Unclassified

AD-A199 940

PORT DOCUMENTATION PAGE

(2)

1a	1b RESTRICTIVE MARKINGS
2a N/A	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 88-0975
6a. NAME OF PERFORMING ORGANIZATION Massachussetts Institute of Technology	6b. OFFICE SYMBOL (If applicable)
6c. ADDRESS (City, State, and ZIP Code) 77 Massachusetts Ave. - Room 33-109 Cambridge, MA 02139	7a NAME OF MONITORING ORGANIZATION AFOSR/NM 7b ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b OFFICE SYMBOL (If applicable) NM
8c. ADDRESS (City, State, and ZIP Code) Building 40 Bolling AFB, DC 20332-6448	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0160 10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
6.1102F	2304	A5	

11 TITLE (Include Security Classification)
Approximate Evaluation of Reliability and Availability Via Perturbation Analysis

12 PERSONAL AUTHOR(S)
Bruce K. Walker, Sin-Kwong Chu and Norman M. Wereley

13a. TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 6/84 TO 9/87	March, 1988	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Markov models, semi-Markov models, reliability, avoidability

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The progress on a three-year effort to examine approximate reliability evaluation techniques for fault tolerant control and sensor systems is described. The motivation for the work is provided by the fact that the reliability models for these systems tend to be finite state semi-Markov models with large dimensions that evolve relatively slowly in time due to the rare occurrence rate of component failures. The transient behavior of these models is of interest because the steady state behavior is trivial and not of practical importance. The evaluation of the transient behavior of such models, however, is intractable even for relatively simple system architectures because of the widely varying rates at which events occur in the model. The research effort concentrates on generating useful limit theorems that approximate the behavior of these models asymptotically well as the small component failure rates become vanishingly small. Using the work of Korolyuk as a starting point, such limit theorems are generated for both continuous and discrete time models that are representative of fault tolerant system behavior. In particular, the limit theorems of Korolyuk are expand to cover models where the classes of the decomposed models include trapping states when the

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian W. Woodruff	22b TELEPHONE (Include Area Code) (202) 767-5027
	22c OFFICE SYMBOL AFOSR/NM

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

88 10 5 140

BLOCK 19. CONTINUED

small parameter vanishes and to cover models where the holding times are not necessarily scale by the small parameter. Furthermore, the sufficient conditions for these theorems are expressed in terms of properties of the unperturbed version of the model that are relatively easy to check and do not involve generating steady state limits of transition operators.

The application of these limit theorems to some selected fault tolerant system models for which analytical results can be generated symbolically is also described. This motivates further work in an effort to expand the applicability of the imite theorem results to the broadest possible class of models that can result from the analysis of fault tolerant systems. Preliminary results from this effort are described.

Approximate Evaluation of Reliability and Availability Via Perturbation Analysis

Final Technical Report on

Grant AFOSR-84-0160

Prof. Bruce K. Walker

Dept. of Aerospace Engineering & Engineering Mechanics
University of Cincinnati
Cincinnati, OH 45221-0070

Siu-Kwong Chu

Norman M. Wereley

Department of Aeronautics & Astronautics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

March, 1988

Covering the Period: June 1, 1984 - September 30, 1987

Prepared for:
Maj. Brian W. Woodruff
AFOSR/NM
Building 410
Bolling AFB, DC 20332

TABLE OF CONTENTS

ABSTRACT

The progress on a three-year effort to examine approximate reliability evaluation techniques for fault tolerant control and sensor systems is described. The motivation for the work is provided by the fact that the reliability models for these systems tend to be finite state semi-Markov models with large dimension that evolve relatively slowly in time due to the rare occurrence rate of component failures. The transient behavior of these models is of interest because the steady state behavior is trivial and not of practical importance. The evaluation of the transient behavior of such models, however, is intractable even for relatively simple system architectures because of the widely varying rates at which events occur in the model.

The research effort concentrates on generating useful limit theorems that approximate the behavior of these models asymptotically well as the small component failure rates become vanishingly small. Using the work of Korolyuk as a starting point, such limit theorems are generated for both continuous and discrete time models that are representative of fault tolerant system behavior. In particular, the limit theorems of Korolyuk are expanded to cover models where the classes of the decomposed models include trapping states when the small parameter vanishes and to cover models where the holding times are not necessarily scaled by the small parameter. Furthermore, the sufficient conditions for these theorems are expressed in terms of properties of the unperturbed version of the model that are relatively easy to check and do not involve generating steady state limits of transition operators.

The application of these limit theorems to some selected fault tolerant system models for which analytical results can be generated symbolically is

also described. This motivates further work in an effort to expand the applicability of the limit theorem results to the broadest possible class of models that can result from the analysis of fault tolerant systems. Preliminary results from this effort are described.

I. INTRODUCTION

1.1 Motivation and Discussion of Problem

Reliability and availability have become two of the prime considerations in the design of control systems for a diverse group of applications that includes flight control systems for both aircraft and spacecraft. Considerable effort is now being devoted to the design of highly reliable control system components and to the design of fault-tolerant processors for online control computations. Despite the success of some of these efforts, the extremely high reliability goals that are becoming commonplace in the Air Force and elsewhere can often be met only by designing control systems with built-in component redundancy. The combination of a redundant system architecture and a redundancy management (RM) algorithm constitutes a fault-tolerant system design.

Predicting the performance of these designs is an important and difficult problem. The performance is judged by such quantities as the reliability, the availability, or some other probabilistic quantity such as the average measurement accuracy or average regulation error. Calculating these quantities is an important problem because they represent the criteria by which the system design is judged. Such calculations are difficult because fault-tolerant systems are subject to random events, such as failures and RM decisions, that change the nature of operation of the system and therefore affect the values of the performance quantities.

Several papers and theses have introduced the concept of modelling the random behavior of a fault-tolerant system by generalized finite-state Markov models [1-3, 17-18]. The states in these models characterize the status of the system in terms of the number of components that are operating, the number of these that are failed, and the status of the RM

decisions. The transition behavior among these states must then be derived from the probabilistic behavior of component failures and of the RM decisions (including errors such as false alarms and missed alarms). Once this characterization is complete, the resulting Markov model (or, more generally, semi-Markov model) can be used to derive the statistics of any relevant quantity that is dependent upon the status of the system. Among these are the reliability and availability of the system, but the statistics of other quantities such as the time to first passage of a particular system status or a performance measure dependent on the system state history can also be calculated.

Despite their obvious utility for fault-tolerant system performance analysis, these models suffer from one serious drawback that has considerably limited their use. That drawback is that they tend to be computationally intractable even for relatively simple fault-tolerant system architectures. This intractability is the result of a number of factors:

1. The number of states in the reliability model can be large, particularly for complex systems comprising many components. Essentially, there are as many states in the model as there are distinct combinations of failed and unfailed components and RM decision statuses for which the system remains operative. Even the exploitation of symmetry and similarities in component behavior to reduce the model order can still leave a very large number of states in the final model.
2. The transient behavior of the model, not the steady state behavior, is of primary interest. Because the components are subject to failure, the steady state condition for all fault-tolerant systems lacking online repair capability is complete system failure. Even when the recovery of failed or fail-indicated components is possible, a steady

state condition for the reliability model may not become established until the elapsed time is greater than the useful lifetime of the system (see comment 4 below). In either case, the transient behavior of the model is of interest and steady state analysis techniques do not apply. This is particularly unfortunate when the model is semi-Markov in nature because the transient analysis of such processes requires the evaluation of convolution quantities (integrals or sums, respectively, for continuous or discrete time models) that require massive amounts of computer memory and computation time.

3. The time horizons of interest are often very long in absolute terms, though they still remain short relative to the time required for the reliability model process to reach steady state. Typically, a fault-tolerant system will be used for operating intervals that are a significant fraction of the expected lifetime of its most failure-prone components. This fraction seldom approaches unity because the redundancy level of the failure-prone components required to satisfy reasonable specifications on the system reliability would drive the price of the system high enough to justify the use of fewer, more reliable (and therefore more expensive) components. On the other hand, extremely short operating times would yield a probability of failure for any component that is so low that the extra investment in fault-tolerance would not be justified by the small increase in reliability. In light of Item 2 above then, the transient behavior of a Markovian process must be examined over time horizons on the order of the mean time to failure of the most failure-prone component. Given the current emphasis on the manufacture of highly reliable components, these time horizons can be extremely long.

4. A time scale separation tends to exist between the component failure process and the RM decision process. Failures tend to occur only rarely and therefore tend to have large time durations between them. RM decisions, however, must occur quickly following a failure and tend to occur very rapidly relative to failure events. This means that the Markovian model of the behavior of the system status exhibits "fast" modes and "slow" modes. This time scale separation provides the motivation for the behavioral decomposition methods that have been investigated by us and by other researchers in the field.

The goal of this research project is to develop a method that generates approximate solutions to the generalized Markov process models that characterize fault-tolerant system behavior without the use of excessive computer memory or computation time. The behavioral decomposition alluded to in Comment 4 above provides the basis for the approach. However, the nature of fault-tolerant system models is such that extensions to existing theory have been necessary in order to exploit the decomposition approach. These extensions and the numerical verification of their validity are the primary results obtained from the work reported here.

1.2 Previous and Related Work

A number of researchers have addressed various aspects of the problem of approximating the behavior of finite state Markov processes with weak interactions between groups of states. In the particular context of fault-tolerant systems, a number of papers by Trivedi and coworkers have examined the use of Markov models for evaluating the reliability of fault-tolerant data processing systems [e.g., 1-3]. Techniques have been developed for decomposing these models along the lines of the behavioral decomposition

strategy discussed above. However, data processing systems have the unique property that all of the signals associated with the system are binary and therefore are rarely affected by noise. As a result, the fault detection system rarely indicates falsely that a fault is present when there is no fault. The model decomposition procedures examined in [1-3] implicitly rely on this fact because the model is always assumed to take the form of a low-order, slow process induced by the component failure process upon which is superimposed a fast process representing the "fault handling" by the RM logic following a fault. This structure is valid only if "fault handling" occurs only following failures. This rules out the possibility of false failure indications in the absence of failures.

Although many of the techniques developed in [1-3] are very powerful and easy to use, the limitation just discussed renders them inapplicable to fault-tolerant systems where false failure indications are likely. This includes essentially all fault-tolerant sensing and control systems because these systems are affected by noise and dynamic error sources that make false failure indications a major concern.

In the general area of Markov processes with weak interactions, one notable recent work is the article by Coderch [4]. This paper is derived from [5], which contains an extensive description of previous work in the area. Much of the work preceding [4] applied only to limited classes of finite state Markov processes and, in particular, was not applicable to semi-Markov processes or to processes with purely transient states. In [4], a method is described by which continuous time, finite state, weakly coupled Markov processes without transient states can be decomposed into transition operators that are valid for increasingly longer time scales. The result is a sequence of operators that describe the transition behavior of the process

at each time scale such that the multiple time scale solution for the process behavior converges to the actual process behavior asymptotically as the small parameter representing the weak interactions converges to zero. Unfortunately, the method does not apply to semi-Markov processes and it has not been extended to apply to discrete time processes. Furthermore, the method requires the solution of very complex linear algebra problems, such as the description of nullspaces of operators, in the generation of the operators that are valid at each time scale.

Another recent effort extended the results of [4] to finite state Markov processes evolving in both discrete and continuous time that include special types of transient states (called "nonsplitting transient states" in [7]). Some preliminary results of this effort are described in [6]. Further results are described in [7]. It should be noted that the results in [6], like those in the previously cited references, are applicable only to Markov processes. Some results on semi-Markov processes are included in [7], but the results are again limited to processes that contain only nonsplitting transient states. This rules out many of the models that represent fault tolerant systems because these models consist almost entirely of transient states (all but the trivial total system failure trapping states), many of which are not "nonsplitting."

It should also be noted that the methods of [6] and [7], like those in [4,5], generate a description of the behavior of the process in sequentially longer time scales. It is frequently the case in fault-tolerant system analysis that the behavior of interest occurs only in the first time scale. This observation, combined with the difficulty that the methods of [6] and [7] have in dealing with transient states and semi-Markov processes,

suggests that an alternative method for dealing with these processes is of interest.

Much of the work reported here is an extension of the work reported by Korolyuk, et. al. [8,9]. These results apply to finite state semi-Markov processes with weak interactions. The continuous time case is treated in [8] and the discrete time case in [9]. The key result of [8,9], which will hereafter be referred to as Korolyuk's limit theorem, is the following:

Theorem: Given a perturbed finite-state semi-Markov process $z^{\epsilon}(t)$ whose transition operator elements $P^{\epsilon}_{ij}(t)$ have the following dependence on ϵ:

$$P^{\epsilon}_{ij}(t) = (p_{ij} - \epsilon q_{ij})h_{ij}(t/e) \qquad \text{if} \quad i,j \epsilon E_k$$

$$= \epsilon q_{ij} h_{ij}(t/\epsilon) \qquad \text{if} \quad i \epsilon E_k, \; j \not\epsilon E_k$$

with $\sum_{j \epsilon E_k} p_{ij} = 1$ and where p_{ij} and q_{ij} are of order 1 and where the set of classes $\{E_k\}_{k=1}^{m}$ is disjoint and exhaustive. If the Markov chains defined by the p_{ij}'s within a single class E_k represent an ergodic Markov process with stationary state probability distribution $\{\pi_i^{(k)}\}$ for each k ($1 \leq k \leq m$), then:

$$\lim_{\epsilon \to 0} \text{Prob}\,\{\text{sojourn time from class } E_k \text{ to class } E_r \leq t\}$$

$$= \gamma_{kr} \cdot (1 - e^{-\lambda_k t})$$

where: $\gamma_{kr} = [\sum_{i \epsilon E_k} \pi_i^{(k)} \sum_{j \epsilon E_r} q_{ij}] \cdot [\sum_{i \epsilon E_k} \pi_i^{(k)} \sum_{j \epsilon E_k} q_{ij}]^{-1}$

$$\lambda_k = \left[\sum_{i \in E_k} \pi_i^{(k)} \sum_{j \in E_k} q_{ij}\right] \cdot \left[\sum_{i \in E_k} \pi_i^{(k)} \sum_{j \in E_k} p_{ij} \bar{\tau}_i\right]$$

$$\bar{\tau}_i = \sum_{j \in E_k} p_{ij} \bar{\tau}_{ij}$$

and where $\bar{\tau}_{ij}$ is the mean holding time for the holding time density $h_{ij}(t)$.

Proof: For the discrete time case, the proof appears in [9]. The corresponding proof for the continuous time case appears in [8].

The conditions and consequences of Korolyuk's limit theorem can be summarized in words as follows. The interactions between the states of the original semi-Markov process are weak in the sense that the transition behavior depends upon a small parameter ϵ such that when ϵ is zero the process decomposes into noncommunicating classes of states. The original process will be referred to hereafter as the perturbed process while the process that is derived from it by setting ϵ to zero will be called the unperturbed process. The form of the transition behavior assumed for the perturbed process is such that the transition probabilities of its imbedded Markov process within a class are independent of ϵ while the interclass imbedded Markov process transition probabilities are all at least first order in ϵ. Also, it is assumed that the holding time densities associated with all transitions for the perturbed process become compressed near the origin as ϵ becomes small. Finally, it is assumed that the decomposed classes of the unperturbed process are all ergodic. When all of these conditions are satisfied, the interclass behavior of the perturbed process over time horizons on the order of t/ϵ can be approximated by a reduced

order Markov process in this "slow" time scale. This process is called the enlarged Markov process. The approximate behavior of the original perturbed process can then be derived by expanding the enlarged process probabilities of occupying each class with the stationary distribution of probability within each class of the unperturbed process that results from the ergodicity of these classes. The parameters of the enlarged Markov process are expressed in terms of the decomposed transition probabilities of the perturbed process and the mean holding times associated with the holding time distributions.

This result is very powerful for approximating the behavior of semi-Markov processes that satisfy all of the conditions. Unfortunately, most models of fault-tolerant system behavior do not satisfy these conditions. This provides the motivation for most of the work on this project.

1.3 Research Goals

The research goals for the project can be summarized as follows:

1. Extend the results of [8,9], if possible, by carefully reviewing the proofs included there and recognizing points where the restrictive assumptions can be relaxed.
2. Extend the results of [8,9] to perturbed semi-Markov models evolving in continuous time where the holding time densities do not depend directly upon the small interaction parameter ϵ but rather on a small time scaling parameter.
3. Conduct investigations on several continuous time models, including some with nonergodic classes. Attempt to identify theoretical results regarding such models.

4. Develop results similar to [8,9] as extended by the two previous goals for discrete time semi-Markov models of fault tolerant systems.
5. Develop a means for generating the exact solution to both continuous and discrete time models of simple fault-tolerant systems for the purpose of comparison with the results generated by the approximate technique.

The next section of this report will discuss the progress made on these goals during the project.

II. PROGRESS SUMMARY

In this section, the work of the past three years is summarized and is related to the goals discussed above. The work will be summarized here in approximately the order in which it was accomplished. Numerous references are made to [10], which is the S.M. thesis of Siu-Kwong Chu that was completed under the support of this grant. This thesis was included as Appendix A of [12]. A draft version of a paper [13] that was derived from this thesis is included as Appendix A of this report. References are also made to [14] and [15], which are the S.M. thesis of Norman Wereley and a paper adapted from it, respectively. The paper [15] is included as Appendix B of this report. The thesis is available from the authors. It was sent to AFOSR previously under separate cover.

2.1 Attempts to Extend Korolyuk's Results

Fault-tolerant system models tend to have two characteristics that violate the conditions imposed on the semi-Markov processes examined in [8,9]. One is that the holding time densities do not compress as the small parameter representing the weak interclass interactions is made smaller. The reason for this is that the holding time densities for fault-tolerant system models are determined by the probability mass functions of the time needed for various sequential fault diagnosis tests to reach decisions. The behavior of the fault diagnosis tests typically occurs in the "fast" time scale, but it is not altered by changes in the failure rate of the components, which is usually the source of the small interaction parameter in these models. This situation is illustrated clearly by the model derived in Chapter 3 of [10], which is the 9-state model referred to in [11]. None

of the holding time densities for this model display the explicit dependence on the scaled time t/ϵ that [8,9] assume (see Appendix C of [10]).

The other manner in which fault-tolerant system models often violate the conditions assumed in [8,9] is with respect to the ergodicity of the classes when $\epsilon=0$. Many fault-tolerant systems include RM logic that shuts off a component permanently once it has been diagnosed as failed. If this diagnosis is the result of a false alarm, the corresponding system status state involves no failures and hence tends to be in the same class upon decomposition of the model as other no-failure states such as the state where no failures and no RM decisions have yet taken place. This means that these classes include "splitting" transient states, and the results of [7] are not applicable. Also, the false alarm state is a trapping state for this class when the failure probability (and hence ϵ) is set to zero. Therefore, this class of the unperturbed process is nonergodic. This tends to be true of many of the classes of states associated with models of fault-tolerant system behavior when irreversible RM logic is used by the system.

A major part of the early research on the project was devoted to extensive study of the results in [8,9] to determine whether they could be directly extended. In [11], it was noted that the ergodicity of the classes is actually a stronger condition than what is sufficient for the proofs presented in [8,9] to hold. In particular, the following Proposition was put forward:

> Proposition: The results of Korolyuk's limit theorem hold if the transition operator P_k of each class k of the unperturbed process is such that either of the following is true:

(1) P_k is the transition operator for an ergodic Markov process with unique stationary state probability distribution $\{\pi_i^{(k)}\}$, or

(2) The inverse operator $[I - P_k + \pi_k]^{-1}$ exists.

In these statements, P_k and π_k are defined as follows:

P_k = transition probability operator for the imbedded Markov process governing transitions within class k of the unperturbed process,

and

$$\pi_k = \lim_{n\to\infty} P_k^n \qquad \text{if this limit exists}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P_k^i \qquad \text{otherwise, if this exists.}$$

Proof: The proof of this Proposition is imbedded in the proof presented in [9] of Korolyuk's limit theorem. In [9], Korolyuk assumes the classes of the unperturbed process are ergodic, which guarantees the existence of the inverse operator. However, it is only the existence of the inverse operator that is required in the proof.

The conclusion that can be drawn from this Proposition is that the results of [8,9] are valid when the weaker sufficient condition represented by (2) in the Proposition is satisfied by each class in the model. This is a considerable generalization of Korolyuk's results. The unperturbed process derived from many fault-tolerant system models satisfy this weakened sufficient condition. In fact, a result will be stated and proven in a later section that generalizes Korolyuk's limit theorem to all perturbed

semi-Markov processes that have a corresponding unperturbed process with a particular property. This property is possessed by all perturbed semi-Markov models, and therefore all semi-Markov fault tolerant system models, that have unperturbed model classes that are aperiodic and contain no more than one trapping state. Such models also satisfy the weakened sufficient condition of this section.

Despite the considerable generality that the weakened sufficient condition implies, several problems still exist in applying the results to models of fault tolerant system behavior. The first problem is the nondependence of the holding time densities on the small parameter ϵ. This will be discussed in the next section.

The other problem is that models for fault tolerant system behavior are rarely specified in the standard form for semi-Markov models. The standard form for each of the transition operator elements for a semi-Markov process is the product of an imbedded Markov process transition probability and a holding time density. For most fault tolerant systems, the transition operator elements are derived directly and therefore do not take this product form, although the decomposition with respect to ϵ is clear. The imbedded Markov process transition probabilities can be determined in these cases only by integrating (or summing) the operator element histories with an infinite upper limit. The weakened sufficient condition requires only the calculation of the "fast" imbedded Markov process transition probabilities p_{ij}, so the calculation typically converges rather quickly. However, the subsequent calculation of π_k for each of the classes can be difficult, especially if the Cesaro limit form must be used. This latter calculation, if it is done numerically, also amplifies any error that may

have been present in P_k. Finally, to check the weakened sufficient condition, it is necessary to find the determinant of the operator $I-P_k+\pi_k$. This may be extremely difficult numerically.

Because some of the calculations leading to a check of the weakened sufficient condition are subject to error, we chose to continue our research to develop other conditions that could be checked under which the results of Korolyuk's limit theorem hold. These will be discussed later in this report.

2.2 Time-scaling of Continuous Time Models

Our research then turned toward circumventing the problem that the holding time densities for fault tolerant system models are not dependent on the small parameter ϵ representing the weak interactions, as is required by Korolyuk's limit theorem. The approach, which was first described in [12], is to introduce a second small parameter that represents time scaling into the model. In this section, we summarize the results of that work.

When the time axis over which a semi-Markov model of fault-tolerant system behavior evolves is scaled by a small parameter δ, the holding time densities in the model take the form that is required for the application of Korolyuk's limit theorem provided the parameter δ is proportional to ϵ. This idea is explained rigorously in section 2.2.1 of [10] and in Section 2.2 of [13]. After introducing this time scaling, it is possible to rederive the results that are of interest for asymptotic approximations to the behavior of these semi-Markov models.

Let E be the state space of a finite state semi-Markov process that evolves in continuous time t. Suppose that the process is observed with respect to the scaled time t/δ. Suppose further that the transition

operator of the process is such that its (i,j) element representing transitions from state i to state j has the form:

$$P_{ij}^{\epsilon}(t') = p_{ij}^{\epsilon} \; F_{ij}(t'/\delta) \qquad i,j \in E$$

where t' represents scaled time and where the imbedded Markov process probabilities p_{ij}^{ϵ} take the form:

$$p_{ij}^{\epsilon} = \begin{cases} p_{ij}^{(k)} - \epsilon \; q_{ij}^{(k)} & i,j \in E_k \\ \epsilon \; q_{ij}^{(k)} & i \in E_k, \; j \notin E_k \end{cases}$$

Here it is assumed that the state space E decomposes into weakly interacting classes $\{E_1, E_2, \ldots, E_n\}$. It is also assumed that the $p_{ij}^{(k)}$ for each E_K sum to unity, hence when $\epsilon=0$ the classes E_i become noninteracting and each describes a valid semi-Markov process.

Now let $\tau_{kr}^{(i)}$ be the sojourn time (in scaled time) of the process in class E_k when it begins from state $i \epsilon E_k$ and transits to class E_r with $r \neq k$. Let $\phi_{kr}^{(i)}(s)$ denote the characteristic function of $\tau_{kr}^{(i)}$. Then, <u>if the $p_{ij}^{(k)}$ for each k represent the transition probabilities of an ergodic Markov chain</u>, the $\phi_{kr}^{(i)}(s)$ are independent of the superscript i and take the form:

$$\phi_{kr}(s) = \frac{\sum_{i \epsilon E_k} \pi_i^{(k)} \sum_{j \epsilon E_r} q_{ij}^{(k)}}{\sum_{i \epsilon E_k} \pi_i^{(k)} \sum_{j \epsilon E_k} (\frac{\delta}{\epsilon} s \; a_{ij} \; p_{ij}^{(k)} + \epsilon \; q_{ij}^{(k)})}$$

$$= p_{kr} \frac{\Lambda_k/\alpha}{\Lambda_k/\alpha + s}$$

where the $\pi_i^{(k)}$ are the stationary probabilities of the ergodic semi-Markov process associated with class E_k, the a_{ij} are the mean holding times associated with the $F_{ij}(t)$ in the original time scale, α is δ/ϵ, and p_{kr} and Λ_k are parameters defined in [10, sec. 2.2.2] and in [13, sec. 2.2]. Note that this expression takes the form of the characteristic function of a Markov process transition operator with imbedded transition probability p_{kr} and transition rate time constant Λ_k/α. Thus, the interclass transitions are Markovian in scaled time.

This result is derived in [10, sec. 2.2.2] and in sec. 2.2 of [13].

The result expressed above makes possible the analysis of continuous time semi-Markov models of fault tolerant system behavior <u>provided the model has ergodic classes</u> (note the underlined condition above). Many fault tolerant system models violate this condition, as was discussed above. However, many fault-tolerant systems that do not employ irreversible fault isolation logic do produce models with ergodic classes. Therefore, this result is a positive step toward analysis of models for these types of systems. Furthermore, as the discussion of the preceding section has shown, models that satisfy the weakened sufficient condition can also be approximated by these results, including all models with unperturbed processes that are aperiodic and contain no more than one trapping state.

The manner in which the result above can be used for approximate analyses is as follows. Suppose a model for a fault tolerant system has been constructed and one is interested in calculating the state

probabilities for the model at some relatively large value of time t in order to assess the reliability (or some other status-related property) of the system. Suppose further that the model satisfies the conditions stated above. Then the approximate class occupancy probabilities at the desired time can be calculated by scaling time appropriately, constructing the enlarged Markov process that approximately governs interclass behavior from the result above and solving this relatively easy, reduced order Markov process problem. It is assumed here that the initial condition is known for the state probabilities and therefore also for the class occupancy probabilities. The results should then be rescaled back to the original time scale. Finally, the approximate state probabilities can be evaluated by weighting the stationary probability distribution associated with each class when $\epsilon=0$ by the appropriate approximate class occupancy probability.

To illustrate the approximate evaluation procedure, a model for a generic fault tolerant system was constructed and solved using both "brute force" numerical convolution techniques and the approximate technique described above. The system consisted of three components where at least one unfailed component must be available for the system to remain operating. It was assumed that the failure diagnosis algorithm used sequential tests in combination with logic that is described in detail in sec. 3.1 of [10]. The tests were assumed to have second order Erlang distributions for their times to decision. The logic included the possibility of recovering components that have previously been diagnosed as failed, thereby leading to a model that has ergodic unperturbed process classes. The complete model is described in secs. 3.3 through 3.5 and Appendix C of [10]. The model has 9 states which decompose into three classes when the small failure rate is set to zero.

The exact state probability histories are obtained numerically and are described in chapter 4 of [10]. It should be noted that a very large amount of computational effort was required to generate these exact solutions. The approximate model is also constructed and solved in chapter 4 of [10]. The approximate solutions were obtained from a relatively short FORTRAN program. They could have been generated using just a hand calculator. Only when complete time histories were desired was it absolutely necessary to resort to the use of a computer. Upon comparison of the results, one finds that the largest error in the evaluation of any of the state probabilities by the approximate method for this example is less than 1% of the value obtained by numerical means (which itself is subject to a small amount of error) for times greater than the longest mean holding time of the sequential tests, where the assumed mean time between failures is 3 orders of magnitude longer than this.

These results are very encouraging, but they are not sufficient to conclude that the approximate technique always works so well. In order to further investigate the properties of the approximate technique with the time scaling included, a number of four-state semi-Markov models were examined. These models were chosen to reflect various characteristics that larger fault tolerant system models tend to possess. By keeping the dimension at 4, however, it is possible to generate the true behavior of the model with relative ease whereas models of larger dimension are extremely difficult to solve exactly (recall the comments above regarding the nine-state model). Even four-state models are difficult enough to solve, however, that symbolic manipulation was necessary to generate the exact solutions. This is true despite the fact that none of the holding time

densities in the models were assumed to be any more complex than second order Erlang.

The five cases of four-state models that were examined are discussed in detail in chapter 5 of [10]. Two of these cases are included in [13]. The approximate method produced very accurate results in every case that was examined. The comparison between the results was almost always exact to 4 decimal places except in the very early time periods before the startup transient of the process has decayed.

One of the cases of four-state models that was examined was a model that did not have ergodic unperturbed process classes (Case IV of [10] or Case 2 of [13]). Because these examples are artificially constructed, it is relatively easy to check the weakened sufficient condition that was discussed in the previous section. A brief calculation shows that it is satisfied, therefore the results of Korolyuk's limit theorem hold for this model as well as the other cases that were examined. This makes the accuracy of the results obtained by the approximate method not surprising.

As was discussed at the end of the preceding section, however, we still desire a means for determining whether the results of Korolyuk's limit theorem hold without calculating the operator π_k for each class of the unperturbed model. The next section discusses our work along these lines.

2.3 Relaxation of Ergodicity Condition

Many fault tolerant systems yield generalized Markovian models of their behavior that decompose into classes that satisfy all of the conditions for applying the approximate technique except the condition that the classes of the unperturbed process must be ergodic. This is typically the result of

irreversible logic structures in the RM algorithm for the system such that diagnostic decisions alone can permanently eliminate a component from use.

However, in the analysis of four-state models discussed above, it was noted that excellent results were obtained when the approximate method was applied to a case where the unperturbed process generated by the model did not possess ergodic classes. As was noted above, this is not surprising because the inverse operator discussed in section 2.1 exists for each of the classes in this case. As we also noted above, however, we desire a simpler method for checking whether the results of Korolyuk's limit theorem are valid than computing the determinant of each of the operators $I-P_k+\pi_k$.

After careful examination of the underlying reasons that the results of Korolyuk's limit theorem hold for cases where the unperturbed process does not possess ergodic classes, we developed the following theorem:

Theorem 1: Let a semi-Markov process depend upon ϵ such that it can be decomposed and time scaled in the manner described in section 2.2. Suppose in addition that the imbedded Markov process transition operator P_k associated with the k^{th} class of the unperturbed process satisfies:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P_k^m = [e \; e \; \cdots \; e]^T$$

where e is some constant vector whose elements are nonnegative and sum to unity. Suppose this is true for each k (with different e in each case, in general). Then the interclass transition behavior of the perturbed process approaches the same enlarged Markov process behavior that was described in section 2.2 as ϵ approaches zero.

Proof: The proof of this theorem appears in chapter 6 of [10] and section 3 of [13].

The sufficient condition of Theorem 1 is weaker than the ergodicity of the classes that was required by Korolyuk. Furthermore, the sufficient conditions of Theorem 1 are more easily checked than the weakened sufficient condition that was discussed in section 2.1 because the determinant of the operator $I-P_k+\pi_k$ need not be computed. Note, however, that the Cesaro limit operator π_k must be computed to check the conditions of Theorem 1. This is still undesirable from a practical computational standpoint.

The analysis leading to Theorem 1 led us to consider the specific situations in which the conditions of the theorem are satisfied. This investigation led to the following refinement:

Theorem 2: Let a semi-Markov process depend on ϵ such that it can be decomposed into classes and time scaled as prescribed in section 2.2. The transition operator P_k of the imbedded Markov process associated with the k^{th} class of the unperturbed process will satisfy the conditions of Theorem 1 if:

1. The k^{th} class is ergodic, or
2. P_k has one and only one eigenvalue of unity.

Proof: The proof of this theorem also appears in chapter 6 of [10] and section 3 of [13].

Theorem 2 provides a more restrictive but more easily checked sufficient condition than Theorem 1 because the Cesaro limit is no longer necessary.

It should be emphasized that Theorems 1 and 2 still represent only sufficient conditions for the approximate technique to yield an accurate approximation to the behavior of the perturbed process as the small parameter ϵ becomes small. In other words, there may exist perturbed semi-Markov models that do not satisfy the conditions stated in Theorem 1 or Theorem 2 whose behavior can still be approximated well by the approximate method.

Some examples of model structures that do and do not satisfy the sufficient conditions of Theorem 1 or 2 are presented in chapter 6 of [10] and at the end of section 3 of [13]. One example in particular that does not satisfy the conditions includes a class in its unperturbed process model that contains multiple trapping states. We shall discuss this situation later in section 2.6.

2.4 Discrete Time Models

All of the results described so far in this report have applied to continuous time models of fault tolerant system behavior. However, the RM algorithms for fault tolerant systems are usually implemented on a digital computer with a significant time interval between successive applications of the diagnosis tests. Therefore, fault tolerant system models are often purely discrete time in nature.

During the course of the project, parallel efforts were made to derive results for perturbed discrete time semi-Markov processes that mimic those discussed above for continuous time processes. This section and the paper [15] reports on these efforts.

Much of the work that has been accomplished during the project for discrete time models has related to the adaptation of Korolyuk's limit

theorem for semi-Markov processes [8] to semi-Markov chains. In addition, a limit theorem with time scaling for semi-Markov chains was also developed. The theorem statements will be summarized below.

An important result that will be referred to in the statement of both theorems is the following:

LEMMA 3: Let $P^{(k)} = [p_{ij}^{(k)}]$ represent an imbedded Markov chain operator of a semi-Markov chain E_k. Consider the system of equations below:

$$\phi_{kr}^{(i)}(z) - \sum_{j \epsilon E_k} \phi_{kr}^{(j)}(z)\, p_{ij}^{(k)} = 0$$

The solution of this system of equations is independent of the superscript, that is:

$$\phi_{kr}(m) = \phi_{kr}^{(i)}(m) \quad \text{for all } i \epsilon E_k$$

if and only if the imbedded Markov chain transition operator $P^{(k)}$ has at most a single unit magnitude eigenvalue.

Proof: This lemma is proved in [14].

Thus, any ergodic imbedded Markov chain operator (for which all eigenvalues have less than unit magnitude) will satisfy Lemma 3. In addition, any monodesmic imbedded Markov chain operator (one that has only one trapping or absorbing state, and hence a single unit magnitude eigenvalue) will also satisfy Lemma 3. This assertion is similar to Theorem 2 of section 2.3 for continuous time models.

We shall now state a theorem that describes how a semi-Markov chain which is dependent on a small parameter ϵ can be approximately described by a Markov chain. This theorem is derived based on the results for semi-Markov processes in [8]. The semi-Markov chains here are assumed to depend on a small parameter ϵ such that the state space can be decomposed into

disjoint classes of states where the probabilities of departure from each class tend to zero along with ϵ. In addition, the total sojourn in each class is assumed to have a non-degenerate distribution in the limit as $\epsilon \to 0$.

THEOREM 4: A Limit Theorem for Semi-Markov Chains

Let the set E of states of the semi-Markov chain be expressible as a union of disjoint classes

$$E = \sum_{k=1}^{N^e} E_k \qquad k \in M = \{1,2,\ldots N^e\}$$

Let $\gamma_{kr}^{(i)}$ be the sojourn of the semi-Markov chain in class E_k when it starts from state $i \epsilon E_k$ and moves to class E_r. The following two sets of conditions are assumed to hold:

1. The elements of the core matrix sequence $\{g_{ij}^{\epsilon}(m) | i,j \epsilon E\}$ specifying the semi-Markov chain depend as follows on the small parameter ϵ:

$$g_{ij}^{\epsilon}(m) = p_{ij}^{\epsilon} \, h_{ij}(\tfrac{m}{\epsilon})$$

where $h_{ij}(0) = 0$. The p_{ij}^{ϵ} may be expanded in a Taylor series about $\epsilon = 0$. Retaining only linear terms in ϵ in explicit form:

$$p_{ij}^{\epsilon} = p_{ij}^{(k)} - \epsilon \, q_{ij}^{(k)} + \ldots + 0(\epsilon); \quad i,j \in E_k$$

$$= \epsilon \, q_{ij}^{(k)} + \ldots + 0(\epsilon); \quad i \in E_k \; ; \; j \notin E_k$$

where $0(\epsilon)$ represents terms such that $\lim_{\epsilon \to 0} \frac{0(\epsilon)}{\epsilon} = 0$.

The imbedded Markov chain obeys the usual Markov chain properties:

$$\sum_{j \epsilon E_k} p_{ij}^{(k)} = 1; \text{ and } p_{ij}^{(k)} \in [0,1];$$

for all $i,j \in E_k$ and for all $k \in M$

and

2. The imbedded Markov chain defined by the transition probability matrices $\{p_{ij}^{(k)} | i,j \in E_k$ for each $k \in M\}$ are ergodic with stationary distributions $\{\pi_i^{(k)} | i \in E_k, k \in M\}$.

Then:

$$\lim_{\epsilon \to 0} \Pr\{\gamma_{kr}^{(i)} \leq t\} = \gamma_{kr} [1 - \exp(-\Lambda_k t/T)]$$

where:

$$\gamma_{kr} = \frac{\sum_{i \in E_k} \pi_i^{(k)} q_i^{(kr)}}{\sum_{i \in E_k} \pi_i^{(k)} q_i^{(k)}} ,$$

$$\Lambda_k = \frac{\sum_{i \in E_k} \pi_i^{(k)} q_i^{(k)}}{\sum_{i \in E_k} \pi_i^{(k)} a_i^{(k)}} ,$$

Here:

$$q_i^{(rk)} = \sum_{j \in E_r} q_{ij}^{(k)} ,$$

$$q_i^{(k)} = \sum_{j \in E_k} q_{ij}^{(k)} ,$$

$$a_i^{(k)} = \sum_{j \in E_r} p_{ij}^{(k)} \bar{\gamma}_{ij} ,$$

$$\bar{\gamma}_{ij} = \sum_{m=0}^{\infty} m \, h_{ij}(m) .$$

Proof: The proof of this theorem appears in [14] and comprises most of section 2 of [15].

Although the above theorem is useful, it is not directly applicable to most fault tolerant system models for two reasons: (1) the imbedded Markov chains for such models are usually non-ergodic (as has been stated on numerous occasions in this report already), and (2) the holding time density functions are usually not dependent on m/ϵ but only on m. Hence, two necessary adjustments must be made in the above theorem. The first of these is to determine what conditions must be satisfied by the imbedded Markov chain in order for the results to be valid. This leads to Lemma 3, which has already been stated. The second of these adjustments is to incorporate time scaling into Theorem 4 in a manner analogous to the time-scaled limit theorem of section 2.2. The following theorem results from this consideration:

THEOREM 5: A Limit Theorem With Time Scaling for Semi-Markov Chains

Let the set E of states of a semi-Markov process be expressible as a union of disjoint classes

$$E = \sum_{k=1}^{N^e} E_k \qquad k \in M \ (= 1,2,\ldots N^e)$$

Let $\gamma_{kr}^{(i)}$ be the sojourn of the semi-Markov chain in class E_k when it starts from state $i \epsilon E_k$ and moves to class E_r. Let the following two sets of conditions hold for the semi-Markov chain:

1. The elements of the core matrix sequence $\{g_{ij}^{\epsilon}(m) | i,j \epsilon E\}$ specifying the semi-Markov chain depend as follows on the small parameter δ:

$$g_{ij}^{\epsilon}(m) = p_{ij}^{\epsilon} \, h_{ij}(\tfrac{m}{\delta})$$

where $h_{ij}(0) = 0$. The p_{ij}^{ϵ} may be expanded in a Taylor series about $\epsilon = 0$ as:

$$p_{ij}^{\epsilon} = p_{ij}^{(k)} - \epsilon\, q_{ij}^{(k)} + \ldots + 0(\epsilon); \quad i,j \in E_k$$

$$= \epsilon\, q_{ij}^{(k)} + \ldots + 0(\epsilon); \quad i \in E_k \; ; \quad j \notin E_k$$

The imbedded Markov chain obeys the usual Markov chain properties:

$$\sum_{j \in E_k} p_{ij}^{(k)} = 1; \text{ and } p_{ij}^{(k)} \in [0,1] \text{ for all } i,j \in E_k \text{ and all } k \in M$$

and

2. The imbedded Markov chains defined by the transition probability matrices $\{p_{ij}^{(k)} | i,j \in E_k \; , \; k \in M\}$ have at most a single unit magnitude eigenvalue (hence, ergodic or monodesmic) with stationary distribution $\{\pi_i^{(k)} | i \in E_k \; , \; k \in M\}$.

Then:

$$\lim_{\epsilon \to 0} \Pr\{\gamma_{kr} \leq t\} = \gamma_{kr}\,[1 - \exp(-\Lambda_k t/\alpha T)]$$

where γ_{kr}, Λ_k, $q_i^{(rk)}$, $q_i^{(k)}$, and $a_i^{(k)}$, were all defined in Theorem 4 and

$$\alpha = \frac{\delta}{\epsilon} \; .$$

Proof: This theorem is proved in [14]. The steps in the proof are nearly identical to those of the proof of Theorem 4.

Once these results had been derived, we turned our attention to demonstrating these results on some models for fault tolerant systems that were simple enough to yield analytical results but still illustrative of the behavior of fault tolerant system models discussed in the preceding sections. The next section summarizes these efforts.

2.5 Application of Discrete Time Results

The results of Theorem 5 were applied to three realistic examples of fault tolerant systems for which semi-Markov chain reliability models can be derived. Two rather simple fault tolerant systems were considered. The first of these systems is a single component monitoring system (SCMS). A single non-essential component is monitored by a sequential test to indicate any detected faults for the information of the pilot. This produces a 3-state semi-Markov reliability model that can be decomposed into two classes. The nature of these classes depends upon the assumptions that are made regarding the monitoring strategy. The second system that was considered is a single-component dual redundant (SCDR) system. This system consists of two identical components operating in parallel with a RM strategy to detect and identify failures and to select the appropriate component for use. Under a particular set of assumptions regarding the RM strategy, the model for the SCDR has eight states and three non-ergodic classes.

The results for these two systems will be summarized below.

2.5.1 SCMS with Continuous Monitoring (SCMS-I)

The results for the SCMS-I model are summarized in sections 5.1 and 5.2 of [15], which is also attached to this report as an Appendix. It should be noted that the reliability model for SCMS-I is simple enough to be solved analytically by z-transform methods. Therefore, an exact analytical answer is available with which to compare the results generated by the approximation implied by Theorem 5. Most reliability models do not have this property, however the SCMS-I system was chosen for examination in part because of this property.

The basic assumptions made in constructing the reliability model for SCMS-I were that a sequential monitoring test is continuously applied to the component and that this test has second order hypergeometrically distributed times to decision under both normal and failed conditions. The second order hypergeometric assumption makes possible the analytical solution discussed in the preceding paragraph. Furthermore, the second order hypergeometric distribution is a reasonable approximation to the decision time distribution that is observed in practice for sequential monitoring tests. It should also be noted that in the construction of the SCMS-I model, it is assumed that a value is known for the probability of an eventual false alarm decision when no failure is present. In practice, this probability would probably not be known and the elements of the core matrix would have to be constructed directly from the probability of a false alarm decision as a function of the elapsed time, which could be found by simulating the monitoring test with no failure present.

Section 5.1 of [15] shows how the model is decomposed and how the approximate results are calculated. The section concludes with the approximate expression for the state occupancy probabilities that are valid for large values of t:

$$\hat{\pi}^e = [\ (1-P_{fa})\ e^{-\epsilon\Lambda_1 t/T},\ P_{fa}\ e^{-\epsilon\Lambda_1 t/T},\ 1 - e^{-\epsilon\Lambda_1 t/T}\ ]$$

where ϵ is the (small) failure probability during each time step, T is the time step, and Λ_1 is determined in terms of the parameters of the decision time distributions and other model quantities and represents the interclass transition rate.

As section 5.2 of [15] shows, the results produced by the approximation are very accurate after an initial transient period that is a small fraction of the mean time between failures of the components (which is essentially proportional to $1/\epsilon$). For the numerical values of the parameters that were selected, the error in each of the elements of $\hat{\pi}^e$ is less than 5% after 1/500 of the MTBF has elapsed. Section 5.2 of [15] also shows that values of ϵ up to approximately 0.001 produce good agreement between the approximation and the exact answer. Finally, it is shown in Section 5.2 of [15] that the class probability predicted by the enlarged Markov process is a first order approximation in ϵ to the exact answer for this example and that the approximation also includes the dominant second order term, although other second order terms are not accounted for.

The excellent results for SCMS-I are not surprising because the reliability model for this system obeys the conditions for Korolyuk's original limit theorem with one exception, namely the condition requiring that the holding time distributions compress as the small parameter ϵ vanishes. The classes of the decomposed model are ergodic in this case, so only time scaling need be used to generate the excellent results that were obtained.

2.5.2 SCMS with Abbreviated Monitoring (SCMS-II)

If the sequential test that is used for monitoring in the SCMS is terminated upon its first indication of a failure, then the resulting reliability model is slightly different from the SCMS-I model. In particular, one of the two classes of the decomposed unperturbed model is nonergodic. This model is described in section 5.3 of [15]. The same assumptions regarding the decision time distributions and the existence of a

value for the eventual false alarm probability are made as in the SCMS-I model. After executing the steps of decomposing the model and calculating the parameters of the approximating Markov process, the result for $\hat{\pi}^e$ is:

$$\hat{\pi}^e = [\ 0,\ \ e^{-t/T},\ \ 1 - e^{-t/T}]$$

The SCMS-II model can also be solved analytically using the same z-transform procedure that was used for SCMS-I.

As section 5.4 of [15] points out, the errors in the approximation to π_2 and π_3 are extremely small (<0.01%) for all values of t. However, the error in the approximation to π_1 is always 100% because the approximation is exactly zero. This results from the stationary distribution that yields the value of zero for $\pi_1^{(1)}$.

Analytically comparing the class probability results generated by the approximation and the analytical solution shows that the SCMS-II class probabilities are accurate through the dominant second order term in ϵ. This is the same result that was observed for the SCMS-I model.

2.5.3 SCDR System

The SCDR system is the simplest redundant system that can be constructed. It was chosen for analysis in the hope that exact results could be obtained for it by the same method that was applied to the SCMS systems discussed above. The SCDR system was assumed to have an independent sequential test monitoring each of the components where the tests are both reset once either of them has reached a no-failure decision (which is commonly done in practice to minimize the number of false alarms). Also, it

was assumed that the system can survive for short periods of time (in this case, one RM time interval) while using a failed component, but not for extended periods of time. The RM strategy is to use one of the components (designated the primary component) until its test indicates that it is failed, then switch to the other (backup) component unless it is also indicated as failed, in which case use of the primary component continues despite the failure indication. Both monitoring tests are discontinued once a failure indication by either has occurred.

The details of the analysis of the SCDR system are presented in [14]. In this report, the results will be briefly summarized.

If the usual techniques are used for constructing the semi-Markov reliability model for the SCDR system, the result is a six-state system referred to as the SCDR-ASL model in [14]. This model does not decompose in the manner required for application of the limit theorems discussed in the preceding section. The reason for this is that the standard reliability model construction technique designates a single aggregated state as the system loss state. The SCDR system with the RM strategy outlined above has several routes by which system loss can occur following the first failure. Some of these routes (like repeated missed detections of the failed component) occur in the "fast" time scale associated with the RM decisions. At least one (the occurrence of a second failure following correct reconfiguration) occurs in the "slow" time scale and therefore contributes a transition probability that is proportional to the small parameter ϵ. The result is that the aggregated system loss state communicates by both fast and slow transitions with states in a single class. This violates the conditions necessary to decompose the model properly.

This difficulty is rather easily overcome. By decomposing the aggregated system loss state into three states with each reflecting a particular route to system loss, an eight-state model referred to in [14] as the SCDR-TS model is produced. This model decomposes into three classes in the manner prescribed by the conditions of the limit theorems. However, one of these classes includes two distinct trapping states. This violates the sufficient conditions for both of the discrete time limit theorems stated in the preceding section. This means that the results of these theorems do not necessarily apply to this model.

Recall that a weakened sufficient condition for the application of Korolyuk's limit theorem was discussed in section 2.1. Let us check this condition for this model. Referring to the description of the SCDR model in [14], we find that the imbedded Markov process transition probability matrix for Class 2 is:

$$P_2 = \begin{bmatrix} P_{fa}(1-P_m) & 0 & (1-P_{fa})P_m & (1-P_{fa})(1-P_m) & P_{fa}P_m \\ 0 & (1-P_{fa})P_m + P_{fa}(1-P_m) & 0 & (1-P_{fa})(1-P_m) & P_{fa}P_m \\ 0 & 0 & 1-P_{fa}\ P_m & 0 & P_{fa}P_m \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where P_{fa} is the eventual probability of a false failure indication by one of the tests, P_m is the probability of an eventual missed detection by one of the tests, and the two tests are assumed to be identical. In practice, of course, P_{fa} and P_m would not necessarily be known explicitly but rather would be implicit in the statistical description of the decision behavior of the tests. For this analysis, however, let us assume that these values are known.

Despite the presence of several zero values in P_2, it is quite difficult to obtain an analytical solution for the steady state operator $\lim_{k\to\infty} P_2^k$ in terms of the variables P_{fa} and P_m. Doing so requires z-transform analysis that includes factoring a fifth order polynomial symbolically. However, when numerical values are substituted for P_{fa} and P_m, it is relatively easy to find the limiting transition operator and thence to check for the existence of the inverse of the operator $I-P_2+\pi_2$. In particular, when $P_{fa}=0.05$ and $P_m=0.1$, we find that the determinant of $I-P_2+\pi_2$ is 0.0049. Therefore, the relaxed sufficient condition is satisfied and Korolyuk's limit theorem results apply. However, for these values (and for several other sets of reasonable values that were tried), the condition number of $I-P_2+\pi_2$ is quite large (nearly 500 for the values cited above, much larger for smaller values of the eventual transition probabilities). This implies that numerical errors are likely in evaluating some of the quantities that are used to describe the enlarged Markov process that approximates the interclass behavior of the perturbed process.

Another problem with this model is that a unique stationary distribution for Class 2 of the unperturbed process does not exist in general. For instance, for the values cited above for P_{fa} and P_m, it is known only that the stationary distribution of the unperturbed process in Class 2 is a linear combination of the distributions [0 0 0 1 0] and [0 0 0 0 1] where the weights in the linear combination depend upon the initial condition for this unperturbed process. This means that the enlarged process cannot be expanded in terms of a stationary distribution to approximate the distribution of the original unperturbed process.

The fact that the SCDR-TS model could not be analyzed using the approximate technique motivated us to pursue further the problem of models with multiple trapping states in the classes of the unperturbed model. The results of that effort are reported in the next section.

Returning to the analysis of the SCDR system, it is possible to construct a reliability model for this system that satisfies the sufficient conditions for Theorem 5 of the preceding section. This is accomplished in section 4.4 of [14] and produces the model referred to there as SCDR-I. Unfortunately, the construction of SCDR-I involves the merging of two of the states in the SCDR-TS model where one of these states represents a working system and the other represents a failed system. Therefore, although the SCDR-I model possesses the mathematical properties that are necessary for the approximate reliability evaluation technique to be applied to it, the results are of questionable value because they cannot be used to generate the system reliability. Nevertheless, the analysis of the SCDR-I model was carried out in an effort to expand our insight on the use of the approximate technique.

Under the assumption that the times to decision for each of the monitoring tests of the SCDR system are distributed according to a second order hypergeometric distribution, the results of approximate analysis of the SCDR-I model are described in [14]. The numerical results indicate excellent approximation to the behavior of the perturbed model by the enlarged process even for relatively large values of ϵ (on the order of .05). Furthermore, it is shown in [14] that the approximate technique applied to the SCDR-I model yields values for the class occupancy probabilities that agree to first order with the exact answer and also include the dominant second order term in the exact answer. This is

significant because it is the same as the results obtained for the SCMS models discussed above despite major differences between the SCMS models and SCDR-I.

The SCDR-I model is sufficiently complicated that the powerful symbolic manipulation language MACSYMA could not generate analytical solutions for the individual state occupancy probabilities, even when the variables in the system were replaced by numerical values. Only the class occupancy probabilities could be found analytically. Thus, all of the numerical results discussed in [14] for the SCDR-I model are limited to class occupancy results. The difficulty encountered in generating an exact solution to the reliability model for a system as simple as the SCDR system with very simple holding time assumptions (second order hypergeometric) puts added focus on two critical aspects of this research. First, the need for approximate techniques is clearly apparent when even the simple SCDR system cannot be analyzed by exact techniques. Second, the fact that exact solutions for simple systems cannot be derived motivates research on techniques, perhaps computationally intensive techniques, for generating the baseline solutions with which approximate results will be compared to determine the validity of the approximations.

2.6 Extension to Models with Multiple Trapping States in the Classes

The SCDR-TS model and some of the continuous time models that do not satisfy the conditions of Theorem 2 in section 2.3 illustrate a type of model for which the results discussed to this point are not applicable. These models include multiple trapping states in some of the classes of the decomposed unperturbed model. When this occurs, the transition operator P_k for some of the classes can have multiple eigenvalues of unity and the

Cesaro limit defined in the sufficient conditions of the theorems does not in general have the necessary form.

In the fault tolerant system context, these types of models occur when the results of RM decisions without additional failures can yield system configurations with widely different performance interpretations. For instance, in the case of the SCDR system, a detected failure results in a working system while a false alarm combined with failure of the other component results in a failed system. Both of the states representing these situations are in the same class of the model and, because the RM tests are terminated upon any failure indication, are trapping states for this class when the failure rate is set to zero.

In this section, we will present an approach that extends the limit theorem results of the previous sections to this case. The results are preliminary because this work had just begun shortly before the termination of the grant. We limit the discussion here to continuous time models. Discrete time models remain to be investigated.

Suppose that a continuous time perturbed semi-Markov model is obtained for the behavior of a fault tolerant system such that it decomposes into classes and can be time-scaled as in section 2.2. Suppose also that at least one of the classes, say E_r, yields an unperturbed imbedded Markov process transition operator P_r that corresponds to a process that has transient states, labelled s_1, s_2, and so on, and multiple trapping sets, where a trapping set can consist either of a single state or a closed communicating class of recurrent states. The procedure for applying the results of the time-scaled limit theorem to this case is as follows.

Let the trapping sets in E_r be labelled R_1, R_2 , and so on and let S be the set of transient states in E_r. For each state $s_i \in S$, calculate the probability that a transition is eventually made to trapping set R_j for each j. This can be accomplished using only the information provided by P_r as follows.

Define an initial condition on the unperturbed process whereby the probability of occupying state s_i is unity. The unperturbed imbedded Markov process transition operator P_r can then be used to calculate the probability $p_{s_i r_j}$ that each of the trapping sets is the one eventually occupied when the process starts in state s_i by successively operating on this initial condition until the probability that any of the states in S is occupied vanishes and then adding together the probabilities of occupying the states within that trapping set when steady state is reached. Since the r^{th} class contains only a subset of the states of the overall model, the dimension of P_r can be small enough that the problem of finding the steady state trapping set probabilities can be solved by transform methods.

Once these probabilities are determined, we decompose class r into as many subdivisions as there are trapping sets. Each of these subdivisions becomes a new class in the modified decomposed model. Let these classes be labelled E_{r_1}, E_{r_2}, and so forth. As before, let $\tau_{kr_j}^{(i)}$ be the sojourn time from state $i \in E_k$ to class E_{r_j} with $E_{r_j} \in E_r \neq E_k$ and let $\phi_{kr_j}^{(i)}(s)$ be its characteristic function. Suppose that P_k has the Cesaro limit property that was the condition for Theorem 1 of section 2.3 (and recall that Theorem 2

showed that ergodicity of E_k or the existence of one and only one eigenvalue of unity for P_k were sufficient for this property). We then have that, in scaled (i.e. slow) time, as ϵ approaches zero:

$$\phi_{kr_j}(s) = \phi^d_{kr_j}(s) + \sum_i p_{s_i r_j} \phi_{ks_i}(s)$$

$$- p_{kr_j} \frac{\Lambda_k}{\Lambda_k + s} + \sum_i p_{s_i r_j} p_{ks_i} \frac{\Lambda_k}{\Lambda_k + s}$$

where the summation is over all i for which $s_i \in S$, $\phi^d_{kr_j}(s)$ is the direct transition limiting sojourn characteristic function, and $\phi_{ks_i}(s)$ is the limiting sojourn characteristic function from class k to state $s_i \in S$. The parameters p_{kr_j}, p_{ks_i}, and Λ_k in the second line are determined in exactly the same fashion as the parameters p_{kr} and Λ_k in the continuous extension of Korolyuk's limit theorem stated in section 2.3 above. Note that this result is independent of the starting state in class k, i.e. the result is independent of the superscript.

This result implies that an enlarged Markov process can be defined that approximates the transition behavior of the perturbed process among the trapping sets of the classes in scaled time. As before, the approximate state occupancy probabilities can then be found by expanding the class probabilities using the steady state distribution of the unperturbed process within each class, if the steady state distribution exists.

The proof of this result is still incomplete. However, it will be outlined here.

There are essentially three steps to the proof. The first is to show that the characteristic equation of the sojourn from any state in E_k to any state in S has the form $p_{ks_i} \frac{\Lambda_k}{\Lambda_k + s}$. Next is to show that the sojourn from any state in E_k directly (i.e. without occupying any of the states in S in the process) to any state in E_{r_j} also has this form with s_i replaced by r_j. Finally, we must show that the linear combination of these terms in the final result produces the characteristic function of the sojourn from any state in E_k to any state in E_{r_j}. The proof outline will describe how this can be done when S comprises only one state. The extension to multiple transient states is straightforward, but had not been completed at this writing.

Since S and E_{r_j} are subsets of E_r, the first two steps in the proof are already complete by applying the results of Theorem 1 in section 2.3. This leaves only the third step. With $\tau_{kr_j}^{(i)}$ defined as above and with $\eta_{kr_j}^{(i)}$ representing the analogous sojourn for transitions directly from states in E_k to states in E_{r_j}, we have that when S is the only transient state in E_{r_j}:

$$\Pr(\tau^{(i)}_{kr_j} \le t) = \sum_{n \epsilon E_k} \Pr(i \to n,\ \delta\zeta_{in} + \eta^{(n)}_{kr_j} \le t)$$

$$+ \sum_{n \epsilon E_{r_j}} \Pr(i \to n,\ \delta\zeta_{in} \le t)$$

$$+ \sum_{n \epsilon E_k} \sum_{l \epsilon E_{r_j}} \Pr(i \to n,\ \delta\zeta_{in} + \tau^{(n)}_{kS} + \delta\beta_{Sl} \le t)$$

$$+ \sum_{l \epsilon E_{r_j}} \Pr(i \to S,\ \delta\zeta_{iS} + \delta\beta_{Sl} \le t)$$

where $i \to n$ represents that the next transition is from i to n, ζ_{in} is the holding time for the transition $i \to n$ within class E_k measured in the "fast" time scale (so that $\delta\zeta_{in}$ is this holding time measured in scaled time where the small parameter δ represents the time scaling from "fast" time to the scaled or "slow" time), and β_{nl} is the holding time for transitions from state n to state l within class E_{r_j} expressed in "fast" time units (hence multiplication by δ scales this to the "slow" time scale). Proceeding as in [8], this expression can be written in terms of the transition operator for the perturbed process as:

$$\Pr\{\tau_{kr_j}^{(i)} \le t\} = \sum_{n\epsilon E_k} \int_0^t \Pr\{\eta_{kr_j}^{(i)} \le t-u\}\, dP_{in}^{\epsilon}(u)$$

$$+ \sum_{n\epsilon E_{r_j}} P_{in}^{\epsilon}(t)$$

$$+ \sum_{n\epsilon E_k} \sum_{l\epsilon E_{r_j}} \int_0^t \left[\int_0^{t-u} \Pr\{\tau_{kS}^{(i)} \le t-u-\tau\}\, dP_{in}^{\epsilon}(\tau) \right] dP_{Sl}^{\epsilon}(u)$$

$$+ \sum_{l\epsilon E_{r_j}} \int_0^t \Pr\{\delta\beta_{Sl} \le t-u\}\, dP_{iS}^{\epsilon}(u)$$

Taking the Laplace transform of the entire expression and using the properties assumed for the perturbed process, we have:

$$\phi^{(i)}_{kr_j}(s) = \sum_{n\epsilon E_k} \phi^{(n)d}_{kr_j}(s)\ [p^{(k)}_{in} - \epsilon\, q^{(k)}_{in}]\ [\,1 - \delta\, a_{in}\, s + o(\epsilon)]$$

$$+ \epsilon \sum_{n\epsilon E_{r_j}} q^{(k)}_{in}$$

$$+ \sum_{n\epsilon E_k} \sum_{l\epsilon E_{r_j}} \phi^{(n)}_{kS}(s)\ [\,p^{(k)}_{in} - \epsilon\, q^{(k)}_{in}]\ [\,1 - \delta\, a_{in}\, s + o(\delta)]$$

$$[\,p^{(k)}_{Sl} - \epsilon\, q^{(k)}_{Sl}]\ [\,1 - \delta\, a_{Sl}\, s + o(\delta)]$$

$$+ \epsilon\, q^{(k)}_{iS}\, p_{Sr_j}$$

where:

$$p_{Sr_j} = \sum_{l\epsilon E_{r_j}} p^{(r)}_{Sl}$$

By moving all of the terms of order one or higher in ϵ and δ to the right hand side of the expression for the characteristic function and moving all of the zero order terms to the left, upon taking the limit as ϵ and δ go to zero we get:

$$\phi^{(i)}_{kr_j}(s) - \sum_{n\epsilon E_k} p^{(k)}_{in}\ [\,\phi^{(n)d}_{kr_j}(s) + p_{Sr_j}\, \phi^{(n)}_{kS}(s)\,] = 0$$

The assumption that class E_k has a unique Cesaro limit sense solution for the steady state distribution of the unperturbed process within class E_k implies that the characteristic functions $\phi^{(i)}_{kr_j}(s)$, $\phi^{(i)d}_{kr_j}(s)$, and $\phi^{(i)}_{kS}(s)$ are all independent of i. This yields:

$$\phi_{kr_j}(s) - \sum_{n \epsilon E_k} p^{(k)}_{in} \left[\phi^{d}_{kr_j}(s) + p_{Sr_j} \phi_{kS}(s) \right] = 0$$

for which the solution is obviously:

$$\phi_{kr_j}(s) = \phi^{d}_{kr_j}(s) + p_{Sr_j} \phi_{kS}(s)$$

This essentially completes the proof for the case of a single transient state S in class E_r. We are hoping to extend this proof technique to the case of multiple transient states in E_r.

III. SUMMARY OF SIGNIFICANT FINDINGS AND FUTURE WORK

3.1 Summary of Significant Findings

The limit theorems derived as the result of the research described above extend the enlarged Markov process approximation to all of the model structures that have been observed by the authors in constructing semi-Markov models for fault tolerant control system behavior. This is of great practical significance because of the tremendous reduction in computational overhead that is realized by using the approximation relative to solving the model exactly.

The key results of the research are the limit theorems cited above. These include the extension of Korolyuk's limit theorem to time-scaled models that do not decompose into ergodic classes (Theorems 1 and 2 of section 2.3) and the discrete time version of these theorems, Theorem 5 of section 2.4. The final key result is the one stated in section 2.6 regarding continuous time models that decompose into classes that have multiple trapping sets. Because fault tolerant system behavior often involves situations where RM decisions alone (without failures) can lead to more than one outcome, multiple trapping sets within a class are a common occurrence in models of this behavior. Therefore, this latter result is crucial to extending the approximation to essentially all of the fault tolerant system models that one might encounter for evaluation.

Another key finding results from the error analysis described in section 2.5. It was found that for the two models considered (the single component monitoring system and the single component dual redundant system)

the class probabilities predicted by the approximation agree with the exact solution to first order in the small parameter ϵ and also through the dominant second order term. This implies that the error in this approximation is not just second order but in some sense "small" second order for these two cases. We hasten to add that there is no indication that this result is true in general, but it is rather interesting that it is true for two widely different simple cases.

In summary, the research conducted under this grant has laid the mathematical groundwork for the practical computation of approximate reliability predictions for a wide range of fault tolerant systems with random dynamics that can be modelled by semi-Markov processes. This is significant because such models are too complex to be solved exactly even with today's extremely powerful computers. The approximate method applies to all discrete or continuous time models of fault tolerant systems that the authors have encountered with one exception: discrete time models with multiple trapping sets in the decomposed classes. It is believed that this single exception can be eliminated (see below).

The availability of this practical reliability evaluation tool can have profound impact on the practice of designing fault tolerant systems that are subject to random RM decision errors as well as random component failures. Informed design decisions can now be made based upon quantitative system performance results that were too difficult to compute before this method was developed. The ease with which results can be computed also makes possible iterative design schemes that use a performance measure calculated by this technique as the design criterion.

3.2 Proposed Future Work

A proposal has been submitted to AFOSR to continue the funding of this research. The proposed research includes two tasks that follow directly from the research reported here. One of these is to extend the limit theorems to discrete time models with multiple trapping sets in the classes in order to eliminate the major exception cited above. This should be a straightforward task. The other is to apply the approximate technique to more sophisticated models than those that have been examined so far. As part of this task, an effort will be undertaken to generate exact results for these complex cases with which to compare the approximate results in order to verify the accuracy of the approximation. Some innovative computational methods on extremely high throughput machines may be required to do this. In addition, another task involves the examination of approximating quantities other than the system reliability and state occupancy probabilities with the approximate method.

IV. PERSONNEL

This research was conducted at the Massachusetts Institute of Technology and the University of Cincinnati under the direction of Bruce K. Walker, Associate Professor. He was assisted at the Massachusetts Institute of Technology by Siu-Kwong Chu and Norman M. Wereley, Graduate Research Assistants. For the last eight months of the grant, the project director at the Massachusetts Institute of Technology was Wallace E. Vander Velde, Professor.

V. PAPERS AND PRESENTATIONS

The following papers and presentations resulted either partly or wholly due to the research reported here:

1. B.K. Walker, "Approximate Methods of Performance Evaluation for Fault-Tolerant Systems," presented at 23rd IEEE Conf. on Decision & Control, Las Vegas, December 1984.

2. B.K. Walker, "Decomposition of Generalized Markovina Models of Fault Tolerant Systems - Motivation, Progress, and Problems," presented at AFOSR Workshop on Reliability, Shenandoah National Park, May 1985.

3. B.K. Walker & D.K. Gerber, "Evaluation of Fault Tolerant System Performance by Approximate Techniques," Proc. of 7th IFAC Symp. on Identification & System Parameter Estimation, York, UK, July 1985.

4. B.K. Walker, N.M. Wereley, R.H. Luppold, & E. Gai, "Effects of Redundancy Management on Reliability Modelling," accepted for publication in IEEE Trans. on Reliability, 1988.

5. S.-K. Chu & B.K. Walker, "Approximate Behavior of Generalized Markovian Fault Tolerant System Models," under revision for submittal to Reliability Engineering.

6. N.M. Wereley & B.K. Walker, "Approximate Evaluation of Perturbed Generalized Markov Chain Models of Fault Tolerant Systems," accepted for publication in Reliability Engineering, 1988.

7. N.M. Wereley & B.K. Walker, "Approximate Evaluation of Perturbed Generalized Markov Chain Models of Fault Tolerant Systems," submitted to 27th IEEE Conf. on Decision & Control, Austin, December 1988.

REFERENCES

[1] K.S. Trivedi, "Reliability Evaluation for Fault Tolerant Systems," in G. Iazeolla, P.J. Courtois and A. Hordijk (eds.), Mathematical Computer Performance and Reliability, North-Holland, Amsterdam, 1984.

[2] A. Bobbio and K.S. Trivedi, "An Aggregation Technique for the Transient Analysis of Stiff Markov Chains," IEEE Trans. on Computers, C-35: 9: 803-814, September 1986.

[3] J.B. Dugan and K.S. Trivedi, "The Hybrid Automated Reliability Predictor," J. of Guidance, Control and Dynamics, 9: 3: 319-331, May-June 1986.

[4] M. Coderch, A.S. Willsky, S.S. Sastry, and D.A. Castanon, "Hierarchical Aggregation of Singularly Perturbed Finite State Markov Processes," Stochastics, 8:259-289, 1986.

[5] M. Coderch, "Multiple Time Scale Approach to Hierarchical Aggregation of Linear Systems and Finite State Markov Processes," Ph.D. Thesis, Dept. of EECS, M.I.T., Cambridge, MA, 1982.

[6] X.-C. Lou, J.R. Rohlicek, P.G. Coxson, G.C. Verghese, and A.S. Willsky, "Time Scale Decomposition: The Role of Scaling in Linear Systems and Transient States in Finite State Markov Processes," Proc. of 1985 American Control Conf., Boston, June 1985.

[7] J.R. Rohlicek, "Aggregation and Time Scale Analysis of Perturbed Markov Systems," Ph.D. Thesis, Dept. of Elect. Eng. & Computer Sci., M.I.T., January, 1987.

[8] V.S. Korolyuk, L.I. Polishchuk and A.A. Tomusyak, "A Limit Theorem for Semi-Markov Processes," Kybernetika, 5:4:144-145, July-August 1969.

[9] V.S. Korolyuk and A.F. Turbin, "Asymptotic Enlarging of Semi-Markov Processes with an Arbitrary State Space," in A. Dold and B. Eckmann (eds.), Lecture Notes in Mathematics 550: Proc. of 3rd Japan - USSR Symp. on Probability Theory, Springer-Verlag, 1972.

[10] S.-K. Chu, "Approximate Behavior of Generalized Markovian Models of Fault-Tolerant Systems," S.M. Thesis, Dept. of Aero. & Astro., M.I.T., Cambridge, MA, February 1986.

[11] B.K. Walker, S.-K. Chu and N.M. Wereley, "Annual Progress Report on Grant AFOSR-84-0160: Approximate Evaluation of Reliability and Availability Via Perturbation Analysis," Dept. of Aero. & Astro., M.I.T., September 1985.

[12] B.K. Walker, S.-K. Chu and N.M. Wereley, "Annual Progress Report on Grant AFOSR-84-0160: Approximate Evaluation of Reliability and Availability Via Perturbation Analysis," Dept. of Aero. & Astro., M.I.T., December 1986.

[13] S.-K. Chu and B.K. Walker, "Approximate Technique for the Transient Analysis of Generalized Markovian Fault Tolerant System Models", submitted to Reliability Engineering, 1987.

[14] N.M. Wereley, "An Approximate Method for Evaluating Generalized Markov Chain Reliability Models of Fault Tolerant Systems," S.M. thesis, Dept. of Aero. & Astro., M.I.T., Cambridge, MA, January 1987.

[15] N.M. Wereley and B.K. Walker, "Approximate Evaluation of Semi-Markov Chain Reliability Models," submitted to 1988 IEEE Conference on Decision & Control and to Reliability Engineering, 1988.

[16] E.E. Lewis and F. Bohm, "Monte Carlo Simulation of Markov Unreliability Models," Nuclear Engineering and Design, 77:49-62, 1984.

[17] B.K. Walker, "A Semi-Markov Approach to Quantifying Fault Tolerant System Performance," Sc.D. Thesis, Dept. of Aero. & Astro., Massachusetts Institute of Technology, July 1980.

[18] B.K. Walker, "Performance Evaluation of Systems that Include Fault Diagnostics," Proc. of 1981 Joint Automatic Control Conf., Charlottesville, VA, June 1981.

Appendix A

Title : Approximate Technique for the Transient Analysis of Generalized Markovian Fault-Tolerant System Models.

Authors : Siu-Kwong Chu, Doctoral Candidate, Department of Aeronautics and Astronautics, M.I.T., Cambridge, MA.

Bruce K. Walker, Associate Professor, Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Cincinnati, OH.

Submitted to : Reliability Engineering.

Date : September, 1987.

Abstract

Problems associated with evaluating state probability vector of large state space models of fault-tolerant systems are explained. Korolyuk's Limit Theorem on semi-Markov processes leads to a solution to these problems by approximating an aggregated version of the original semi-Markov process by a reduced order Markov chain. The Theorem is modified and extended to apply to fault-tolerant system models in a slow time scale. The approximate technique is then completed by expanding the approximate Markov chain states probabilities with the limiting probability vector that apply to each decoupled aggregate class of states of the original semi-Markov process. The approximate technique is demonstrated on a couple of 4-state models that mimic

the class-to-class transition structure of typical fault-tolerant system models. The results show that accurate approximation is achieved for these examples after a short transient period. In addition, the ergodicity sufficient condition imposed on the classes of the original, decoupled semi-Markov process by Korolyuk's theorem is relaxed. As a result, fault-tolerant system models with certain types of non-ergodic classes can also be treated by the approximate technique.

Keywords:

Semi-Markov Process

Fault-Tolerant System

Enlarged Process

Reliability Evaluation

Approximate Solution

Transient Analysis

1 INTRODUCTION

1.1 Background

A fault-tolerant system is a system designed with redundant capacity to perform its function. That is, it can do its job using more than one configuration of its components, e.g. sensors, actuators and information processing components. The on-line detection and isolation of failed components and the subsequent reconfiguration of the system's operating architecture is performed by the system's Redundancy Management (RM) scheme.

The fault-tolerant design approach enhances system reliability and performance. There are many application areas where ultra-high system reliability is necessary or desirable. One such area is the control of nuclear power plants where the consequences of improper control system behavior may be serious indeed. There are space missions for which the desired operational lifetime of the spacecraft is many years during which time many component failures are probable. The air traffic control system and many military systems are also subject to very high reliability requirements. There is also a desire for increased reliability in computerized banking systems, chemical process control systems, medical monitoring systems, transportation systems, and many more.

Growing attention is being given to the design of components for long life, to quality control during manufacture, and testing and maintenance policies which enhance reliable system operation. Despite these efforts to improve the reliability of individual components, the resulting system reliability is still often inadequate for some reliability requirements. As a result, there

is increasing interest in fault-tolerant system designs which allow components to fail but still provide a means for the system to continue to function.

The growing use of fault-tolerant system designs has in turn spurred interest in methods for assessing the reliability and performance of such systems. The traditional methods of reliability evaluation are based on combinatorial analysis of combinations of component failures[1] amd these analyses seldom account for the probabilistic nature of the outcomes of the on-line monitoring tests that are used to detect and identify failures and to reconfigure the system . In addition, classical reliability analysis produces as its sole result the probability that the system will maintain its integrity over the duration of its operating time. No information is provided on the performance of the system during the transient period of the mission.

Since classical reliability analysis fails to quantify fault-tolerant system time behavior, other alternatives must be considered. Naturally, Monte Carlo simulation is one option. However, for the systems we are interested in, complex and with low component failure rate, a huge number of simulations is required to generate statistically significant results, and it is often prohibitively costly.

The use of Markov chain theory[2,3] has shown promise as a means for evaluating the performance of fault-tolerant systems that employ Fault Detection and Isolation (FDI) tests of the single sample variety, i.e., the information that is used for FDI is gathered and discarded at each time sample. Methods have been proposed to deal with the problems associate with large Markvo chain

models, aggregation/disaggregation technique by Takahashi[4], decomposition technique proposed by Courtois[5] and aggregation technique by Bobbio[6]. In addition to the problem of large state space, these efforts also consider the issue of stiffness, i.e. the simultaneous presence of transition rate of different orders of magnitude in the model. The former two techniques are for solving the steady state probability vector and only the latter is for transient analysis of large stiff Markvo chains. However, single sample FDI tests generally have a relatively high likelihood of decision errors, particularly in noisy signal environments. In such situations, the FDI tests are usually based on several samples of the monitoring data at each time sample, e.g. moving window tests and sequential tests[7,8,9]. Such tests are not memoryless. Therefore, Markov chain analysis does not apply to systems employing these types of tests.

Some effort[8,10] has been made to analyze such systems and it appears that generalized Markovian (or semi-Markov[11,12]) modeling methods are applicable to some systems of this type. Semi-Markov processes are very similar to Markov chains, but have an extra degree of freedom that makes them well-suited for capturing the random delay behavior of RM decisions for nonmemoryless FDI tests. However, a problem with this reliability evaluation method is that the large number of states in the model causes the computation of results to involve excessive amounts of computer storage and computation time. The reason for this is that standard time-invariant semi-Markov theory requires the solution of a matrix convolution integral equation to find the interval transition probability matrix.

For complex systems with a large number of states in their model (for example, model for a dual-redundant engine controller[13] examined has 30 states and flight control system models will have many more), it becomes intractable to obtain a solution either analytically or numerically. More specifically, consider a discrete time semi-Markov model with state probability vector $\underline{\pi}(k)$, at time step k. If $\underline{\pi}(0)$ is known, then $\underline{\pi}(k)$ can be expressed as,

$$\underline{\pi}(k) = \underline{\pi}(0)\ \underline{\Phi}(k) \tag{1}$$

where $\underline{\Phi}(k)$ is recursively generated[12] by,

$$\underline{\Phi}(k) = {}^{>}W(k) + \sum_{m=0}^{k} [P \square H(m)]\ \underline{\Phi}(k-m)\ , \qquad \underline{\Phi}(0)=I \tag{2}$$

It can be seen that a convolution sum is involved. This implies that for a system with N states, approximately $2kN^2$ values must be stored in order to compute $\underline{\Phi}(k)$ and hence $\underline{\pi}(k)$. For N = 20 and k = 100,000, as might be the case for a simple flight control system operating with RM updates at a rate of 50Hz for 35 minutes, the storage required is approximately 80×10^6 values or 640 megabytes of storage for accurate single precision state probability distribution calculations. The number of floating point multiplications required for calculating $\underline{\Phi}(100{,}000)$ is approximately 7×10^{12}. The same problem arises for continuous time models. Several authors[14,15,16,17] have developed algorithms for transient analysis of special semi-Markov process. They allow the aggregates of fast states to constitute a semi-Markov or a more general stochastic process and Trivedi et al. extend Stiffler's approach to allow for slow transition out of fast states. However, these attempts in the area of reliability analysis were tailored to particular system structures and provide no analytical approximate solution in term of parameters of kernel elements of the large and stiff semi-Markov process.

This paper discuss an aggregation approximation technique explicitly intended to aggregate states so that the resulting system is small state space and the transitions between aggergate states approximated by Markov chain. The technique proposed in this paper can be applied to stiff semi-Markov models of fault-tolerant systems. Generally speaking, the states of the model each represent a different system operating configuration in terms of number of working components, components in use and failure monitoring status and states are aggregated according to the number of working components but with different RM configuration. The algorithm proceeds by first establishing the Markvoian behavior between aggregates of states of a semi-Markov process. The resulting Markov chain is then analyzed by a standard analytical or numerical technique. Approximate solution is completed by expanding the total probability in each aggregate of states by the stationary probability vector in that aggregate of states.

The rest of the paper is organized as follows. Section 2 derives how the aggregates of states of a semi-Markov process can be approximated by a Markov chain. Section 3 relaxes the sufficient condition imposed on the aggregates of states by the Theorem. The approximate technique is then demonstrated with two numerical examples in Section 4.

2 THEORY OF APPROXIMATE AGGREGATE TECHNIQUE

2.1 Introduction

Assuming stiff semi-Markov models of fault-tolerant systems of interest have transition kernel elements of the following form (generally they are, see

9-state model of a three components redundant fault-tolerant system in the literature[18]) :

$$P_{ij}(t) = p_{ij}^{\epsilon} F_{ij}(t) \qquad i,j \in E; \tag{3}$$

$$p_{ij}^{\epsilon} = \begin{cases} p_{ij}^{(k)} - \epsilon q_j^{(k)} & i,j \in E_k \\ \epsilon q_{ij}^{(k)} & i \in E_k,\ j \notin E_k \end{cases} \tag{4}$$

$$\text{where} \sum_{j \in E_k} p_{ij}^{(k)} = 1 \qquad i \in E_k, \quad 1 \le k \le m.$$

where ϵ is small parameter associates with failure rate of components, and we can aggregate the model into classes of states E_i, $1 \le i \le m$, that is there is no transition across each aggregate of states when ϵ tends to zero. It was shown[19] that state i probability can be approximated as:

$$\pi_i(t) \simeq \pi_i^{(k)} \pi_k^{e}(t) \tag{5}$$

where $\pi_k^{(k)}$ is the stationary probability of state i in class E_k and $\pi_k^{e}(t)$ is the probability of kth aggregate of states. So we have an approximate solution if we can find the probability of aggregates of states, since stationary probability in each class can be evaluated by standard semi-Markov theory.

2.2 Korolyuk's Limit Theorem for Semi-Markov Processes

Literatures[20,21] describe sufficient conditions under which a perturbed semi-Markov process can be approximated by a Markov chain. There are essentially two conditions. First, the kernel of the semi-Markov process must depend on a small positive parameter ϵ in such a way that the state space of the semi-Markov process E can be split into disjoint classes of states

$E=\sum_{k=1}^{m} E_k$, such that the probabilities of departure from each class and of the sojourn time in a given state both tend to zero with ϵ. The total sojourn time in each class is assumed to have a nondegenerate distribution in the limit as $\epsilon \to 0$. When $\epsilon=0$, the resulting process will be referred to as the *non-perturbed* semi-Markov process while the original process will be referred to as the *perturbed* semi-Markov process. This condition can be expressed by the following equations,

$$P_{ij}^{\epsilon}(t) = p_{ij}^{\epsilon} F_{ij}(t/\epsilon) \qquad i,j \in E; \tag{6}$$

$$p_{ij}^{\epsilon} = \begin{cases} p_{ij}^{(k)} - \epsilon q_j^{(k)} & i,j \in E_k \\ \epsilon q_{ij}^{(k)} & i \in E_k, \quad j \notin E_k \end{cases} \tag{7}$$

$$\text{where} \sum_{j \in E_k} p_{ij}^{(k)} = 1 \qquad i \in E_k, \quad 1 \leq k \leq m.$$

where p_{ij} is the eventual transition probability of the original process from state i to state j and $F_{ij}(t/\epsilon)$ is the Cumulative Distribution Function (CDF) of the holding time for transitions from state i to state j.

Second, the Markov chains defined by the transition probabilities $p_{ij}^{(k)}$ $(i,j \in E_k, \ 1 \leq k \leq m)$, must be ergodic with stationary probabilities $\pi_{M_i}^{(k)}$ $(i \in E_k, \ 1 \leq k \leq m)$. When these conditions are satisfied by a perturbed semi-Markov process, then its behavior can be approximated by a Markov chain. More specifically, if $\tau_{kr}^{(i)}$ is the sojourn of the semi-Markov process in class E_k when it begins from state $i \in E_k$ and moves to class E_r, then the Theorem shows that the cumulative distribution function of $\tau_{kr}^{(i)}$ approaches an exponential function as ϵ becomes vanishingly small:

$$\lim_{\epsilon \to 0} \Pr\{ \tau_{kr}^{(i)} < t \} = p_{kr} (1 - e^{-\Lambda_k t}) \tag{8}$$

As can be seen from the above equation, the dependence on i disappears on the right hand side of the equation. That is, each state in class E_k has the same asymptotic exponential holding time density function for transitions to class E_r. Therefore, all the states in class E_k can be merged together and the aggregated model has the characteristics of a Markov chain.

In a sophisticated fault-tolerant system there is Built-In Test Equipment (BITE) included in the RM system in order to recover a working component that is incorrectly detected as failed and isolated, so a component that was previously isolated as failed by the RM can be brought back on line. For this kind of system, the imbedded Markov chain within each class is usually ergodic. Then the ergodic condition is satisfied. Moreover, by comparing the fault-tolerant system model decribed by Eq.(3)-(4) and the semi-Markov model decribed by Eq.(6), our system model of interest satify all the conditions imposed by the Theorem except the condition defined by Eq.6. Usually, this condition is <u>not</u> satisfied by a fault-tolerant system model. The reason for this is as follows: If ϵ is small, i.e. the *Mean Time To Failure* (MTTF) of the components is large, say hundreds of hours, then the holding time of the transition, particularly those within a class, is determined only by the noise in the signals and the threshold set by the FDI test designer. So, as the failure rate tends to zero, the RM decision delay will <u>not</u> be affected by the failure rate. Therefore, the transition kernel of a fault-tolerant system model will not take on the form implied by Eq.(6)

2.3 Derivation of $\phi_{kr}(s)$ of a Time-Scaled Perturbed Semi-Markov Process

However, instead of viewing the semi-Markov process holding time from state i

to state j depend on the parameter ϵ, we could view the holding times are being in a different time scale compared to the aggregated model class to class transition holding times. The constant relating the two time scale is the small parameter ϵ. So Eq.(6) will represent the semi-Markov process in a different time scale compared with the Markov chain of the aggregates of state and the time scale could be relaxed to any small parameter δ,

$$P^{\epsilon}_{ij}(t) = p^{\epsilon}_{ij} F_{ij}(t/\delta) \tag{9}$$

With the properties of the semi-Markov process already stated in last section, we could proceed with the derivation of parameters for the Markov chain of the aggregated model.

Let $\delta\zeta_{ij}$ denote the sojourn of the semi-Markov process in state i, with the CDF $F_{ij}(t)$, while δ^{ϵ}_{ij} be binary indicators of transition from state i to the state j, so that $E\{\delta^{\epsilon}_{ij}\}=p^{\epsilon}_{ij}$. Then, the random quantities $\tau^{(i)}_{kr}$ can be obtained by using the expression for total probability :

$$P\{\ \tau^{(i)}_{kr} \le t\ \} = \sum_{j\in E_k} P\{\ \delta^{\epsilon}_{ij}=1,\ \ \delta\zeta_{ij} + \tau^{(j)}_{kr} \le t\ \} + \sum_{j\in E_r} P\{\ \delta^{\epsilon}_{ij}=1,\ \ \delta\zeta_{ij} \le t\ \} \tag{10}$$

Hence

$$P\{\ \tau^{(i)}_{kr} \le t\ \} = \sum_{j\in E_k} \int_0^t P\{\ \tau^{(j)}_{kr} \le t-u\ \}\, dP^{\epsilon}_{ij}(u) + \sum_{j\in E_r} P^{\epsilon}_{ij}(t) \tag{11}$$

Using the Laplace transform,

$$\phi^{(i)}_{kr}(s) = E\ \{\ e^{-s\tau^{(i)}_{kr}}\ \} \tag{12}$$

$$p^{\epsilon}_{ij}(s) = \int_0^{\infty} e^{-st}\, dP^{\epsilon}_{ij}(t) \tag{13}$$

then, Eq.(10) becomes,

$$\phi^{(i)}_{kr}(s) = \sum_{j\in E_k} \phi^{(j)}_{kr}(s)\ p^{\epsilon}_{ij}(s) + \sum_{j\in E_r} p^{\epsilon}_{ij}(s) \tag{14}$$

Combining the Laplace transform of Eq.(7) and Eq.(9) :

$$p_{ij}^{\epsilon}(s) = \begin{cases} \{p_{ij}^{(k)} - \epsilon q_{ij}^{(k)}\}\ \{1 - \delta s a_{ij} + 0(\delta)\} & i,j \in E_k \\ \delta q_{ij}^{(k)} + 0(\delta) & i \in E_k,\ j \notin E_k \end{cases} \tag{15}$$

Substituting these expressions in Eq.(14), it becomes,

$$\phi_{kr}^{(i)}(s) - \sum_{j \in E_k} p_{ij}^{(k)}\ \phi_{kr}^{(j)}(s) = -\sum_{j \in E_k} (\ \delta s a_{ij} p_{ij}^{(k)} + \epsilon q_{ij}^{(k)})\ \phi_{kr}^{(j)}(s) + \epsilon \sum_{j \in E_r} q_{ij}^{(k)} + 0(\epsilon\ \delta) \tag{16}$$

Passing to the limit as ϵ and $\delta \to 0$, the functions $\phi_{kr}^{(i)}(s)$ are found to satisfy the system of equations,

$$\phi_{kr}^{(i)}(s) - \sum_{j \in E_k} p_{ij}^{(k)}\ \phi_{kr}^{(j)}(s) = 0 \tag{17}$$

It follows from this and the assumption that the imbedded Markov chain defined by the transition probabilities $p_{ij}^{(k)}$ ($i, j \in E_k$) is ergodic, that the solution[22] of system Eq.(17) is independent of the superscript, i.e. for all $i \in E_k$, $\phi_{kr}^{(i)}(s)=\phi_{kr}(s)$. Multiplying Eq.(16) by the stationary probabilities $\pi_{M_i}^{(k)}$ and summing over all $i \in E_k$, then cancelling ϵ, the following is obtained,

$$\sum_{i \in E_k} \pi_{M_i}^{(k)} \sum_{j \in E_k} (\frac{\delta}{\epsilon}\ s\ a_{ij}\ p_{ij}^{(k)} + q_{ij}^{(k)})\ \phi_{kr}^{(j)}(s) = \sum_{i \in E_k} \pi_{M_i}^{(k)} \sum_{j \in E_r} q_{ij}^{(k)} \tag{18}$$

On passing to the limit as $\epsilon \to 0$, noting that all the $\phi_{kr}^{(j)}(s)$ have the limit function $\phi_{kr}(s)$, we obtain:

$$\phi_{kr}(s) = \frac{\sum_{i \in E_k} \pi_{M_i}^{(k)} \sum_{j \in E_r} q_{ij}^{(k)}}{\sum_{i \in E_k} \pi_{M_i}^{(k)} \sum_{j \in E_k} (\frac{\delta}{\epsilon}\ s\ a_{ij}\ p_{ij}^{(k)} + q_{ij}^{(k)})} \tag{19}$$

Table 1: State definitions and class decompositions for SCMS-I,II

State	State Definition	Class
1	Component is working	1
2	Component has a false alarm	1
3	System loss - component failed	2

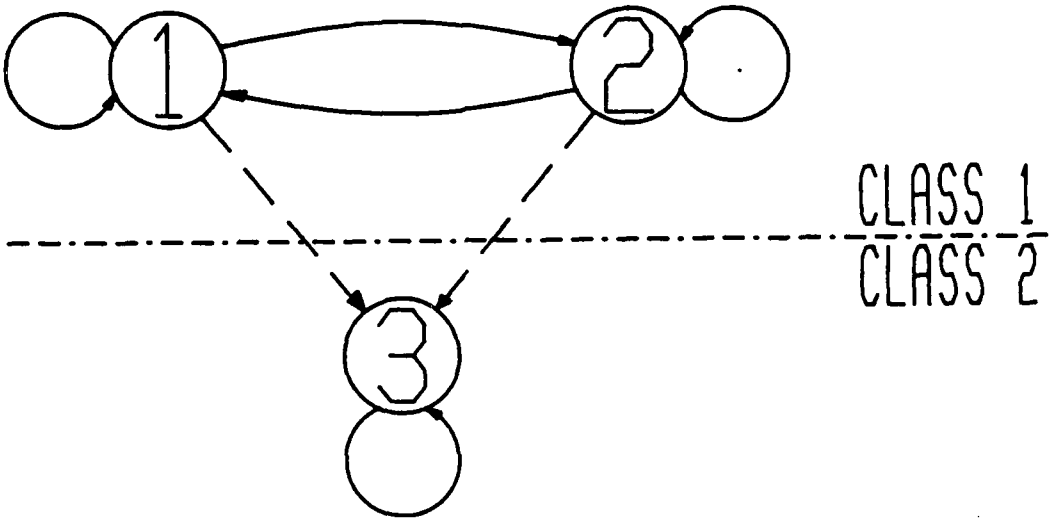


Figure 1: Semi-Markov transition diagram for SCMS-I

5.1 SCMS with continuous monitoring

Table 1 enumerates and defines the states of a semi-Markov chain reliability model of the SCMS-I. The dashed line in the table distinguishes the class decomposition of the model: class 1 contains states 1 and 2, class 2 contains only state 3.

The semi-Markov transition diagram for the SCMS-I is presented in Figure 1. Two aspects of this diagram should be noted. Given that the chain has entered a state, the lines directed out of that state represent transitions after the chain has remained in that state for a period of time, namely, the holding time. Secondly, the dashed lines represent transitions whose transition PMFs are proportional to ϵ. Thus, a dashed line represents the condition that no such transition occurs when $\epsilon = 0$. This is a convenient way of depicting the class decomposition of a semi-Markov chain reliability model.

aggregated model can be approximated by a Markov chain. The parameters of this process are given by Eq.(21)-(27).

Note that the approximate Markov chain evolves in a longer time scale, i.e. $1/\delta$ times that of the original process. For instance, if $\delta=1/3600$ and if the original semi-Markov model evolves in seconds then the approximate enlarged Markov process will evolve in hours.

One of the sufficient conditions in the derivation in this section for the enlarged process is that all the disjoint classes must be ergodic when $\epsilon=0$. This condition is usually not satisfied by all fault-tolerant system models. However, relaxation of this condition is discussed in the next section

3 RELAXATION OF ERGODICITY CONDITION

The second sufficient condition stated in the last section for the approximate Markov chain to be non-trivial is that the imbedded Markov chain of the non-perturbed process within each class must be ergodic. However, this condition can be relaxed and Korolyuk's Theorem can be modified as follows.

Theorem 1. If a semi-Markov process depends on a small parameter ϵ such that its state space can be partitioned according to Eq (7) and is time-scaled according to Eq.(6) and additionally if the transition probability operators P_k for the imbedded Markov chain of the k-th class of the non-perturbed semi-Markov process satisfy:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l = [\, \underline{e}\ \underline{e} \cdots \underline{e}\,]^T \qquad (28)$$

where $\underline{e}$ is column vector with sum of its elements equals to 1. Then the aggregated semi-Markov process can be approximated by the Markov chain defined by Eq.(20).

Proof. The proof follows an identical line of reasoning to the proof in section 2 until the point where the functions $\phi_{kr}^{(i)}(s)$ are shown to satisfy the system Eq.(17). The system equation can be rewritten in linear equation vector form:

$$\underline{\Phi}_{kr}(s) = P_k \, \underline{\Phi}_{kr}(s)^T \tag{29}$$

Premultiplying the above equation by the operator P_k and using Eq.(29) on the result gives:

$$\underline{\Phi}_{kr}(s) = P_k^2 \, \underline{\Phi}_{kr}(s)^T \tag{30}$$

By successively premultiplying the system of equations and replacing the left hand side by $\underline{\Phi}_{kr}(s)$, and averaging an infinite number of these equations

$$\underline{\Phi}_{kr}(s) = \left[\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l \right] \underline{\Phi}_{kr}(s) \tag{31}$$

Since the operator P_k defined by $p_{ij}^{(k)}$ satisfies Eq.(28), then, by linear equation theory, the solution of the system of equations in Eq.(31) is a vector with all its elements being the same, that is for all $i \in E_k$, $\phi_{kr}^{(i)}(s) = \phi_{kr}(s)$.

The remainder of the proof that the aggregated model is Markovian and the derivation of parameters of the approximate Markov chain will be exactly the same as that of the remainder of the proof in section 2.

This extended Theorem is a relaxation of the ergodicity sufficient condition stated earlier in Chapter 2 imposed on the semi-Markov process to be approximated.

It is of interest to find conditions under which Eq.(3.1) is satisfied. Along these lines, the following theorem is established:

Theorem 2 If the imbedded Markov chain which is defined by the transition operator P_k of the k-th class of the non-perturbed semi-Markov process is (1) ergodic, or (2) is non-ergodic with one and only one eigenvalue of unity, then the operator P_k satisfies Eq.(27).

Proof. (1) By ergodic theorem,

$$\lim_{l \to \infty} P_k^l = \Pi_k = [\, \underline{e}\ \underline{e} \cdots \underline{e}\,]^T \tag{32}$$

and,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l = \lim_{n \to \infty} \left\{ \frac{1}{n} \sum_{l=1}^{r} P_k^l + \frac{1}{n-r} \sum_{l=r+1}^{n} P_k^l \right\} \tag{33}$$

where r is finite but large such that,

$$P_k^r \approx \Pi_k \tag{34}$$

Therefore, Eq.(33) can be reduced to :

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l = \lim_{n \to \infty} \frac{1}{n-r} \sum_{l=r+1}^{n} P_k^l \tag{35}$$

By Eq.(34) it follows:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l = \Pi_k = [\, \underline{e}\ \underline{e} \cdots \underline{e}\,]^T \tag{36}$$

which proves the Theorem for this case.

(2) The operator P_k can be put in Jordan form by the following transformation:

$$P_k = T \Lambda_k T^{-1} \tag{37}$$

where T is a square invertible matrix with columns made up of the right eigenvectors (or generalized right eigenvectors) of the operator P_k. By a proper ordering, Λ_k has the form:

$$\Lambda_k = \left[\begin{array}{ccc|c} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & 0 \\ & & & \lambda_p \\ \hline & 0 & & J \end{array}\right] \tag{38}$$

where $\{ \lambda_1 \cdots \lambda_p \}$ are the unit magnitude eigenvalues and J is a Jordan form matrix containing all the eigenvalues of less than unit magnitude on its main diagonal. (This form is known to exist for a stochastic matrix P_k because the unit magnitude eigenvalues must have a full set of linearly independant eigenvectors.) Therefore:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l = \lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \left[T \Lambda_k T^{-1} \right]^l$$

$$= T \left[\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \Lambda_k^l \right] T^{-1} \tag{39}$$

Since Λ_k has one and only one eigenvalue of one:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \Lambda_k^l = \text{diagonal matrix with a single non-zero element of unity on its main diagonal} \tag{40}$$

because $\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} J^l = 0$ and $\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \lambda_k^l = 0$. Because P_k is a stochastic matrix, the right eigenvector appearing in the column of T corresponding to the unit eigenvalue is a column vector with all its elements equals to 1, i.e. $[\underline{1}]$. Therefore:

$$T \left[\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \Lambda_k^l \right] = [\ \underline{0} \cdots \underline{1} \cdots \underline{0}\ ] \tag{41}$$

Therefore:

$$T \left[\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} \Lambda_k^l \right] T^{-1} = T\ [\ \underline{0} \cdots \underline{1} \cdots \underline{0}\ ] = [\ \underline{e}\ \underline{e} \cdots \underline{e}\ ]^T \tag{42}$$

That is,

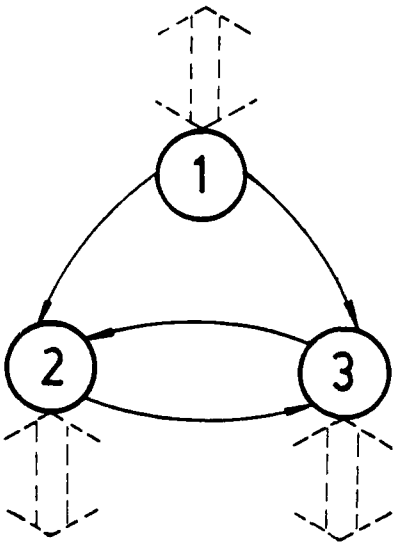
$$\lim_{n \to \infty} \frac{1}{n} \sum P_k^l = T\ \Lambda_k\ T^{-1} = [\ \underline{e}\ \underline{e} \cdots \underline{e}\ ]^T \tag{43}$$

which completes the proof.

As an illustration of the implication of the sufficient condition stated in the Theorem 2, valid and invalid examples of state transition structures are shown in Figure 1. Note that one of the invalid examples in Figure 1b has 2 eigenvalues of one because 2 trapping sets of states are present in single class.

As a result of the Theorems above, the state probability values for

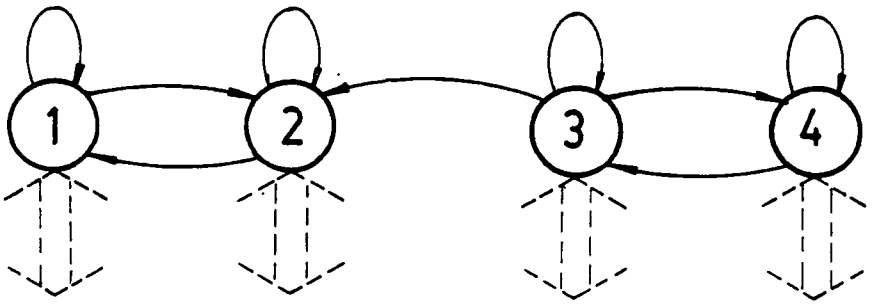
fault-tolerant system models with non-ergodic classes that satisfy the condition stated in Eq.(28) will be approximated well by the approximate technique developed in this paper. Note that there may exist fault-tolerant system models that can also be treated by the approximation technique even when the conditions are violated because the Theorem is a sufficient but not necessary condition.



to and from E_i, $E_i \in E$ and $i \neq j$

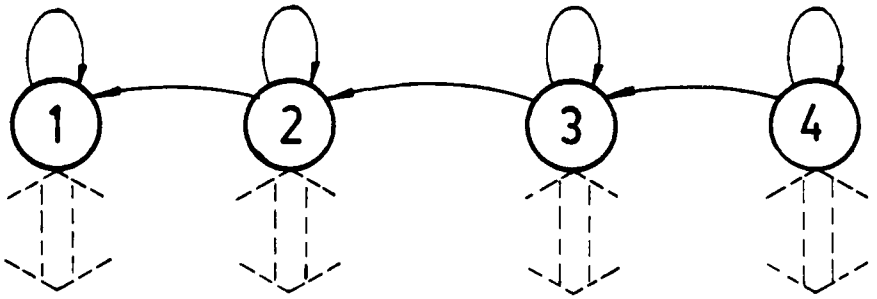
E_j

to and from E_i, $E_i \in E$ and $i \neq j$



E_j

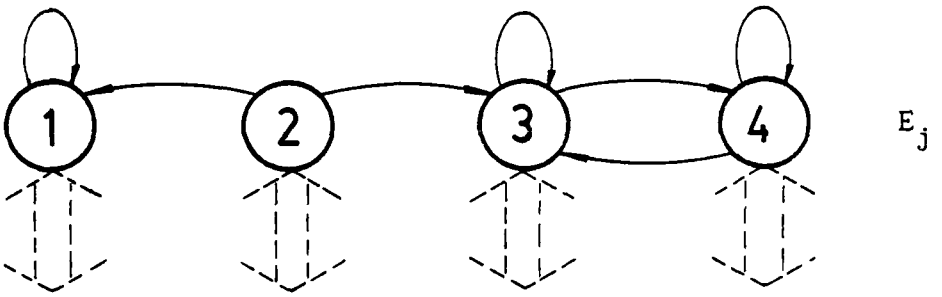
to and from E_i, $E_i \in E$ and $i \neq j$



E_j

to and from E_i, $E_i \in E$ and $i \neq j$

Figure 1a: Valid non-ergodic classes



to and from E_i, $E_i \in E$ and $i \neq j$

Figure 1b: Invalid non-ergodic class

4. TESTS OF APPROXIMATE TECHNIQUE WITH 4-STATE MODLES

In this section, we use 4-state models to demonstrate the use of the approximate technique with relaxed ergodicity condition developed in the preceding sections. Four-state models are the dimensionally smallest models that can be aggregated and can include class-to-class behavior. The approximate technique has also been successfully applied to a 9-state model[18], but the computation required to generate the numerical exact solution in such case starts very costly.

Two cases will be considered here. In case 1, there are two ergodic classes where the second class is a trapping class. The next example, Case 2, consists

of two non-ergodic classes where class 2 is a trapping class.

4.1 Case 1

The schematic state transition diagram for the semi-Markov process is shown in Figure 2.

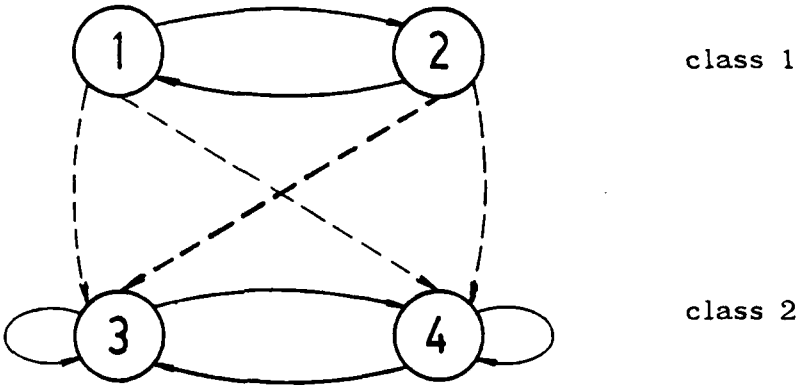


Figure 2: State transition diagram for Case 1

The process can be decomposed into two classes, namely class 1 and class 2, when $\epsilon=0$. Class 1 comprises states 1 and 2 and class 2 comprises states 3 and 4. The transition from class 1 to class 2 is through the small eventual transition probability ϵ from states 1 and 2 to states 3 and 4. However, state 3 and state 4 cannot transit back to any of the states in class 1, hence

class 2 is a trapping class. The governing transition kernel matrix is given by the following:

$$P(t) = \begin{bmatrix} 0 & (1-6\epsilon)\lambda_2 e^{-\lambda_2} & 2\epsilon\lambda_1^2 te^{-\lambda_1 t} & 4\epsilon\lambda_2^2 te^{-\lambda_2 t} \\ (0.3-7\epsilon)\lambda_1 e^{-\lambda_1 t} & (0.7-2\epsilon)\lambda_2 e^{-\lambda_2 t} & 6\epsilon\lambda_1^2 te^{-\lambda_1 t} & 3\epsilon\lambda_2^2 te^{-\lambda_2 t} \\ 0 & 0 & 0.4\lambda_1^2 te^{-\lambda_1 t} & 0.6\lambda_2^2 te^{-\lambda_2 t} \\ 0 & 0 & 0.5\lambda_1^2 te^{-\lambda_1 t} & 0.5\lambda_2^2 te^{-\lambda_2 t} \end{bmatrix} \quad (44)$$

where λ_1=0.2, λ_2=0.1, ϵ=2.5x10^{-6}, (all units are in sec^{-1}).

It is assumed that the initial condition on the state probability vector is,

$$\underline{\pi}(0) = [\ 1 \quad 0 \quad 0 \quad 0\ ] \quad (45)$$

One point about this model should be emphasized. That is that the holding time density functions for the transitions from states in class 1 to states in class 2 and those within class 2 are 2nd order Erlang PDFs. These are non-exponential holding time density functions. Therefore, this is a semi-Markov model.

4.1.1 Stationary probability distribution of the non-perturbed semi-Markov process

By setting ϵ=0 and dropping all the holding time density functions in the transition kernel matrix, the transition probability matrix of the

non-perturbed Markov chain is found to be :

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix} \tag{46}$$

Note that each class of the nonperturbed process is ergodic for this case.By raising the single step transition probability matrix successively to higher powers, the stationary interval transition probability matrix is found.

The stationary probability vectors of the non-perturbed imbedded Markov chain in class 1 and 2, respectively, are:

$$\underline{\pi}_M^{(1)} = [\ 0.2308 \quad 0.7692\ ] \tag{47}$$

$$\underline{\pi}_M^{(2)} = [\ 0.4545 \quad 0.5455\ ] \tag{48}$$

The mean waiting times for the states in class 1 are,

$$\bar{\tau}_1 = p_{12}\,\frac{1}{\lambda_2} = 10 \tag{49}$$

$$\bar{\tau}_2 = p_{21}\,\frac{1}{\lambda_1} + p_{22}\,\frac{1}{\lambda_2} = 8.5 \tag{50}$$

Therefore the mean waiting time for class 1 is

$$\bar{\tau} = \sum_{i \in E_1} \pi_{M_i} \bar{\tau}_i = 8.8462 \tag{51}$$

Hence, the stationary probabilities in class 1 are

$$\pi_1^{(1)} = \frac{\pi_{M_1} \bar{\tau}_1}{\bar{\tau}} = 0.2609 \tag{52}$$

$$\pi_2^{(1)} = \frac{\pi_{M_2} \bar{\tau}_2}{\bar{\tau}} = 0.7391 \tag{53}$$

or in vector form,

$$\underline{\pi}^{(1)} = [\ 0.2609 \quad 0.7391\ ] \tag{54}$$

The mean waiting times for the states in class 2 are:

$$\bar{\tau}_3 = p_{33}\frac{2}{\lambda_1} + p_{34}\frac{2}{\lambda_2} = 16 \tag{55}$$

$$\bar{\tau}_4 = p_{43}\frac{2}{\lambda_1} + p_{44}\frac{2}{\lambda_2} = 15 \tag{56}$$

Therefore the meaning waiting time of the process in class 2 is,

$$\bar{\tau} = \sum_{i\in E_2} \pi_{M_i}\bar{\tau}_i = 15.4545 \tag{57}$$

Hence, the stationary probabilities in class 2 are,

$$\pi_3^{(2)} = \frac{\pi_{M_3}\bar{\tau}_3}{\bar{\tau}} = 0.4705 \tag{58}$$

$$\pi_4^{(2)} = \frac{\pi_{M_4}\bar{\tau}_4}{\bar{\tau}} = 0.5295 \tag{59}$$

or in vector form,

$$\underline{\pi}^{(2)} = [\ 0.4705 \quad 0.5295\ ] \tag{60}$$

4.1.2 Approximate Markov chain

In both of the cases in this section, the time scale factor δ is assumed to equal ϵ for evaluating the approximate Markov chain. The Laplace transform of the kernel element for transition from aggregated "state" 1 to aggregated "state" 2 is given by Eq.(20) with parameter defined by Eq.(21)-(27).

From the transition kernel matrix in Eq.(44),

$$q_1^{(1)} = 6 \tag{61}$$

$$q_2^{(1)} = 7+2 = 9 \tag{62}$$

Substituting all the numerical quantities in Eq.(23):

$$\Lambda_1 = \frac{0.2308\times6 + 0.7692\times9}{0.2308\times10 + 0.76992\times8.5} = 0.9391 \tag{63}$$

Obviously from the structure of the class to class transitions,

$$p_{12} = 1 \tag{64}$$

Therefore,

$$\phi_{12}(s) = \frac{0.9391}{s + 0.9391} \tag{65}$$

or in the scaled time domain,

$$\phi_{12}(t') = 0.9391\ e^{-0.9391\ t'} \tag{66}$$

Since there are only two classes and class 2 cannot transit to class 1, the approximate probability in class 2 is given in scaled time by,

$$\begin{aligned}\pi_2^e(t') &= \int_0^{t'} \phi_{12}(\tau)\ d\tau \\ &= 1 - e^{-0.9391\ t'}\end{aligned} \tag{67}$$

and the approximate probability in class 1 in scaled time is,

$$\begin{aligned}\pi_1^e(t') &= 1 - \pi_2^e(t') \\ &= e^{-0.9391\ t'}\end{aligned} \tag{68}$$

Converting to the original time scale by using $\delta=2.5\times10^{-6}$, this becomes

$$\pi_1^e(t) = e^{-0.9391\ \times\ 2.5\times10^{-6}\ t} \tag{69}$$

$$\pi_2^e(t) = 1 - e^{-0.9391\ \times\ 2.5\times10^{-6}\ t} \tag{70}$$

4.1.3 Exact solution of the original semi-Markov process

In order to determine the accuracy of the approximate results above, it is necessary to generate the exact solution to the original semi-Markov model. The exact solution will be calculated analytically by using the Laplace

transform technique as in Eq.(11.5.6)[12]. Although there are only four states in the model to be solved, the manipulation will require the use of a powerful symbolic manipulation program called MACSYMA.

Two of the elements of the interval transition probability matrix are obtained by this procedure as follows:

$$\pi_1(t) = 0.030115\,[\, e^{-2.35\times10^{-6}\,t} - e^{-0.22999815\,t}\,] + 0.230749\,[\, e^{-2.35\times10^{-6}\,t} - e^{-0.22999815\,t}\,] + 5.384780\times10^{-7}\,t\,e^{-0.1\,t} + 0.538474\,e^{-0.1\,t} + 2.833533\times10^{-5}\,e^{-0.2\,t} \quad (71)$$

$$\pi_2(t) = 0.939775\,[\, e^{-2.35\times10^{-6}\,t} - e^{-0.22999815\,t}\,] + 0.538533\,[\, e^{-2.35\times10^{-6}\,t} - e^{-0.22999815\,t}\,] - 5.769362\times10^{-7}\,t\,e^{-0.1\,t} - 0.538483\,e^{-0.1\,t} - 5.000000\times10^{-5}\,e^{-0.2\,t} \quad (72)$$

The total probability in class 1 is given by:

$$P_{E_1}(t) = \pi_1(t) + \pi_2(t) \quad (73)$$

4.1.4 Comparison of results

The approximate and exact total probabilities in class 1 at different time

points are compared in Table 1.

Table 1
Comparison of approximate and exact probability in class 1

t/sec.	approximate class probability $\pi_1^e(t)$	exact class probability $P_{E_1}(t)$
1	0.99999	1.00000
5	0.99999	1.00000
10	0.99998	0.99999
50	0.99988	0.99990
100	0.99977	0.99978
500	0.99883	0.99884
1000	0.99765	0.99767
5000	0.98833	0.98834
10000	0.97680	0.97697

The results indicate that errors in the aproximation occur only at the fifth decimal place up to t=10000 sec. with the maximum relative percentage error occuring at t=1000 sec. with a value of only 0.0002%. This shows that the class probability is well approximated by the approximate Markov chain.

After the class probability results have been compared, the transient normalized probability vector within class 1, as shown in Table 2, is compared with the stationary probability vector of the non-perturbed semi-Markov process in that class which is given in Eq.(54).

Table 2

Normalized probability distribution in class 1

t/sec.	state 1	state 2
1	0.9075	0.0925
5	0.6510	0.3490
10	0.4791	0.5209
40	0.2707	0.7293
100	0.2609	0.7391
200	0.2609	0.7391

It is easy to see that there is no error up to 4 decimal places between the stationary normalized probability vector and stationary probability vector after a transient period of 100 seconds. This implies that the original semi-Markov process solution is approximated to within 0.0002% error by the approximate solution after a transient period of 100 seconds at the beginning of the mission. The transient period is about 12 times the minimum mean waiting time among the states of the non-perturbed process in class 1 and 0.025% of the MTTF.

4.2 Case 2

For some fault-tolerant system semi-Markov models, there may be trapping

states among some classes of states. Under these circumstances, the imbedded Markov chain of those classes will be non-ergodic. In order to demonstrate the relaxed approximate technique, in this example, a model with two non-ergodic classes is created where each class consists of two states. The state transition diagram is shown in Figure 2 and the process is governed by the transition kernel matrix in Eq.(74)

$$P(t) = \begin{bmatrix} (0.5-\epsilon)\lambda_1 e^{-\lambda_1 t} & (0.5-5\epsilon)\lambda_2 e^{-\lambda_2 t} & 2\epsilon\lambda_3 e^{-\lambda_3 t} & 4\epsilon\lambda_4 e^{-\lambda_4 t} \\ 0 & (1-9\epsilon)\lambda_2 e^{-\lambda_2 t} & 6\epsilon\lambda_3 e^{-\lambda_3 t} & 3\epsilon\lambda_4 e^{-\lambda_4 t} \\ 0 & 0 & 0.4\lambda_3 e^{-\lambda_3 t} & 0.6\lambda_4 e^{-\lambda_4 t} \\ 0 & 0 & 0 & \lambda_4 e^{-\lambda_4 t} \end{bmatrix} \quad (74)$$

where $\lambda_1=0.2$, $\lambda_2=0.1$, $\lambda_3=0.4$, $\lambda_4=0.5$, $\epsilon=2.5\times10^{-6}$, (all units are in sec^{-1}).

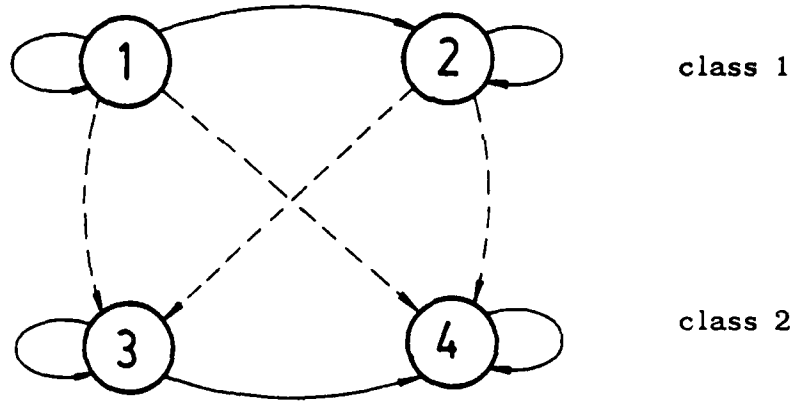


Figure 3: State transition diagram for Case 2

4.2.1 Stationary probability distrubution of the non-perturbed semi-Markov process

By decomposing the transition kernel matrix, the transition probability matrix of the non-perturbed imbedded Markov chain is as follows:

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{75}$$

By raising the transition probability matrix to successively higher powers until stationarity is established, the stationary probability vector of the non-perturbed imbedded Markov chain in classes 1 and 2 are found to be:

$$\underline{\pi}_M^{(1)} = [\, 0 \quad 1 \,] \tag{76}$$

$$\underline{\pi}_M^{(2)} = [\, 0 \quad 1 \,] \tag{77}$$

Hence, the stationary probability vectors of the non-perturbed semi-Markov process are,

$$\underline{\pi}^{(1)} = [\, 0 \quad 1 \,] \tag{78}$$

$$\underline{\pi}^{(2)} = [\, 0 \quad 1 \,] \tag{79}$$

4.2.2 Approximate Markov chain

The Laplace transform of the transition kernel for transition from aggregated "state" 1 to "state" 2 is again given by Eq.(20). From the transition kernel matrix in Eq.(73),

$$q_1^{(1)} = q_{22}^{(1)} = 9 \tag{80}$$

$$\tau_2^{(1)} = p_{22}\, \tau_{22} = 0.9 \tag{81}$$

Substituting the above quantities and Eq.(78) into Eq.(23) gives

$$\Lambda_1 = \frac{0 \times q_1^{(1)} + 9}{0 \times \tau_1^{(1)} + 10} = 0.9 \tag{82}$$

From the structure of the model,

$$p_{12} = 1 \tag{83}$$

So, the transition kernel element for transitions from aggregated "state" 1 to 2 in the new time scale is,

$$\phi_{12}(s) = \frac{0.9}{s + 0.9} \tag{84}$$

or, in the scaled time domain,

$$\phi_{12}(t') = 0.9\, e^{-0.9\, t'} \tag{85}$$

Because there are only two classes, therefore:

$$\pi_1^e(t') = e^{-0.9\, t'} \tag{86}$$

$$\pi_2^e(t') = 1 - e^{-0.9\, t'} \tag{87}$$

In the original time scale, this becomes

$$\pi_1^e(t) = e^{-0.9 \times 2.5 \times 10^{-6}\, t} \tag{88}$$

$$\pi_2^e(t) = 1 - e^{-0.9 \times 2.5 \times 10^{-6}\, t} \tag{89}$$

This completes the derivation of the approximate results for this model. Clearly, it is rather trivial to expand the class probabilities by the stationary probabilities in each class. The results are

$$\pi_1(t) \approx 0 \tag{90}$$

$$\pi_2(t) \approx \pi_1^e(t) \tag{91}$$

$$\pi_3(t) \approx 0 \tag{92}$$

$$\pi_4(t) \approx \pi_2^e(t) \tag{93}$$

4.2.3 Exact solution of the original semi-Markov process

Exact solutions in closed form of the state probabilities in class 1 were evaluated with the help of MACSYMA similar to that of Case 1. The results for this model are:

$$\pi_1(t) = 1.0039\times10^{5}\ e^{-9.9999\times10^{-2}\ t} - 1.00338\times10^{5}\ e^{-1.00000\times10^{-1}\ t} + 3.33334\times10^{-6}\ e^{4.0\times10^{-1}\ t} + 7.50000\times10^{-6}\ e^{-5.0\times10^{-1}\ t} - 5.6449\times10^{-9} \quad (94)$$

$$\pi_2(t) = 1.54588\times10^{-12}\ e^{-5.00014\times10^{-12}t}\ [6.49386\times10^{16}\ \cosh(4.99991\times10^{-12}t) - 6.49373\times10^{16}\ \sinh(4.99991\times10^{-12}\ t)] - 1.00382\times10^{5}\ e^{-9.99999\times10^{-2}t} - 1.24998\times10^{-6}\ e^{-4.0\times10^{-1}t} - 5.62498\times10^{-7}\ e^{-5.0\times10^{-1}t} - 1.47474\times10^{-4} \quad (95)$$

The total probability of occupying class 1 is then $\pi_1(t)+\pi_2(2)$ and the normalized probabilities in this class aare simply π_1 and π_2 normalized by their sum.

4.2.4 Comparison of results

The total probability of occupying class 1 obtained from the approximate Markov chain and from the analytical solution of the original semi-Markov process are compared in Table 3. The exact and approximate solutions listed in the table agree to four decimal places, except after one million seconds have elapsed where the error occurs in the fourth decimal place.

The exact normalized transient probability vector within class 1 is shown in Table 4. It can be seen from these results that the stationary normalized probability vector agrees with the stationary probability vector of the

non-perturbed semi-Markov process in class 1. The approximate state probabilities converges to the exact solution within 0.0003 absolute error after t=100 seconds.

This example, which consists of two non-ergodic classes, shows that the original process aggregated transient probability vector is well approximated by the approximate Markov chain. Furthermore, the normalized probability vector in class 1 converges to the stationary probability vector of the non-perturbed process after a brief transient period. Notice that this example demonstrates the relaxation of the ergodicity condition described in section 3.

Table 3

Compaarison of approximate and exact probability in class 1

t/sec.	approximate class 1 probability $\pi_1^e(t)$	exact class 1 probability $\pi_1^e(t)$
10	0.99997	0.99998
100	0.99976	0.99978
200	0.99954	0.99955
500	0.99886	0.99888
1000	0.99774	0.99775
5000	0.98880	0.98881
10000	0.97773	0.97775
1000000	0.10527	0.10540

Table 4

Normalized probability distribution in class 1

t/sec.	state 1	state 2
10	0.55183	0.44817
100	0.00027	0.99973
200	0.00000	1.00000
500	0.00000	1.00000

5 CONCLUSIONS AND CONTRIBUTIONS

The approximate technique presented in this paper can be used to quantify the performance of those fault-tolerant systems with component failure rates small relative to the fault detection and isolation decision rates. This paper has shown that the approximate technique can be a practical tool to simplify the quantification of large complex fault-tolerant system performance and might also be an efficient tool in the synthesis of such system designs. The particular contributions of this paper can be summarized as follows:

1. Korolyuk's limit Theorem was extended by generalizing the form that the transition kernel elements may take. In particular, they may depend through the holding time distribution on a time scale factor δ in addition to depending on the small parameter ϵ that divides the state space of the system into classes.

2. An approximate technique based on this extended Theorem was then presented, by which the state probability vector of a fault-tolerant system semi-Markov model can be approximated by expanding a reduced order Markov chain state probabilities by the stationary probability vectors of the non-perturbed processes within the disjoint classes. The direct benefit if this approximate technique is a large reduction of hte computational cost of generationg results for large models. Therefore, models of large complex fault-tolerant systems become tractable.
3. An extended theorem with the relaxation of the ergodicity condition stated in Korolyuk's original work was also presented and proved in section xx. As a result, the approximate technique can be applied to a wider scope of fault-tolerant system models, including those with certain types of non-ergodic classes. One of the examples in section 4 demonstrated the use of this relaxed condition.

Acknowledgement

The authors wish to thank the Air Force Office of Scientific Research, which provided the support for this work under Grant AFOSR-84-0160.

REFERENCES

1. Shooman, M.L., *Probabilistic Reliability: An Engineering Approach*. McGraw-Hill, New York, 1968.

2. Walker, B.K. and Gai, E., Fault Detection Threshold Determination Technique Using Markov Theory, *Journal of Guidance and Control* 2(4)(1979), pp. 313-319.

3. Harrison, J.V., Daly, K.C., and Gai, E., Reliability and accuracy prediction for a redundant strap down navigator, *Proc. of AIAA Guidance & Control Conf.*, 1980, pp. 403-413.

4. Takahashi, Weak D-Markov chain and its application to a queuing network, *Mathematical Computer Performer and Reliability*, G. Iazeolla, P.J. Courtois, and Hordijk, Eds, Amsterdam, The Netherlands: North-Holland, 1984, pp. 153-165.

5. Courtois, P.J., *Decomposability: Queueing and Computer System Applications*, New York, Academic, 1977.

6. Bobbio A. and Trivedi K.S., An aggregation technique for the transient analysis of stiff Markov chains, *IEEE Trans. Computers*, C-35(1986), pp. 803-814.

7. Wald A., *Sequential Analysis*, Wiley, New York, 1947; reprinted by Dover, New York, 1973.

8. Walker, B.K., A *semi-Markov Approach to Quantifying Fault-Tolerant System Performance*. ScD. thesis, Department of Aeronautics and Astronautics, M.I.T., Cambridge, MA, July, 1980.

9. Willsky A.S., A survey of design methods for failure detection in dynamics systems, *Automatica*, 12(6)(1976), pp. 601-611.

10. Walker, B.K. and Gai, E. Semi-Markov Theory and the Performance of Redundant Systems with Sequential Fault Tests. Submitted.

11. Levy P., Systemes Semi-Markoviens ayant au plus une infinite denombrable d'etats possibles, *Proc. of Int. Congress on Math.*, Amsterdam, 2(1954), pp. 294.

12. Howard, R.A., *Dynamic Probabilistic Systems, Volume 2: Semi-Markov and Decision Processes*. Wiley & Sons, New York, 1971

13. Gai, E., Harrison, J. and Luppold, R., Reliability analysis of a dual redundant engine controller. *Proc. of SAE Aerospace Congress and Explosition*. Anaheim, CA Oct., 1981.

14. Stiffler, J.J., Computer-aided reliability estimation, *Proc.*

AIAA/NASA/IEEE/ACM Comput. Aerospace Conf., (1977), pp. 427-434.

15. Geist R., Trivedi K., Dugan J.B., and Smotherman M, Modeling imperfect coverage in fault-tolerant systems, *Proc. IEEE 14th Fault-Tolerant Comput. Symp.*, (1984), pp.77-82.

16. Trivedi K., Reliability evaluating of fault-tolerant systems, *Mathematical Computer Performance and Reliability*, G. Iazeolla, P.J. Courtois, and A. Hordijk, Eds. Amsterdam, The Netherland: North-Holland, (1984), pp. 403-414.

17. Trivedi K., Geist R. Smotherman M., and Dugan J.B., Hybrid modeling of fault-tolerant systems, *Comput. Elec. Eng. Int. J.*, 11(2,3)(1985), pp. 87-108.

18. Chu, S.K., *Approximate Behavior of Generalized Markovian Models of Fault-Tolerant Systems*. Master's thesis, Department of Aeronautics and Astronautics, M.I.T., Cambridge, MA, Feburary, 1986.

19. Walker, B.K., Approximat Evaluation of Reliability and Availability via Perturbation Analysis. Proposal submitted to Air Force Office of Scientific Research Reliability Initiative Program, August, 1983.

20. Korolyuk, V.S., Polishchuk, L.I. and Tomusyak, A.A., A Limit Theorem for Semi-Markov Processes. *Kybernetika* 5(4)(1969), pp. 144-155.

21. Korolyuk, V.S. and Turbin, A.F., Asymptotic Enlarging of Semi-Markov Processes with an Arbitrary State Space, A. Dold and B. Eckmann (eds.), *Lecture Notes in Mathematics* 550: *Proceedings of the 3rd Japan-USSR Symposium on Probability Theory*, Springer-Verlog, 1972.

22. Chung, K.L., *Markov Chains with Stationary Transition Probabilities*. Springer, New York, 1967.

Appendix B

Approximate Evaluation of Semi-Markov Chain Reliability Models

Norman M. Wereley
Dept. of Aeronautics and Astronautics
Massachusetts Institute of Technology
77 Massachusetts Ave., Rm 33-105
Cambridge, MA 02139

Bruce K. Walker
Dept. of Aerospace Engineering
and Engineering Mechanics, ML-70
University of Cincinnati
Cincinnati, OH 45221

February 22, 1988

Abstract

A property observed in high reliability fault tolerant control systems is the relatively rare occurrence of component failures compared to the frequent occurrence of redundancy management decision events. This property leads to a temporal decomposition of the semi-Markov chain reliability model into two time scales: a slow time scale for failure events, a fast time scale for FDI events. Conditions are described under which a perturbed semi-Markov chain can be approximated by an enlarged Markov process, the parameters of which are derived from the parameters of the semi-Markov chain.

1 Introduction

A typical fault-tolerant control system (FTCS) is composed of many highly reliable redundant components including sensors, actuators, power supplies and computers. These components are networked in a hierarchical architecture, and their use is governed by a redundancy management (RM) policy. Failure detection and isolation (FDI) logic is implemented to indicate to the RM system which components are no longer safely usable.

It has been demonstrated [1,2] that the reliability and availability of an FTCS can be computed using a finite-state generalized Markov (that is, Markov or semi-Markov) reliability model. These calculations are often difficult or impossible to accomplish by classical combinatorial methods due to time-ordered event sequences that are a consequence of the RM policy and FDI logic. If sequential tests are used to detect failures [3], then a semi-Markov chain reliability model must be used to predict the system reliability.

Many methods exist for the simplified analysis of the steady state behavior of generalized Markov chain models. However, generalized Markov chains model of FTCS invariably contain one or more trapping states that represent system loss. Thus, the steady state behavior is of no interest because the steady state condition will certainly be system loss. It is the transient behavior of these models that is of interest.

A generalized Markov chain is characterized by a discrete set of states and an arbitrary distribution of the holding or sojourn time for each transition. The semi-Markov chain specializes to a Markov chain when the holding times are geometrically distributed and identically distributed for all transitions exiting a particular state.

The result that must be routinely computed in analyzing the reliability model is the interval transition probability, $\phi_{ij}(n)$, which is the probability that the model occupies state j at time n given that it entered state i at the initial time. For FTCS, the states represent a complete characterization of the condition of the system. Thus, if all of the $\phi_{ij}(n)$ that correspond to system loss configurations for j can be computed for n corresponding to the finite duration of the mission, then the probability of an unsuccessful mission can be computed.

Once the interval transition probabilities have been determined for a particular time n, the probability of occupying each state can be determined if the initial state occupancy probabilities are known. Let $\underline{\pi}(n)$ be the state probability distribution at time m. If $\underline{\pi}(0)$ is known, then

$$\underline{\pi}(n) = \underline{\pi}(0)\Phi(n) \tag{1}$$

In the context of the FTCS, the first state is routinely chosen to represent the situation where all components are working. Usually, the system occupies the first state with probability one at the initial time.

The interval transition probabilities are generated by the semi-Markov chain recursion formula [4]:

$$\mathbf{\Phi}(n) = {}^{>}\mathbf{W}(n) + \sum_{m=0}^{n} \mathbf{G}(n)\mathbf{\Phi}(n-m); \;\; IC: \; \mathbf{\Phi}(0) = \mathbf{I} \tag{2}$$

Taking z-transforms of both sides of (2) and solving for $\mathbf{\Phi}(z)$:

$$\mathbf{\Phi}(z) = [\mathbf{I} - \mathbf{G}(z)]^{-1>}\mathbf{W}(z) \tag{3}$$

The z-transform of the state occupancy probability distribution is

$$\underline{\pi}(z) = \underline{\pi}(0)[\mathbf{I} - \mathbf{G}(z)]^{-1>}\mathbf{W}(z) \tag{4}$$

which follows directly from (2). The inverse matrix of $[\mathbf{I} - \mathbf{G}(z)]$ always exists for a semi-Markov chain. The inverse transform of either $\underline{\pi}(z)$ or $\mathbf{\Phi}(z)$ can be found using standard partial fraction expansion techniques. However, for all but the simplest of situations, transform methods are useless in a practical sense.

In practice, the interval transition probability matrix is nearly always found by performing the semi-Markov recursion numerically. For a model with N states, computation of $\mathbf{\Phi}(n)$ requires storage of $2nN^2$ values because both $\mathbf{\Phi}(n)$ and $\mathbf{G}(n)$ must be stored for all times prior to and including time n. A reliability model for a typical inertial navigation system might have twenty states, a sampling period of 200ms, and a two hour mission time. This would require storage of 2.88×10^7 single precision values and require 230 megabytes of storage. Moreover, the number of floating point multiplications required to compute $\underline{\pi}(n)$ from $\underline{\pi}(0)$ is about n^2N^2 - which is 2.59×10^{11} for the example described above. Thus, the computational burden and memory requirements are tremendous even for a simple system.

The problem to be addressed in this paper is to substantially reduce the computational burden while preserving the accuracy of reliability and availability calculations.

One possible means for doing this is direct Monte Carlo techniques. If a sufficient number of Monte Carlo simulations are made of system operations to account correctly for all possible random events that bear on the reliability calculation, then any aspect of system performance can be evaluated. To obtain meaningful results for high reliability systems with events that occur with probabilities as low as 1×10^{-9} (typical of the probability of a component failure over a single time step), over one billion simulations must be performed.

This task is as formidable as evaluating the semi-Markov chain recursion for large values of the time index. Consequently, reliability calculations via direct Monte Carlo methods also have prohibitive computational costs.

Lewis suggested in [5,6] that a modified Monte Carlo approach be used for high reliability systems. Again, failure events are assumed to be extremely rare relative to other events that occur in the system. Thus, the vast majority of simulations will be those for which no failures occur. Lewis assumes that all events have exponentially distributed times of occurrence and can be modeled by a Markov chain. It is possible to sample the failure distributions before a simulation is initiated to determine if any failures will occur during the mission. If all failures occur after the mission has been completed (which is usually the case), then a normal simulation results. If a failure occurs during the mission, then the complete simulation must be performed including FDI decisions, decision errors, and repairs. However, this approach does not apply to semi-Markov chains because FDI events arising from a sequential FDI test are not exponentially distributed. In these cases, a complete simulation must always be run and no benefits are derived from the modified technique.

Another approach that exploits the rare occurrence of failure events is suggested by Trivedi in [7,8]. The model is based upon a time-scale decomposition of the system into virtually disjoint fault-occurrence and fault handling submodels. The fault-handling submodels represent aggregated states and the failure occurrence submodels dictate the behavior between these aggregated states. The reliability of the system predicted by the aggregated model is then computed using Markov or Monte Carlo techniques. However, the only fault-handling events that are accounted for are detections and missed detections following actual faults. A common FDI event that cannot be treated by these hybrid models is the false alarm, which occurs in the absence of a fault. Therefore, this approach is limited to systems where false alarms cannot occur.

In this paper, the relatively rare occurrence of component failures relative to RM decision events will be exploited in the development of an approximate method for evaluating semi-Markov chain reliability models of fault tolerant control systems.

2 A Limit Theorem for Semi-Markov Chains

Theorem 1 describes how a perturbed semi-Markov chain, which is dependent on a small parameter ϵ in a certain way, can be described asymptotically by an enlarged Markov process as $\epsilon \to 0$. This theorem is an extension of the results for discrete parameter semi-Markov processes stated in [9].

The semi-Markov chain depends on a small parameter ϵ such that the entire state space of the semi-Markov chain can be decomposed into disjoint classes of states where the probabilities of departure from each class tend to zero with ϵ. Also, the total sojourn in each class is assumed to have a non-degenerate distribution in the limit as $\epsilon \to 0$. (When $\epsilon = 0$, the chain will be referred to as the unperturbed semi-Markov chain while the ϵ–dependent chain will be referred to as the perturbed semi-Markov chain.)

Theorem 1 (Limit Theorem for Semi-Markov Chains) *Let the set E of states of the semi-Markov chain be expressible as a union of disjoint classes:*

$$E = \sum_{k=1}^{N^e} E_k \quad k \in M \equiv \{1, 2, \ldots, N^e\}\,. \tag{5}$$

Let $\tau_{kr}^{(i)}$ be the sojourn of the semi-Markov chain in class E_k when it starts from state $i \in E_k$ and moves to class E_r where $r \neq k$. If the following two conditions hold for the semi-Markov chain E:

1. The elements of the core matrix sequence $\{g_{ij}^{\epsilon}(n) \mid i, j \in E\}$ specifying the semi-Markov chain depend as follows on the small parameter ϵ:

$$^{\leq}g_{ij}^{\epsilon}(n) = p_{ij}^{\epsilon} {}^{\leq}h_{ij}\left(\frac{n}{\epsilon}\right) \tag{6}$$

where $^{\leq}h_{ij}(0) = 0$. The p_{ij}^{ϵ} can be expanded in a Taylor series about ϵ=0. Retaining terms that are linear in ϵ:

$$p_{ij}^{\epsilon} = \begin{cases} p_{ij}^{(k)} - \epsilon q_{ij}^{(k)} + O(\epsilon) & \text{if } i, j \in E_k \\ \epsilon q_{ij}^{(k)} + O(\epsilon) & \text{if } i \in E_k \text{ and } j \notin E_k \end{cases} \tag{7}$$

The embedded Markov chain for ϵ=0 obeys the usual Markov chain properties:

$$\sum_{j \in E_k} p_{ij}^{(k)} = 1; \text{ and } p_{ij}^{(k)} \in [0, 1];\ \forall k \in M \tag{8}$$

2. The embedded Markov chains defined by the matrices $\left\{p_{ij}^{(k)} \mid i,j \in E_k \; \forall k \in M\right\}$ *are ergodic with stationary distributions* $\left\{\pi_i^{(k)} \mid i \in E_k \; \forall k \in M\right\}$.

Then:

$$\lim_{n\to\infty} Pr\left\{\tau_{kr} \le t\right\} = \gamma_{kr}\left\{1 - \exp\left[\frac{-\Lambda_k t}{T}\right]\right\} \tag{9}$$

where:

$$\gamma_{kr} \equiv \frac{\sum_{i\in E_k} \pi_i^{(k)} q_i^{(kr)}}{\sum_{i\in E_k} \pi_i^{(k)} q_i^{(k)}} \tag{10}$$

$$\Lambda_k \equiv \frac{\sum_{i\in E_k} \pi_i^{(k)} q_i^{(k)}}{\sum_{i\in E_k} \pi_i^{(k)} a_i^{(k)}} \tag{11}$$

Here:

$$q_i^{(kr)} \equiv \sum_{j\in E_r} q_{ij}^{(k)} \tag{12}$$

$$q_i^{(k)} \equiv \sum_{j\in E_k} q_{ij}^{(k)} \tag{13}$$

$$a_i^{(k)} \equiv \sum_{j\in E_k} p_{ij}^{(k)} \bar{\tau}_{ij} \tag{14}$$

$$\bar{\tau}_{ij} \equiv \sum_{n=0}^{\infty} n\, h_{ij}(n) \tag{15}$$

PROOF: Let $\epsilon\zeta_{ij}$ denote the integer valued sojourn of the semi-Markov chain in state i with next transition to state j with the holding time distribution ${}^{\le}h_{ij}(n/\epsilon)$ while the δ_{ij}^{ϵ} are the transition indicators from state i to state j. The probability distribution of the random quantities $\tau_{kr}^{(i)}$ can be expressed in terms of total probability as

$$Pr\left\{\tau_{kr}^{(i)} \le n\right\} = \sum_{j\in E_k} Pr\left\{\delta_{ij}^{\epsilon} = 1;\, \epsilon\zeta_{ij} + \tau_{kr}^{(j)} \le n\right\} + \sum_{j\in E_r} Pr\left\{\delta_{ij}^{\epsilon} = 1;\, \epsilon\zeta_{ij} \le n\right\} \tag{16}$$

Defining the interval transition CDF as

$${}^{\le}\phi_{kr}^{(i)}(n) = Pr\left\{\tau_{kr}^{(i)} \le n\right\} \tag{17}$$

then

$${}^{\le}\phi_{kr}^{(i)}(n) = \sum_{j\in E_k} \sum_{m=0}^{n} g_{ij}^{\epsilon}(m)\, {}^{\le}\phi_{kr}^{(j)}(n-m) + \sum_{j\in E_r} {}^{\le}g_{ij}^{\epsilon}(n) \tag{18}$$

Taking z-transforms of both sides yields:

$${}^{\le}\phi_{kr}^{(i)}(z) = \sum_{j\in E_k} g_{ij}^{\epsilon}(z)\, {}^{\le}\phi_{kr}^{(j)}(z) + \left\{\frac{z}{z-1}\right\} \sum_{j\in E_r} g_{ij}^{\epsilon}(z). \tag{19}$$

The z-transforms of the $g_{ij}^{\epsilon}(z)$ must be evaluated to first order in ϵ. From (6) and the definition of the z-transform [10]:

$$g_{ij}^{\epsilon}(z) = p_{ij}^{\epsilon} \sum_{n=0}^{\infty} h_{ij}\left(\frac{n}{\epsilon}\right) z^{-n} \tag{20}$$

Note that p_{ij}^{ϵ} has been moved in front of the summation sign because it does not depend on time. Let $m = n/\epsilon$ and expand $z^{-\epsilon m}$ in a Taylor series about ϵ=0. Then:

$$g_{ij}^{\epsilon}(z) = p_{ij}^{\epsilon} \sum_{m=0}^{\infty} \{1 - \epsilon m \log z\} h_{ij}(m) + O(\epsilon) \tag{21}$$

where $O(\epsilon)$ represents terms such that in the limit as $\epsilon \to 0$, the quantity $O(\epsilon)/\epsilon$ approaches zero. Noting that:

$$\sum_{n=0}^{\infty} h_{ij}(n) = 1 \tag{22}$$

$$\sum_{n=0}^{\infty} n\, h_{ij}(n) = \bar{\tau}_{ij}. \tag{23}$$

and substituting p_{ij}^{ϵ} from (7) and combining terms of $O(\epsilon)$ yields:

$$g_{ij}^{\epsilon}(z) = \begin{cases} p_{ij}^{(k)} \{1 - \epsilon \bar{\tau}_{ij} \log z\} - \epsilon q_{ij}^{(k)} O(\epsilon) & \text{if } i,j \in E_k \\ \epsilon q_{ij}^{(k)} + O(\epsilon) & \text{if } i \in E_k \text{ and } j \notin E_k \end{cases} \tag{24}$$

Incorporating these results into (19) and placing all terms proportional to ϵ on the RHS:

$$\begin{aligned} {}^{\leq}\phi_{kr}^{(i)}(z) - \sum_{j \in E_k} p_{ij}^{(k)} {}^{\leq}\phi_{kr}^{(j)}(z) &= -\epsilon \sum_{j \in E_k} {}^{\leq}\phi_{kr}^{(j)}(z) \left\{ q_{ij}^{(k)} + p_{ij}^{(k)} \bar{\tau}_{ij} \log z \right\} \\ &\quad + \epsilon \left\{ \frac{z}{z-1} \right\} \sum_{j \in E_r} q_{ij}^{(k)} + O(\epsilon) \end{aligned} \tag{25}$$

Now, passing to the limit as $\epsilon \to 0$, the RHS vanishes and the ${}^{\leq}\phi_{kr}(z)$ are found to satisfy the system of equations below:

$${}^{\leq}\phi_{kr}^{(i)}(z) - \sum_{j \in E_k} p_{ij}^{(k)} {}^{\leq}\phi_{kr}^{(j)}(z) = 0 \tag{26}$$

Let $\mathbf{P}_k = \left[p_{ij}^{(k)}\right]$ represent the embedded Markov chain operator in class E_k of the unperturbed semi–Markov chain E. The system of equations in (26) can be expressed as:

$${}^{\leq}\phi_{kr}^{(i)}(z)^T = \mathbf{P}_k\, {}^{\leq}\phi_{kr}^{(j)}(z)^T \tag{27}$$

After successive premultiplication by $\mathbf{P}_k$, and taking the limit as $n \to \infty$:

$$^{\leq}\phi_{kr}(z)^T = \left\{ \lim_{n\to\infty} \mathbf{P}_k^n \right\} {}^{\leq}\phi_{kr}(z)^T \tag{28}$$

Under Condition 2, the ergodic theorem for Markov chains [11] implies that:

$$\lim_{n\to\infty} \mathbf{P}_k^n = \mathbf{P}_k^\infty = \begin{bmatrix} \pi_M^{(k)} \\ \vdots \\ \pi_M^{(k)} \end{bmatrix} \tag{29}$$

so that the solution to (27) is independent of the superscript:

$$^{\leq}\phi_{kr}^{(i)}(z) = {}^{\leq}\phi_{kr}(z) \;\; \forall\, i \in E_k, \; \forall\, k \in M \tag{30}$$

Now, (25) is of the form $f(x) = g(x, \epsilon)$, that is, the LHS is not a function of ϵ and is therefore constant with respect to ϵ. However, as $\epsilon \to 0$, the RHS approaches zero so that the LHS must be zero for all values of ϵ. Canceling ϵ from the result, multiplying by the stationary probabilities of the unperturbed semi-Markov chain in class k, $\pi_i^{(k)}$, and summing over $i \in E_k$ yields:

$$\sum_{i\in E_k} \pi_i^{(k)} \sum_{j\in E_k} {}^{\leq}\phi_{kr}^{(j)}(z) \left\{ q_{ij}^{(k)} + p_{ij}^{(k)} \bar{r}_{ij} \log z \right\} + \left\{ \frac{z}{z-1} \right\} \sum_{i\in E_k} \pi_i^{(k)} \sum_{j\in E_r} q_{ij}^{(k)} + O(\epsilon) \tag{31}$$

On passing again to the limit as $\epsilon \to 0$, noting that all of the $^{\leq}\phi_{kr}^{(i)}(z)$ have the limit function $^{\leq}\phi_{kr}(z)$, and solving for $^{\leq}\phi_{kr}(z)$, the z-transform of the class-to-class transition PMF becomes:

$$\phi_{kr}(z) = \gamma_{kr}\, \Lambda_k \frac{1}{\log z + \Lambda_k} \tag{32}$$

The mapping from the z domain to the s domain (Laplace) is given by $s = (\log z)/T$. Dividing top and bottom by the sampling period T, and applying the transformation concludes the proof. □

In summary, Theorem 1 describes the conditions under which a perturbed semi-Markov chain can be approximated by an enlarged Markov process that evolves in the slow time–scale, and also states how the parameters of the Markov process are determined from the parameters of the semi-Markov chain. In the context of FTCS, the fast time scale behavior within a class would represent FDI decision and RM events while the slower class–to–class

behavior would represent the occurrence of failures. The class–to–class interval transition CDF $\Phi_{kr}(t)$ that results is a continuous time envelope of the behavior between the classes. This interpretation is intuitively satisfying since failures are invariably assumed to have exponentially distributed times of occurrence over continuous time.

However, two problems occur in the application of Theorem 1 to FTCS models: (1) the embedded Markov chains for each class of the unperturbed model are rarely ergodic, and (2) the holding time PMFs are usually functions of n, not n/ϵ, that is, the holding times are typically not on the order of the mean time to a component failure. The requirement that the embedded Markov chains of the unperturbed classes be ergodic is important in producing (26) and guarantees the existence of the stationary probabilities $\left\{\pi_i^{(k)} \mid i \in E_k \forall\, k \in M\right\}$. The ergodicity condition can be relaxed in much the same way as was done in [12] for semi-Markov processes. This will be accomplished in Lemma 2 and Lemma 3. The second problem can be mitigated by introducing time-scaling into Theorem 1, as will be done in Theorem 4.

3 Relaxation of the Ergodicity Condition

Lemma 2 discusses how the existence of the Caesaro limit of the embedded Markov chain operator leads to a relaxation of the ergodicity condition.

Lemma 2 *Consider a semi-Markov chain state space E that can be expressed as a sum of disjoint classes according to (5) and (7). Let $\mathbf{P}_k = \left[p_{ij}^{(k)}\right]$ represent the embedded Markov chain operator for class E_k. The solution of (26) is independent of the superscript (and the results of Theorem 1 hold), if the Caesaro limit exists:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \mathbf{P}_k^l = \Pi_k = \begin{bmatrix} \pi_c^{(k)} \\ \vdots \\ \pi_c^{(k)} \end{bmatrix} \tag{33}$$

PROOF: The system of equations in (26) can be expressed in matrix form as is done in (27). Successively premultiplying both sides by $\mathbf{P}_k$, and averaging an infinite number of these terms:

$$^{\leq}\phi_{kr}(z)^T = \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P_k^l {}^{\leq}\phi_{kr}(z)^T \tag{34}$$

Because the operator P_k satisfies the Caesaro limit from (33), the solution of (26) is independent of the superscript. □

The relaxation due to Lemma 2 demonstrates that the ergodicity condition of Theorem 1 was sufficient, but not necessary. Thus, the conditions under which the Caesaro limit exists should be determined in hopes of finding a necessary condition.

Lemma 3 *Consider a semi-Markov chain state space E that can be expressed as a sum of disjoint classes according to (5) and (7). Let $\mathbf{P}_k = \left[p_{ij}^{(k)}\right]$ represent the embedded Markov chain operator of the unperturbed chain for class E_k. If the embedded Markov chain represented by the operator $\mathbf{P}_k$ is: 1) ergodic, or 2) non-ergodic with one and only one unit eigenvalue, then the Caesaro limit in (34) exists.*

Proof: The proof of this lemma is essentially similar to that in [12]. For details of this proof, see [13]. □

4 Limit Theorem with Time Scaling

In FTCS with small single step component failure probabilities, the holding time PMFs associated with the core matrix sequence elements do not depend on ϵ but only on the FDI decision delay. If a semi-Markov chain is observed in another time scale that is $1/\delta$ times that of the original time scale, then the PMF $h_{ij}(n)$ will be affected but the eventual transition probabilities, p_{ij}^{e}, will remain the same because they characterize the transition probability from state i to state j regardless of when the transition takes place. However, the holding time PMFs in the new time scale are not obtained by simply changing the argument of $h_{ij}(\cdot)$ from n to n/δ. This is because the summation of $h_{ij}(n/\delta)$ for all non-negative values of the time index would not be unity and so would not yield a proper holding time function. The CDF ${}^{\leq}h_{ij}(n)$ associated with the PMF $h_{ij}(n)$ must be determined and the argument of the CDF replaced by n/δ. The new PMF $h'_{ij}(n)$ observed in the new time scale would have most of its probability mass close to the origin. The statistics of the process in the new time scale will depend on the small parameter δ - the time scaling factor.

Theorem 4 (Limit Theorem With Time Scaling) *Let the set E of states of the semi-Markov chain be expressible as a sum of disjoint classes as in (5). Let $\tau_{kr}^{(i)}$ be the sojourn*

of the semi-Markov chain in class E_k when it starts from state $i \in E_k$ and moves to class E_r for $r \neq k$. If the following two conditions hold for the semi-Markov chain E:

1. The elements of the core matrix sequence $\{g_{ij}^{\epsilon}(n) \mid i,j \in E\}$ specifying the semi-Markov chain depend as follows on the small parameters δ and ϵ:

$$^{\leq}g_{ij}^{\epsilon}(n) = p_{ij}^{\epsilon}{}^{\leq}h_{ij}(\frac{n}{\delta}) \tag{35}$$

Here, $^{\leq}h_{ij}(\cdot)$ is the transition CDF of the semi- Markov chain in the original time scale and $^{\leq}h_{ij}(0) = 0$. The p_{ij}^{ϵ} can be expanded in a Taylor series about ϵ=0 as in (7). The embedded Markov chain obeys the usual Markov chain properties described in (8).

2. The embedded Markov chains defined by the matrices $\left\{p_{ij}^{(k)} \mid i,j \in E_k \forall k \in M\right\}$ are ergodic or non-ergodic with one and only one unit eigenvalue with the stationary probabilities (in the Caesaro limit sense) $\left\{\pi_i^{(k)} \mid i \in E_k \forall k \in M\right\}$.

Then:

$$\lim_{n\to\infty} Pr\left\{\tau_{kr} \leq t'\right\} = \gamma_{kr}\left\{1 - \exp\left[\frac{-\Lambda_k t'}{\alpha T}\right]\right\} \tag{36}$$

where the parameters of the enlarged Markov process were defined in Theorem 1 and $\alpha = \delta/\epsilon$.

PROOF: The proof of this theorem is essentially identical to that of Theorem 1. For details of this proof, see [13]. □

It should be noted that an explicit analytical expression of the core matrix sequence, $G^{\epsilon}(n)$, is not required to expand the eventual transition probabilities of the perturbed semi-Markov chain, p_{ij}^{ϵ} in a Taylor series about ϵ=0. The eventual transition probabilities may be evaluated numerically, which is what would be done in practice. This is fortunate because the direct form of the core matrix is not always available [3]. In many cases, the decision time PMFs are tabulated numerically and no functional form is available.

Also, the time scale decomposition of the semi-Markov chain is crucial to the use of this technique. A simple way of characterizing each class is as follows: the first class contains states for which no failures have occurred, the second class contains states for which a single failure has occurred, the third class contains states for which two failures have occurred, etc. These classes arise by setting ϵ=0 and observing which groups of states of the unperturbed semi-Markov chain do not communicate.

Finally, estimates of the original semi-Markov chain state probabilities can be recovered from the enlarged Markov process. The asymptotic behavior of the unperturbed semi-Markov chains in each class are the stationary probabilities (or Caesaro limit probabilities) for that class. The class–to–class behavior is determined by the enlarged process. The approximate state probabilities in each class are:

$$\hat{\pi}_i^{(k)}(n) = \pi_i^k \hat{\pi}_k^e(n) \tag{37}$$

where the approximate class probabilities of the enlarged process are found from its interval transition probability matrix.

5 Performance Evaluation of the SCMS

Two simple semi-Markov reliability models of a single component monitoring system (SCMS) will be developed. The SCMS uses a sequential FDI test to monitor the status (failed or working) of a single component. The two models will differ in monitoring policy. The first example, SCMS-I, models an FDI test that operates continuously over the entire mission duration. The second example, SCMS-II, models an FDI test that is discontinued after the first failure indication (namely, abbreviated monitoring).

In this section, the performance of the SCMS will be evaluated through application of the approximate method to a semi-Markov model. The procedure follows: (1) semi-Markov transition diagrams are constructed describing all of the random events that can take place, (2) the direct form of the core matrix sequence is derived, (3) the core matrix is placed in standard form, (4) the performance is evaluated through application of Theorem 4.

In addition, z-transforms will be used to determine an analytical expression for the state and class occupancy probabilities, $\underline{\pi}(n)$ and $\underline{\pi}^e(n)$ respectively. The results of the z-transform analysis will be used to evaluate the accuracy of the approximate method. This is possible here because the models are relatively simple. In more general cases, this would not be practical.

Table 1: State definitions and class decompositions for SCMS-I,II

State	State Definition	Class
1	Component is working	1
2	Component has a false alarm	1
3	System loss - component failed	2

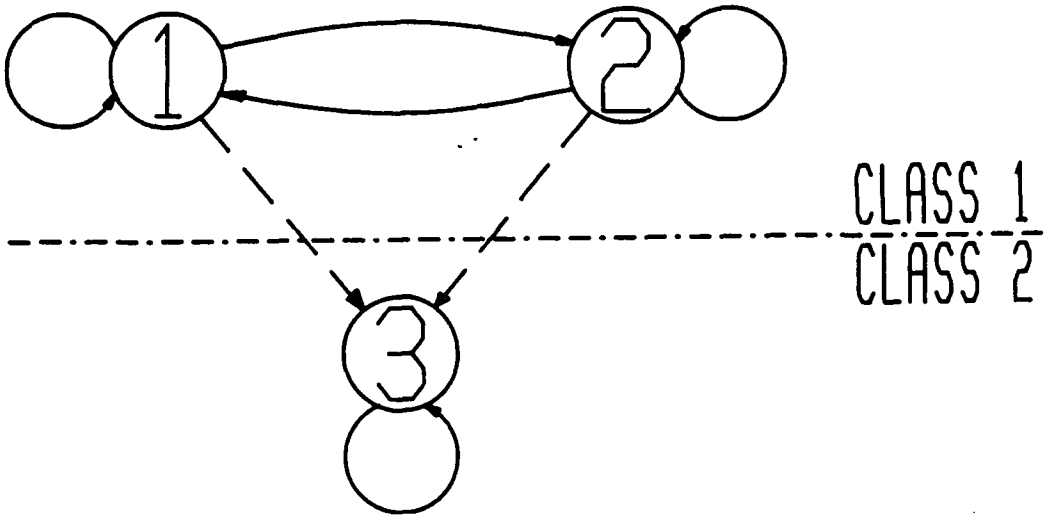


Figure 1: Semi-Markov transition diagram for SCMS-I

5.1 SCMS with continuous monitoring

Table 1 enumerates and defines the states of a semi-Markov chain reliability model of the SCMS-I. The dashed line in the table distinguishes the class decomposition of the model: class 1 contains states 1 and 2, class 2 contains only state 3.

The semi-Markov transition diagram for the SCMS-I is presented in Figure 1. Two aspects of this diagram should be noted. Given that the chain has entered a state, the lines directed out of that state represent transitions after the chain has remained in that state for a period of time, namely, the holding time. Secondly, the dashed lines represent transitions whose transition PMFs are proportional to ϵ. Thus, a dashed line represents the condition that no such transition occurs when $\epsilon = 0$. This is a convenient way of depicting the class decomposition of a semi-Markov chain reliability model.

A complete statistical description of the sequential test used in the FDI process requires knowledge of the conditional PMFs of the time to decision of the test. The following two functions are required:

$f_D^0(n)$ PMF of time to a decision that no failure is present when no failure is present.

$f_D^0(n)$ PMF of time to a failure indication when no failure is presen' (false alarm).

In these PMFs, the fault monitoring event at time n must be conditioned on the failure events that take place prior to and including time $n-1$. Thus, it is assumed that there is a delay of at least a single time step between when a failure takes place and when it can be detected.

Another necessary function is the sum of all probabilities of all possible test outcomes - nominal decision, failure indication, and decision not yet available - at a given time n. Itcan be specified in terms of the decision time PMFs as:

$$Q_0(n) = 1 - \sum_{k=1}^{n-1} \left\{ f_D^0(k) + f_D^0(k) \right\}; \quad n \geq 1 \tag{38}$$

Note that $Q_0(n)$ is defined only for positive values of the time index n and is defined to be zero for $n = 0$. Thus, one of the necessary criteria for a permissible holding time PMF is maintained - there is no probability mass at the initial time.

The core matrix sequence, $\mathbf{G}^\epsilon(n)$, for SCMS-I can be expressed in matrix form as:

$$\mathbf{G}^\epsilon(n) = \begin{bmatrix} (1-\epsilon)^n f_D^0(n) & (1-\epsilon)^n f_D^0(n) & \epsilon(1-\epsilon)^{n-1} Q_0(n) \\ (1-\epsilon)^n f_D^0(n) & (1-\epsilon)^n f_D^0(n) & \epsilon(1-\epsilon)^{n-1} Q_0(n) \\ 0 & 0 & \delta(n-1) \end{bmatrix} \tag{39}$$

Any reasonable PMF may be used for the decision time PMFs. However, a closed form solution for $\underline{\pi}(n)$ is desired. A simple but realistic choice for the decision time PMFs is the hypergeometric PMF [13]. This PMF is a good approximation to the holding time behavior of many sequential tests, as demonstrated by Table 6.6 of [3]. Choosing an appropriate eventual transition probability yields the hypergeometric decision time PMFs below:

$$f_D^0(n) = A_D^0 \left(a^n - b^n\right) \quad ; \quad A_D^0 = (1 - P_{fa}) \frac{(1-a)(1-b)}{(a-b)} \tag{40}$$

$$f_D^0(n) = A_D^0 \left(a^n - b^n\right) \quad ; \quad A_D^0 = P_{fa} \frac{(1-c)(1-d)}{(c-d)} \tag{41}$$

where $0 < b < a < 1$ and $0 < d < c < 1$. The parameter P_{fa} is the eventual false alarm probability of the sequential test. The core matrix can now be expressed in terms of these PMFs.

A z-transform analysis of the semi-Markov recursion formula using the above core matrix sequence yields the state occupancy probability vector $\underline{\pi}(n)$:

$$\pi_1(n) = (1 - P_{fa})\, R^n + P_{fa}\frac{c(1-d)}{(c-d)}(cR)^n - P_{fa}\frac{(1-c)d}{(c-d)}(dR)^n \tag{42}$$

$$\pi_2(n) = P_{fa}\, R^n - P_{fa}\frac{c(1-d)}{(c-d)}(cR)^n + P_{fa}\frac{(1-c)d}{(c-d)}(dR)^n \tag{43}$$

$$\pi_3(n) = 1 - R^n \tag{44}$$

and the class occupancy probability vector $\underline{\pi}^{\epsilon}(n)$:

$$\underline{\pi}^{\epsilon}(n) = [(1-\epsilon)^n\ ,\ 1 - (1-\epsilon)^n] \tag{45}$$

Availability of these analytical results permits comparisons to be made with the approximate results that exploit the class decomposition to be described below. It should be emphasized again that the existence of analytical is rare, and occurs only because the system is very simple.

In order to derive the enlarged Markov process for this model, $\mathbf{G}^{\epsilon}(n)$ must be placed in standard form. For an in–class transition, the decomposition is obtained from the first two terms of the Taylor series expansion of the eventual transition probability about $\epsilon = 0$. In addition, the mean waiting times, $\bar{\tau}_{ij}$, must be derived. For an out– of–class transition, the decomposition is obtained by dividing the eventual transition probability by ϵ and then taking the zeroth order term in the Taylor series expansion about $\epsilon = 0$.

Consider an in-class transition from state 1 to state 1. First, the eventual transition probability is found:

$$p^{\epsilon}_{11} = A^0_D \left\{ \frac{aR}{(1-aR)} - \frac{bR}{(1-bR)} \right\} \tag{46}$$

The decomposition for the transition is:

$$p^{\epsilon}_{11} = 1 - P_{fa} + \epsilon(1 - P_{fa})\frac{(1-ab)}{(1-a)(1-b)} \tag{47}$$

To satisfy the requirements for a permissible holding time function, the holding time function for this transition must be expressed as:

$$h_{11}(n) = \frac{(1-aR)(1-bR)}{(a-b)R}\{(aR)^n - (bR)^n\} \tag{48}$$

From (15), the mean holding time can be found:

$$\bar{r}_{11} = \frac{(1-abR^2)}{(1-aR)(1-bR)} \tag{49}$$

Thus, all of the parameters required to place this in–class transition PMF in standard form have been derived.

A second type of core matrix element that must be placed in standard form is one corresponding to an out–of–class transition such as a transition from state 1 to state 3. First, the eventual transition probability must be found.

$$p^{\epsilon}_{31} = \epsilon(1-P_{fa})\frac{(1-abR)}{(1-aR)(1-bR)} + \epsilon P_{fa}\frac{(1-cdR)}{(1-cR)(1-dR)} \tag{50}$$

The sole parameter required for the approximation technique from this eventual transition probability is found from:

$$q_{31} = \left[\frac{1}{\epsilon}p^{\epsilon}_{31}\right|_{\epsilon=0} = (1-P_{fa})\frac{(1-ab)}{(1-a)(1-b)} + P_{fa}\frac{(1-cd)}{(1-c)(1-d)} \tag{51}$$

The eventual transition probabilities of each row of $\mathbf{G}^{\epsilon}(n)$ sum to unity. Thus, this is a proper semi-Markov chain [4].

The next step in the procedure is to determine the eventual transition probability matrix of the unperturbed semi-Markov chain. This is found by setting $\epsilon = 0$ and ignoring all time varying terms in the core matrix:

$$\mathbf{P} = \begin{bmatrix} 1-P_{fa} & P_{fa} & 0 \\ 1-P_{fa} & P_{fa} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{52}$$

By raising $\mathbf{P}$ to successively higher powers, the stationary interval transition probability matrix is found to be identical to (52). The embedded stationary probability distribution in partitioned form is thus:

$$\underline{\pi}_M = [\,1-P_{fa} \;\; P_{fa} \mid 1\,] \tag{53}$$

With knowledge of this and of the mean holding times for transitions from state i to j, $\bar{\tau}_{ij}$, it is possible to determine the stationary probability distribution of the unperturbed semi-Markov chain, $\underline{\pi}^{(k)}$. This probability distribution is needed to approximate the state probability distribution of the original perturbed semi-Markov chain.

From semi-Markov theory, the stationary probability distribution for each unperturbed class E_k is given by

$$\pi_i^{(k)} = \pi_{M_i}^{(k)} \bar{\tau}_i^{(k)} / \bar{\tau}^{(k)} \tag{54}$$

where $\bar{\tau}^{(k)}$ is the mean waiting time of the chain in class E_k:

$$\bar{\tau} = \sum_{i \in E_k} \pi_{M_i}^{(k)} \bar{\tau}_i^{(k)} \tag{55}$$

$\underline{\pi}_M^{(k)}$ was determined above, and $\bar{\tau}_i^{(k)}$ is the mean holding time in state i:

$$\bar{\tau}^{(k)} = \sum_{j \in E_k} p_{ij}^{(k)} \bar{\tau}_{ij}^{(k)} \tag{56}$$

where $\bar{\tau}_{ij}^{(k)}$ is determined from the limit of $\bar{\tau}_{ij}$, defined in (15), as $\epsilon \to 0$.

The stationary probability distribution of the unperturbed semi-Markov chain will now be determined. The mean holding times of the unperturbed semi-Markov chain in the first class are:

$$\bar{\tau}_{11}^{(1)} = \bar{\tau}_{12}^{(1)} = \frac{(1-ab)}{(1-a)(1-b)} \tag{57}$$

$$\bar{\tau}_{21}^{(1)} = \bar{\tau}_{22}^{(1)} = \frac{(1-cd)}{(1-c)(1-d)} \tag{58}$$

The mean holding time in class 1 starting from state i is thus

$$\bar{\tau}_1^{(1)} = (1 - P_{fa}) \frac{(1-ab)}{(1-a)(1-b)} + P_{fa} \frac{(1-cd)}{(1-c)(1-d)} \tag{59}$$

Similarly, $\bar{\tau}_2^{(1)} = \bar{\tau}_1^{(1)}$. The mean waiting time of the semi-Markov chain in class 1 is:

$$\bar{\tau}^{(1)} = \sum_{i \in E_1} \underline{\pi}_{M_i}^{(1)} \bar{\tau}_i^{(1)} = \bar{\tau}_1^{(1)} \tag{60}$$

Hence, for this situation (but not in general): $\underline{\pi} = \underline{\pi}_M$.

The time scale factor δ is set equal to ϵ for convenience. It should be noted that δ must be of the same order as ϵ, but not necessarily equal.

All parameters required to describe the enlarged Markov process have now been stated. The parameters of the approximate class–to–class interval transition CDF can be found as described in Theorem 4: $\gamma_{21} = 1$, $\Lambda_1 = q_{31}^{(1)}/a_1$. So, the class–to–class interval transition CDF expressed in the slow time scale is:

$$\leq \hat{\phi}_{12}(t') = 1 - \exp\left\{-\frac{\Lambda_1 t'}{T}\right\} \tag{61}$$

To return to the original time scale, let $t' = \delta t$, and recall that δ was chosen to be equal to ϵ in this case. The rows of the interval transition probability matrix of the enlarged process must sum to unity. Since the semi-Markov chain is always in state 1 at the initial time, the enlarged process is always in class 1 at the initial time. Hence, approximate class occupancy probabilities can be stated directly from the first row of the interval transition probability matrix since $\hat{\underline{\pi}}^e(t) = \underline{\pi}^e(0) \leq \hat{\underline{\Phi}}(t)$:

$$\hat{\underline{\pi}}^e(t) = \left[\exp\left\{-\frac{\epsilon\Lambda_1 t}{T}\right\} \mid 1 - \exp\left\{-\frac{\epsilon\Lambda_1 t}{T}\right\}\right] \tag{62}$$

By expanding the approximate Markov process in terms of the stationary probabilities of the unperturbed semi-Markov chain as in (48), approximate expressions for the state occupancy probabilities of the original process can be stated as follows:

$$\hat{\underline{\pi}}^e(t) = \left[(1 - P_{fa})\exp\left\{-\frac{\epsilon\Lambda_1 t}{T}\right\} P_{fa}\exp\left\{-\frac{\epsilon\Lambda_1 t}{T}\right\} \middle| 1 - \exp\left\{-\frac{\epsilon\Lambda_1 t}{T}\right\}\right] \tag{63}$$

The approximate expressions above will be compared to the analytical expressions derived using z-transform techniques.

5.2 Discussion of Results for SCMS-I

This section examines sources of error associated with the approximate technique for a specific set of system parameters: a=0.95, b=0.94, c=0.89, d=0.88 and P_{fa}=0.05. This set of parameters implies a time to detection in the absence of a failure of 16 time steps (3.2 seconds), and a time to a nominal decision in the absence of a failure of 36 time steps (7.2 seconds) for a sample period of 200 milliseconds.

The relative error (in percent), $\Delta_i = | \pi_i(n) - \hat{\pi}_i(n) | / \pi_i(n)$ will be used to compare the approximate and the analytical state occupancy probabilities.

The approximate state probability time histories, $\hat{\underline{\pi}}(n)$, are compared to those obtained analytically, $\underline{\pi}(n)$, in Figure 2 for each of the three states. These results are for ϵ=0.00005, implying an $MTBF$ of 20,000 time steps (4000 seconds or just over an hour). In this figure, the state probabilities are propagated for a period of one component $MTBF$. Time is normalized by the $MTBF$.

The largest error occurs early, especially in the first class. This is due to the fact that the normalized state probabilities in class 1 have not converged to the class 1 stationary probabilities of the unperturbed semi-Markov chain. For example, at the tenth time step the normalized probabilities in class 1 are

$$\underline{\pi}_N^{(1)}(10) = [0.9817,\ \ 0.0183]\,. \tag{64}$$

These differ substantially from the class 1 stationary probabilities of the unperturbed semi-Markov chain:

$$\underline{\pi}^{(1)} = [0.9500,\ \ 0.0500]\,. \tag{65}$$

The approximate method accurately estimates the state probabilities when the normalized probabilities have converged to the stationary probabilities in each class. This occurs as early as time step 200, and the relative errors for states 1 and 2 have dropped to $\Delta_1 = \Delta_2 = 8.62 \times 10^{-4}\%$, which indicates that the estimate is closely tracking the exact solution. Until time step 200, use of the approximate method is not valid resulting in large relative errors in the state probabilities.

Another source of error is due to non-zero value of ϵ since Theorem 4 describes $\overset{\leq}{\Phi}(t)$ in the limit as $\epsilon \rightarrow 0$. Obviously, the ϵ chosen in Figure 2 was "small enough" because the state probabilities were estimated adequately. Figure 3 examines the class 2 (or state 3) probability at 100%, 50% and 25% of an MTBF for a range of values of ϵ. The relative error decreases markedly with decreasing ϵ for all three choices of mission time. For large ϵ, ($\epsilon > .01$), the "slow" time scale represented by failure events and the "fast" time scale represented by fault monitoring events are nearly indistinguishable from each other resulting in poor estimates of the state probabilities. In contrast, for small ϵ, ($\epsilon < .001$) the two time scales are distinct. For ϵ=0.00005, the time to a decision is about 36 seconds and the $MTBF$ is 4000 seconds, or, the "slow" time scale is approximately 100 times slower than

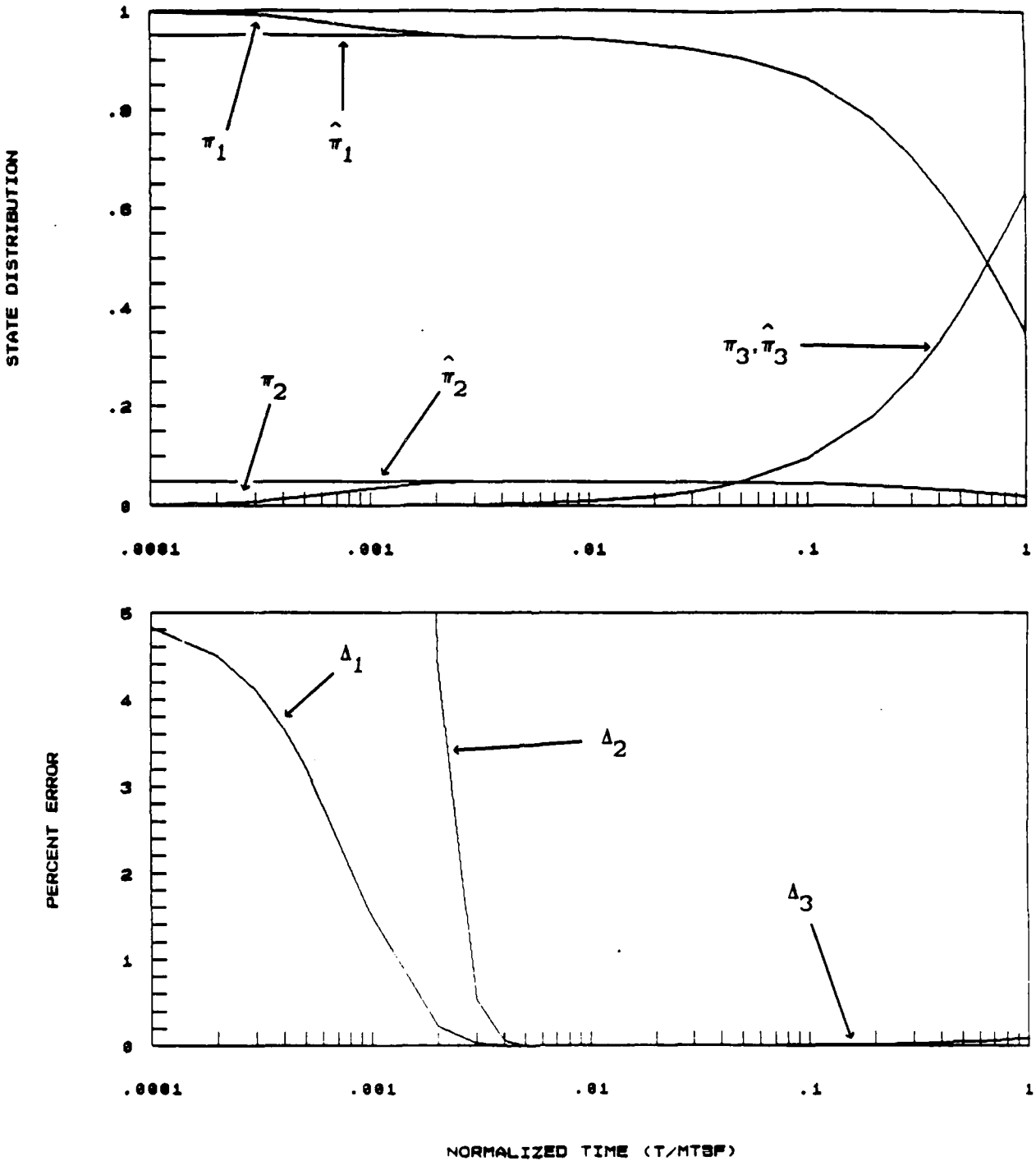


Figure 2: State probability time histories for SCMS-I. (a) Analytical and approximate solutions, (b) Relative error

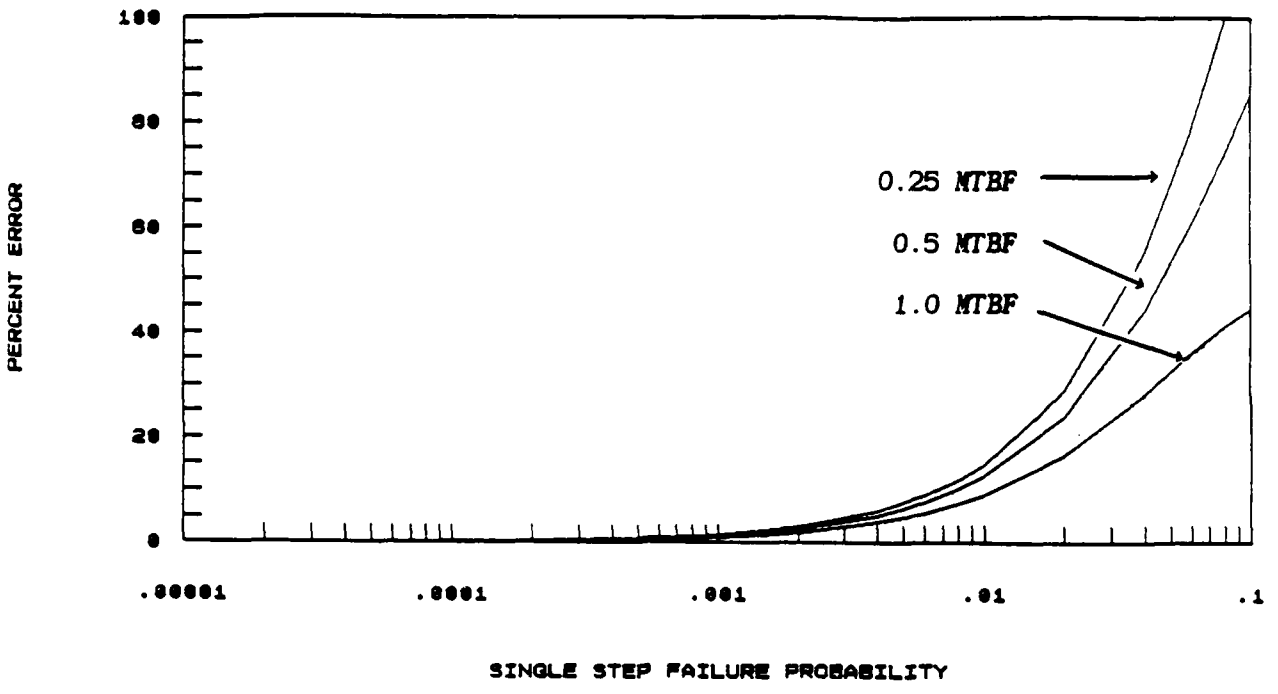


Figure 3: Sensitivity to ϵ for SCMS-I. The relative error is plotted versus the single-step probability, ϵ, for mission times of one $MTBF$, $0.5 \times MTBF$, and $0.25 \times MTBF$.

the fast time scale. Therefore, to obtain accurate estimates of the state probabilities, it is imperative that the fast and slow time scales be distinctly separated in terms of their mean holding times. A possible rule of thumb is suggested by these results for determining whether the time scales are distinct. That is, compute the holding time of the slowest FDI event. For the approximation to be valid, the $MTBF$ of the fastest failure should be at least 100 times longer than this calculated FDI holding time.

The analytical and approximate solutions of the class 2 probability can also be compared by expanding each in a Taylor series about $\epsilon = 0$. If the two are the same to first order in ϵ then the estimate is a first order perturbation solution. If they differ, this would suggest that an alternative estimate could be derived. Expanding $\pi_2^\epsilon(n)$ and $\hat{\pi}_2^\epsilon(n)$ Taylor series about ϵ=0:

$$\pi_2^\epsilon(n) = n\epsilon - \frac{1}{2}(n^2 - n)\epsilon^2 + O(\epsilon^2) \tag{66}$$

$$\hat{\pi}_2^\epsilon(n) = n\epsilon - \frac{1}{2}\left\{n^2 - 2n\left[\frac{\partial}{\partial\epsilon}\Lambda_1(\epsilon)\Big|_{\epsilon=0}\right\}\epsilon^2 + O(\epsilon^2)\right. \tag{67}$$

To first order in ϵ:

$$\pi_2^\epsilon(n) = \hat{\pi}_2^\epsilon(n) = n\epsilon + O(\epsilon) \tag{68}$$

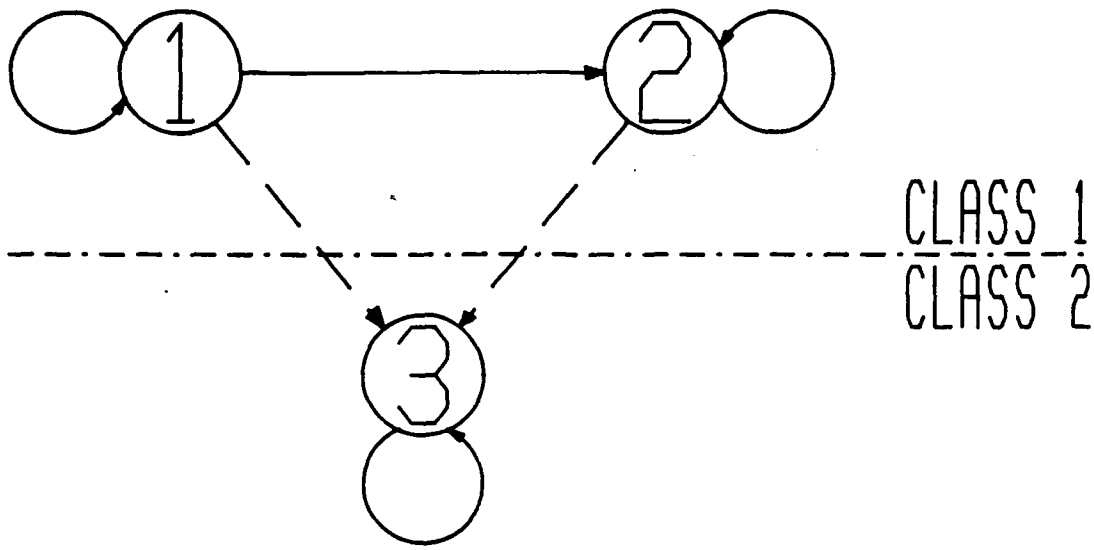


Figure 4: Semi-Markov transition diagram for SCMS-II

So, the approximation developed in Theorem 4 produces a first order perturbation solution in ϵ for this model. Therefore, the error between the analytical and approximate class 2 probabilities begins with the order ϵ^2 terms. Note that the dominant second order term $(n^2\epsilon^2)$ is also the same. It can be shown [13] that the error is due to a difference in a second order term with a small coefficient, namely a term that is proportional to elapsed time. Although this observation is strongly model dependent, it may also be true for other models as well.

5.3 The SCMS with abbreviated monitoring

A second method of fault monitoring is to deploy a sequential test that monitors the status of a component until a failure is indicated, at which point the sequential test is discontinued. An SCMS of this type will be denoted by SCMS–II.

The states for the semi-Markov model of the SCMS-II are enumerated in Table 1. The semi-Markov transition diagram of the SCMS-II is depicted in Figure 4. The class decomposition of the SCMS-II is similar to SCMS-I. However, in this case, the embedded Markov chain in class 1 is non-ergodic.

The direct form of the core matrix sequence can be developed in the same manner as for

the SCMS-I. A notable difference is in the transition probabilities out of state 2. Because the fault monitoring test is discontinued upon a failure indication, only failure events cause such transitions. A reset of state 2 occurs when no failure occurs. A transition from state 2 to state 3 occurs only if a failure takes place. Assuming geometrically distributed failures, the core matrix can be stated:

$$\mathbf{G}^{\epsilon}(n) = \begin{bmatrix} (1-\epsilon)^n f_D^0(n) & (1-\epsilon)^n f_D^0(n) & \epsilon(1-\epsilon)^{n-1} Q_0(n) \\ 0 & (1-\epsilon)\delta(n-1) & \epsilon\,\delta(n-1) \\ 0 & 0 & \delta(n-1) \end{bmatrix} \tag{69}$$

As for the SCMS-I, $\underline{\pi}(n)$ can be found using z-transforms. The state probability time histories could not be obtained, however, because the partial fraction expansions could only be done numerically. These results are decribed fully in Appendix B of [13]. However, the class probabilities were found and are stated below:

$$\pi^{\epsilon}(n) = [(1-\epsilon)^n \ , \ 1-(1-\epsilon)^n] \tag{70}$$

Again, these analytical expressions for $\underline{\pi}(n)$ and $\hat{\underline{\pi}}^{\epsilon}(n)$ will be compared to the approximate results derived using the approximate technique in the next section.

To generate the approximate solutions, the core matrix must be placed in standard form. However, all of the required quantities are known based on the manipulations performed for the SCMS-I. The eventual transition probability matrix of the unperturbed semi-Markov chain is obtained by setting $\epsilon = 0$ and ignoring the holding time PMFs:

$$\mathbf{P} = \begin{bmatrix} 1-P_{fa} & P_{fa} & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{71}$$

By raising this matrix to successively higher powers, the stationary interval transition probability matrix can be found, and the embedded stationary probability distribution in partitioned form is:

$$\underline{\pi}_M = [0 \ \ 1 \ \mid \ 1] \tag{72}$$

Because of the model structure, it is clear that the stationary probabilities for each class of the unperturbed semi–Markov chain are: $\underline{\pi} = \underline{\pi}_M$. For this analysis, the time scale factor

δ is again set equal to ϵ. Finally, $\gamma_{21} = 1$, and $\Lambda_1 = 1$, so that the approximate expressions for the class probabilities can be found:

$$\hat{\pi}^e(t) = \left[\exp(-\frac{t}{T}) \ , \ 1 - \exp(-\frac{t}{T})\right]. \tag{73}$$

By expanding the enlarged Markov process in terms of the stationary probabilities of the unperturbed semi-Markov chain, approximate expressions for the state occupancy probabilities of the original process can be stated:

$$\hat{\pi}(t) \approx \left[0 \ \ \exp(-\frac{t}{T}) \ , \ 1 - \exp(-\frac{t}{T})\right] \tag{74}$$

5.4 Discussion of results for SCMS-II

The approximate state probability time histories, $\hat{\underline{\pi}}(n)$, are compared to those obtained analytically, $\underline{\pi}(n)$, in Figure 5 for each of the three states. These results are for the same parameter set as SCMS-I. The largest absolute errors occur in estimating state 1 and do not attenuate until 50% of an $MTBF$ has passed. The approximation estimates the state 1 probability to be zero because the class 1 embedded Markov chain is non-ergodic and yields zero for the stationary state 1 probability. The estimated state probabilities in states 2 and 3 are very accurate with relative errors of less than 0.01% for all time steps.

The relative error in state 1 is 100% at all times because the normalized probabilities in class 1 cannot converge to the stationary probabilities of the unperturbed semi-Markov chain. This is because the state 1 probability will never be exactly zero. For example, at the tenth time step in class 1 the normalized state probabilities are

$$\underline{\pi}_N^{(1)}(10) = [0.981175 \ , \ 0.014111], \tag{75}$$

and the unperturbed stationary probabilities are:

$$\underline{\pi}^{(1)} = [0 \ , \ 1]. \tag{76}$$

The approximate method requires that the normalized probabilities converge to the stationary probabilities for each class in order to obtain accurate state probability estimates.

The other source of error is due to non-zero ϵ. In Figure 5, the value of ϵ was small enough to provide accurate results because the state 2 and 3 probabilities were estimated

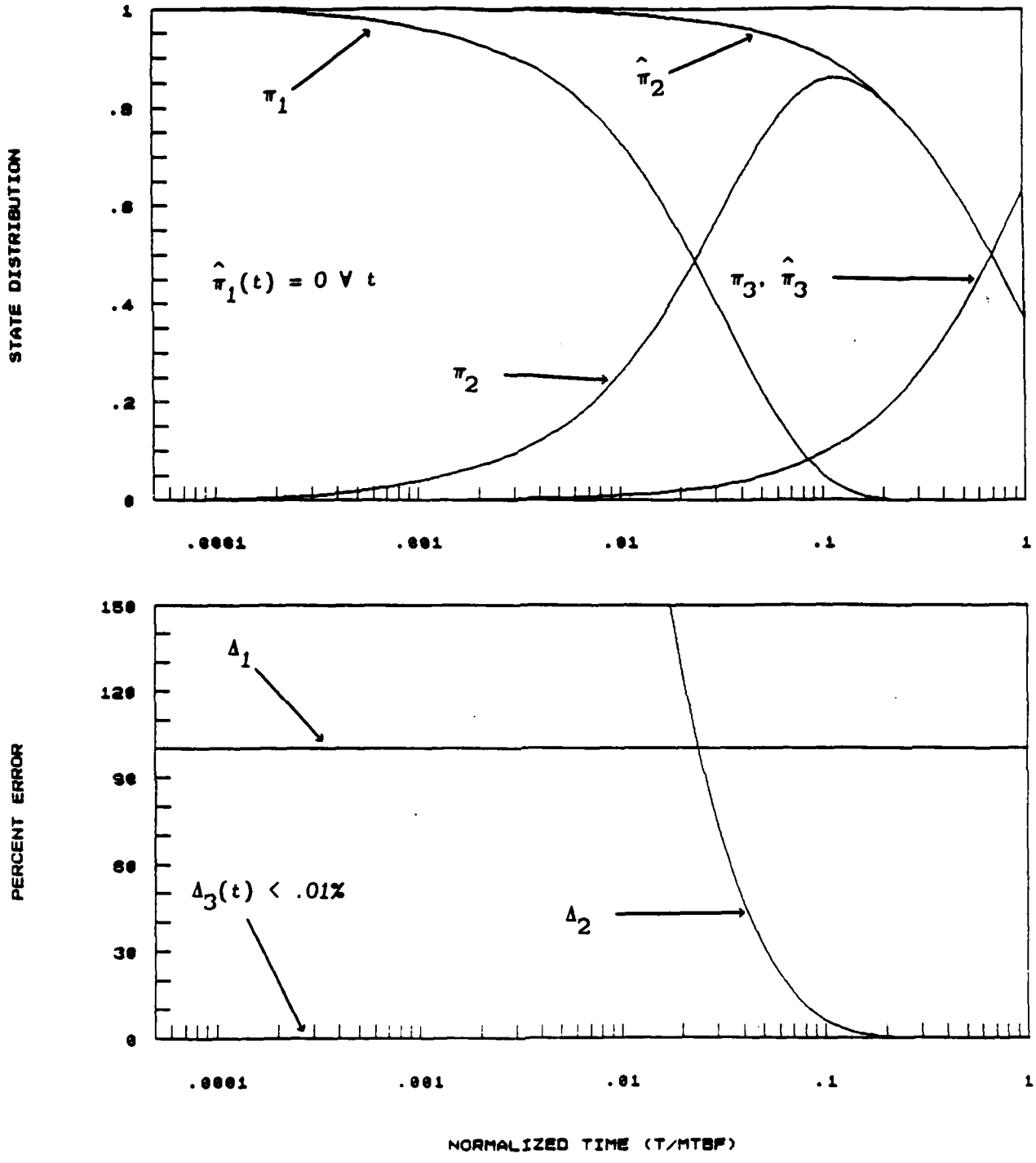


Figure 5: State probability time histories for SCMS-II (a) Analytical and approximate solutions. (b) Relative error.

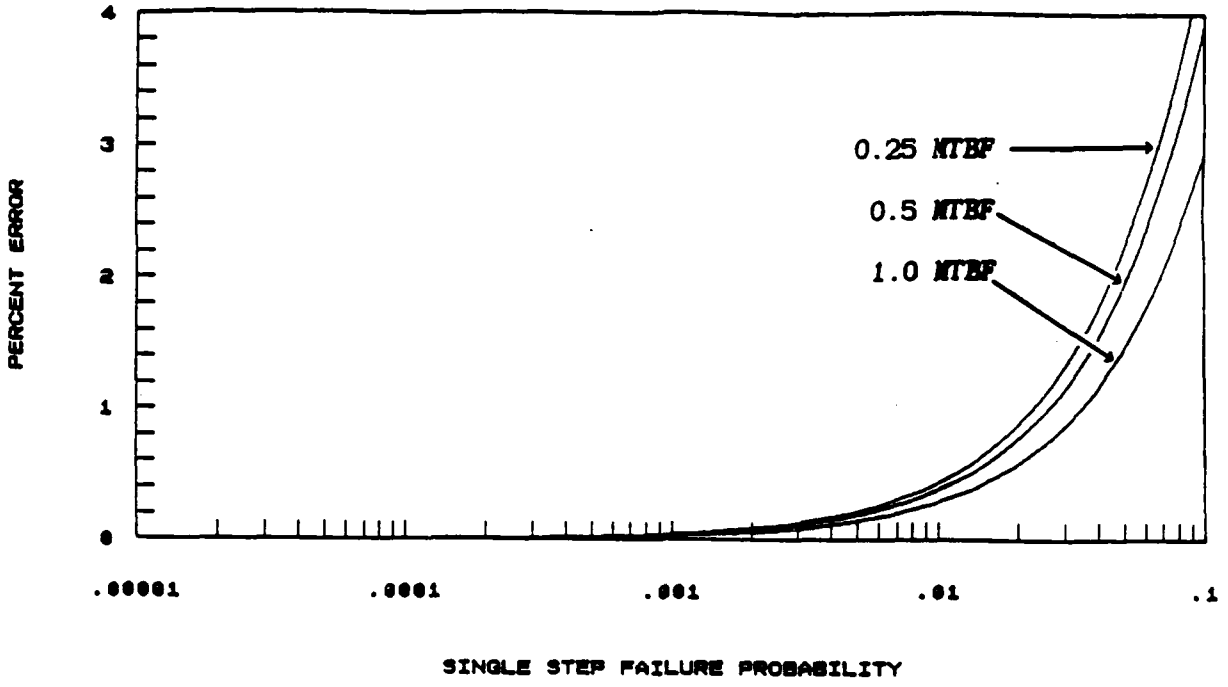


Figure 6: Sensitivity to ϵ for SCMS-II. The relative error is plotted versus the single-step probability, ϵ, for mission times of 1 $MTBF$, $0.5 \times MTBF$, and $0.25 \times MTBF$.

adequately. Figure 6 presents the class 2 (or state 3) occupancy probability for mission times of 100%, 50% and 25% of an $MTBF$ for a range of values of ϵ corresponding to a component $MTBF$ ranging from 4 seconds to 5555 hours. As was the case for the SCMS-I, the relative error decreases markedly with decreasing ϵ for the three choices of mission time. This reiterates the observation that the fast and slow time scales must be distinct in terms of their mean holding times in order to obtain accurate estimates of the state probabilities. This analysis also demonstrates the usefulness of the rule of thumb suggested earlier.

The Taylor series expansions for the analytical and approximate class 2 probability will again be compared. Expanding the class 2 occupancy probability in a Taylor series about $\epsilon = 0$ yields

$$\pi_2^\epsilon(n) = n\epsilon - \frac{1}{2}(n^2 - n)\epsilon^2 + O(\epsilon^2) \quad (77)$$

$$\hat{\pi}_2^\epsilon(n) = n\epsilon - \frac{1}{2}n^2\epsilon^2 + O(\epsilon^2) \quad (78)$$

To first order, $\pi_2^\epsilon(n)$ and $\hat{\pi}_2^\epsilon(n)$ are identical. This proves that the approximate method produces a first order perturbation solution in ϵ for this model. The two expressions begin to differ starting with the ϵ^2 terms, but the dominant second order term $(n^2\epsilon^2)$ is the same.

Hence, the error can be expressed as:

$$\tilde{\pi}_2^{\epsilon}(n) = \frac{1}{2} n\epsilon^2 + O(\epsilon^2) \tag{79}$$

which is second order in ϵ and proportional to time, which emphasizes the asymptotic nature of the approximation. Again, this observation is model dependent. However, the same behavior was found for the SCMS-I.

6 The SCDR System Model

The single-component dual-redundant (SCDR) system consists of two identical components, a primary and a backup, operating in parallel. An independent sequential test monitors the status of each component. The reliability of this system was evaluated using the approximate technique in [13]. However, in the interest of brevity and clarity, the interested reader is referred to [13].

7 Conclusions

A primary contribution of this work is the extension of Korolyuk's limit theorem for semi–Markov processes to semi– Markov chains in Theorem 1, which describes the conditions under which a perturbed semi–Markov chain can be approximated by an enlarged Markov process. Moreover, Theorem 1 describes how the parameters of the enlarged Markov process are derived from the parameters of the semi–Markov chain.

Two problems arise in applying Theorem 1 to fault tolerant control system (FTCS) models. First, the non-perturbed embedded Markov chains in each class are usually non-ergodic. This was required in Theorem 1, but was relaxed to the existence of the Caesaro limit probabilities in Lemma 2. These were found to exist in Lemma 3 if the embedded Markov chain was either ergodic, or non-ergodic with one and only one unity eigenvalue.

Second, the transition PMFs are typically not functions of the perturbation parameter ϵ. This problem was mitigated by introducing the concept of time scaling in Theorem 4. The form of the transition PMFs was generalized to include those common to FTCS reliability models. This generalization included a dependence on a time scaling factor δ and on a

small parameter ϵ that determined the state space partitioning of the original semi-Markov chain.

Use of the approximate technique was demonstrated by two simple examples. Accurate estimates of the state probabilities were determined for situations where ϵ was "small enough" and where the normalized probabilities in each class had converged to the stationary probabilities of the non-perturbed semi-Markov chain. In the two examples presented, the approximate technique yielded a first order perturbation solution in ϵ to the analytically obtained class probabilities.

The approximation error was found to be insignificant if the slow and fast time scales were distinct. Finally, a rule of thumb was suggested by the error analysis: the slow and fast time scales are distinct if the MTBF of the fastest failure is 1000 times longer than the mean decision time of the slowest FDI event.

8 Acknowledgments

This research was wholly supported by the U.S. Air Force Office of Scientific Research under grant AFOSR-84-0160.

References

[1] B. Walker, N. Wereley, R. Luppold, and E. Gai, "Effects of redundancy management on reliability modeling," 1988. Submitted.

[2] E. Gai, J. Harrison, and R. Luppold, "Reliability analysis of a dual redundant engine controller," in *Proceedings of the SAE Aerospace Congress and Exposition*, 1982.

[3] B. Walker, *A Semi-Markov Approach to Quantifying Fault Tolerant System Performance*. PhD thesis, Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, 1980.

[4] R. Howard, *Dynamic Probalistic Systems, Volume 2: Semi-Markov and Decision Processes*. Wiley and Sons, 1971.

[5] E. Lewis and F. Bohm, "Monte Carlo simulation of Markov unreliability models," *Nuclear Engineering and Design*, vol. 77, no. 1, pp. 49–62, 1984.

[6] T. Zhuguo and E. Lewis, "Component dependency models in Markov Monte Carlo simulation," *Reliability Engineering*, vol. 13, no. 1, pp. 49–62, 1985.

[7] K. Trivedi and J. Dugan, "Hybrid reliability modeling of fault-tolerant computer systems," *Computer and Electrical Engineering*, vol. 11, no. 2/3, pp. 87–108, 1984.

[8] K. Trivedi, J. Dugan, R. Geist, and M. Smotherman, "Modeling imperfect coverage in fault tolerant systems," in *Proceedings of the Fourteenth International Conference of Fault-Tolerant Computing*, pp. 77– 82, 1984.

[9] V. Korolyuk, L. Polishchuk, and A. Tomusyak, "A limit theorem for semi-Markov processes," *Cybernetics*, vol. 5, no. 4, pp. 524–526, 1969.

[10] G. Korn and T. Korn, *Mathematical Handbook for Scientists and Engineers*. McGraw-Hill, 1968.

[11] J. Kemeny and J. Snell, *Finite Markov Chains*. Springer-Verlag, 1976.

[12] S. Chu, *Approximate behavior of generalized Markovian models of fault tolerant systems*. Master's thesis, Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, 1986.

[13] N. Wereley, *An approximate method for evaluating generalized Markov chain reliability models of fault tolerant systems*. Master's thesis, Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, 1987.