-1}$. With that value of $N_f$, the bounding models have 213,104 states. The reported results are identical for the lower and the upper bounding models. For split regenerative randomization we take for $r$ and $E$ the natural selections, i.e., $r$ is...
Figure 6. Architecture of the fault-tolerant multiprocessor system.

Table 1
Speedups of the multiprocessor system as a function of the maximum number of connected operational processors

<table>
<thead>
<tr>
<th>Processors</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1.96667</td>
</tr>
<tr>
<td>3</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>3.8</td>
</tr>
<tr>
<td>5</td>
<td>4.66667</td>
</tr>
<tr>
<td>6</td>
<td>5.5</td>
</tr>
<tr>
<td>7</td>
<td>6.3</td>
</tr>
<tr>
<td>8</td>
<td>7.06667</td>
</tr>
<tr>
<td>9</td>
<td>7.8</td>
</tr>
<tr>
<td>10</td>
<td>8.5</td>
</tr>
<tr>
<td>11</td>
<td>9.16667</td>
</tr>
<tr>
<td>12</td>
<td>9.8</td>
</tr>
<tr>
<td>13</td>
<td>10.4</td>
</tr>
<tr>
<td>14</td>
<td>10.96667</td>
</tr>
<tr>
<td>15</td>
<td>11.5</td>
</tr>
<tr>
<td>16</td>
<td>12</td>
</tr>
</tbody>
</table>
the single state $o$ without failed components and $E$ includes the states in $S$ without repair. With that natural selection, we have $x_E = 0$ and $x_T = 0$ for Case 1 and $x_E > 0$ and $x_T > 0$ for Case 2. For regenerative randomization we use the selection $r = o$. All CPU times are measured on a Sun-Blade 1000, 4 GB workstation running each method with a unique target time $t$. For all methods we use $\varepsilon = 10^{-10}$.

We start by discussing the dependence on $t$ of the truncation parameters of split regenerative randomization. Table 2 gives the values of the truncation parameters $K$, $L$, and $M$, $K_T = \sum_{k \in T} K_k$, and $L_T = \sum_{k \in T} L_k$ for the method for the $ETRR(t)$ measure; Table 3 gives the corresponding values for the method for the $EARR(t)$ measure. We can note that for both measures and in all cases the truncation

### Table 2
Truncation parameters as a function of $t$ for $ETRR(t)$

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th></th>
<th>Case 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td>$K_T$</td>
<td>$K$</td>
<td>$K_T$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>111</td>
<td>2</td>
<td>121</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>174</td>
<td>3</td>
<td>194</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>202</td>
<td>3</td>
<td>220</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>292</td>
<td>4</td>
<td>319</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>339</td>
<td>4</td>
<td>367</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>500</td>
<td>5</td>
<td>533</td>
</tr>
<tr>
<td>1,000</td>
<td>6</td>
<td>603</td>
<td>6</td>
<td>645</td>
</tr>
<tr>
<td>5,000</td>
<td>8</td>
<td>904</td>
<td>8</td>
<td>961</td>
</tr>
<tr>
<td>10,000</td>
<td>9</td>
<td>1,050</td>
<td>9</td>
<td>1,116</td>
</tr>
<tr>
<td>50,000</td>
<td>10</td>
<td>1,276</td>
<td>11</td>
<td>1,380</td>
</tr>
<tr>
<td>100,000</td>
<td>11</td>
<td>1,359</td>
<td>11</td>
<td>1,443</td>
</tr>
</tbody>
</table>

### Table 3
Truncation parameters as a function of $t$ for $EARR(t)$

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th></th>
<th>Case 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td>$K_T$</td>
<td>$K$</td>
<td>$K_T$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>102</td>
<td>2</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>155</td>
<td>3</td>
<td>173</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>182</td>
<td>3</td>
<td>201</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>263</td>
<td>4</td>
<td>287</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>308</td>
<td>4</td>
<td>335</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>456</td>
<td>5</td>
<td>489</td>
</tr>
<tr>
<td>1,000</td>
<td>6</td>
<td>547</td>
<td>6</td>
<td>586</td>
</tr>
<tr>
<td>5,000</td>
<td>7</td>
<td>805</td>
<td>8</td>
<td>884</td>
</tr>
<tr>
<td>10,000</td>
<td>8</td>
<td>949</td>
<td>9</td>
<td>1,032</td>
</tr>
<tr>
<td>50,000</td>
<td>10</td>
<td>1,214</td>
<td>10</td>
<td>1,290</td>
</tr>
<tr>
<td>100,000</td>
<td>10</td>
<td>1,276</td>
<td>11</td>
<td>1,380</td>
</tr>
</tbody>
</table>
parameters increase smoothly with $t$. Also, the truncation parameters $K$ and $L$ have very small values. This is because, having the system many components with quite similar failure rates, the output rates from states in $E$ are very similar and, therefore, $R_E$ is only slightly larger than 1 and $q_E$ is very small. The truncation parameters $M$, $K_k$, and $L_k$ have also reasonably small values. In all cases, the truncation parameters for the method for the $EARR(t)$ measure are non-greater than the truncation parameters in the method for the $ETRR(t)$ measure. This can be explained by recalling that the model truncation error bounds for the method for the $EARR(t)$ measure are non-greater than the respective model truncation error bounds for the method for the $ETRR(t)$ measure.

We compare next the performance of split regenerative randomization (SRR) with those of standard randomization (SR), regenerative randomization (RR), randomization with quasistationarity detection (RQD), and, for the $ETRR(t)$ measure, adaptive uniformization (AU). For AU we choose the AU layered uniformization variant for AU processes with converged rate described in Moorsel and Sanders (1994), since this ensures for AU the same numerical stability as all other three methods have. Figure 7 gives the CPU times for the $ETRR(t)$ measure; Fig. 8 gives the CPU times for the $EARR(t)$ measure. We start discussing the results
for Case 1. Although not clearly seen in Fig. 7, for $ETRR(t)$, AU is, with few exceptions, the fastest method for $t$ not larger than about 1,000 h. Compared with SR, there is a crossing point at about 5,000 h below which AU is faster and above which AU is slower. This fact is in accordance with the known behavior of AU with respect to SR (Moorsel and Sanders, 1994). RR performs not much worse than SR for both $ETRR(t)$ and $EARR(t)$. In addition, since the size of the truncated transformed model built in RR is logarithmic in $t$ and the number of steps required in SR grows linearly with $t$, for $t$ large enough RR will eventually become faster than SR. In the example, RR becomes faster than SR for $t$ larger than about 50,000 h for both $ETRR(t)$ and $EARR(t)$. For the considered values of $t$, RQD is the more expensive method, but it would outperform also SR for larger $t$’s. Finally, SRR is the fastest method for $t$ beyond approximately 1,000 h. For $t = 100,000$ h, SRR is, for the $ETRR(t)$ measure, about 18.2 times faster than the fastest of the other methods (RR) and, for the $EARR(t)$ measure, about 19.3 times faster than the fastest of the other methods (RR). In Case 2, there is almost no difference in performance between AU and SR for the $ETRR(t)$ measure. This is because, in that case, the adapted randomization rate used in AU is large from the initial steps. In that case RR compares worse with SR than it did in Case 1. The reason is that when the initial probability distribution is not concentrated in the regenerative state (the state without failed components), the truncated transformed model built in RR is larger than when that initial probability distribution is concentrated in the regenerative state (Carrasco, 2003). The performance of RQD is, however, very similar to the performance of that method in Case 1. As in Case 1, for $t$ large enough, SRR is the fastest method. However, the time beyond which SRR is the fastest method is now about 5,000 h for both measures, larger than in Case 1. The reason is that the truncated transformed model is larger than in Case 1 because of the presence of the comb having as back the states $s'_0, s'_1, \ldots, s'_L$ and the string of states $s''_0, s''_1, \ldots, s''_M$. The gain in performance of SRR over the other methods is significant albeit smaller than in Case 1. Thus, for $t = 100,000$ h, SRR is, for the $ETRR(t)$ measure, about 15.4 times faster than the fastest of the other methods (SR) and, for the $EARR(t)$ measure, also about 15.4 times faster than the fastest of the other methods (SR). For the example, $R \approx 8$. Were the repair rates more different, $R$ would be greater, $M, K$, and $L$ would be greater and split regenerative randomization would be relatively more costly.

5. Conclusions

We have generalized a method called split regenerative randomization which is specifically targeted at the transient analysis of rewarded CTMC models of fault-tolerant systems with deferred repair. The generalized method covers a slightly wider type of CTMC models and allows to compute two transient measures: the expected transient reward rate and the expected averaged reward rate. The method has the same good properties as the randomization method (numerical stability, well-controlled computation error, and ability to specify the computation error in advance) and can be significantly less costly than that method. The method requires the selection of a subset of states and a regenerative state and its performance depends on those selections. For a class of rewarded CTMC models, class $C^r$, including typical failure/repair models with exponential failure and repair time distributions and deferred repair, natural selections for the subset of states and
the regenerative state exist and, for those natural selections, theoretical results are available assessing approximately the computational cost of the method in terms of “visible” model characteristics. Using a large class \( C_2 \) model, we have shown that, for models in that class, the method can be significantly faster than other randomization-based methods.

References


