Reliability Engineering and System Safety 68 (2000) 159–168

RELIABILITY ENGINEERING & SYSTEM SAFETY

www.elsevier.com/locate/ress

Reliability modeling of a hard real-time system using the path-space approach

Hagbae Kim

Department of Electrical and Computer Engineering, Yonsei University, 134 Sinchon-dong Sudaemoon-ku, Seoul 120-749, South Korea

Received 2 August 1999; accepted 31 January 2000

Abstract

A hard real-time system, such as a fly-by-wire system, fails catastrophically (e.g. losing stability) if its control inputs are not updated by its digital controller computer within a certain timing constraint called the hard deadline. To assess and validate those systems' reliabilities by using a semi-Markov model that explicitly contains the deadline information, we propose a path-space approach deriving the upper and lower bounds of the probability of system failure. These bounds are derived by using only simple parameters, and they are especially suitable for highly reliable systems which should recover quickly. Analytical bounds are derived for both exponential and Weibull failure distributions encountered commonly, which have proven effective through numerical examples, while considering three repair strategies: repair-as-good-as-new, repair-as-good-as-old, and repair-better-than-old. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: System reliability; Real-time system; Hard deadline; Path-space approach; Semi-Markov process

1. Introduction

A "hard" real-time system is characterized by a stringent timing requirement, which should be met to avoid any catastrophe [1]. This timing information must, therefore, be accounted for when the reliability of a hard real-time system is modeled or measured. By embedding this information, reliability modeling of those systems can also handle temporary malfunctions caused, for example, by ElectroMagnetic Interference (EMI) [2]. One class of examples is a real-time control system where the dynamics of the controlled plant/process (robots, nuclear reactors, or paper mills) keep the plant within the safe region if the controller malfunctions do not persist too long. In real-time control systems such as aircraft or satellites, the system should be directed by an appropriate controller computer in a timely manner; that is, its control input should be updated by the controller computer within a time limit called the *hard deadline* [3]. For safety-critical applications this property has led to highly redundant/reconfigurable controllers.

In some conventional reliability models for digital control systems, temporal periods of controller misbehaviors were ignored while assuming that a (perfect) controller should always be failure-free to manage the underlying controlled plant. Other models have captured the details of such systems by focusing only on the states of fault-tolerant controller computers treating a temporary controller failure as a total system failure regardless of the requirements of the controlled plant [4,5]. That is, they erred on the safe side by ignoring the "system inertia" or system resilience in tolerating temporary loss of the controller. However, it is possible for a system/plant to survive repeated controller perturbations temporarily because of plant dynamics and inertia.

In contrast, in the paper we deal primarily with a system failure resulting from temporal controller upsets in consideration of system inertia specified by the deadline information. In other words, the system failure is caused due to slow recoveries of controller misbehaviors taking more than the hard deadline that intrinsically depends upon the plant dynamics [3,6], where neither of the inter-arrival time of controller failures nor the recovery time is always exponentially distributed and the failure rate is substantially affected by the holding time in the state of controller failure(s). Note that the failure rate is also dependent on the "global time" (the total operating time of the system) in more general systems.

There were also some previous works that considered the deadline/timing information for reliability modeling. In Ref. [7], a Markov model, which does not only describe component-failure behaviors but also incorporates deadline violations as simple transitions, was used to measure system reliability by deriving only the probability of missing a deadline, while needing another computation in a different

E-mail address: hbkim@bubble.yonsei.ac.kr (H. Kim).

0951-8320/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0951-8320(00)00014-4

'lower-level' model. The authors of Ref. [8] considered non-failure-critical cases, where some system-down time can be tolerated if it is recovered within a certain deadline. They derived the mean value of the system lifetime and the cumulative operational time for the case of bounded repair time (restricted by the deadline). However, it is difficult to derive the distribution from the Laplace–Stieltjes transform of the system lifetime, although it is easy to compute the mean value. Hence, it is intractable to derive system reliability using these results. Moreover, none of these considered such general cases as when the time-to-failure and/or time-for-repair are not exponentially distributed. Although these general cases were modeled by a time-non-homogeneous Markov chain [9], a semi-Markov process [10] or a Markov regenerative process [11], none of these dealt with the case when the failure rate depends on the total operation time of the system. These general models can be computed by the Monte Carlo method, but, since the Monte Carlo method is just a statistical estimation through numerical simulation, it is computationally very expensive.

To overcome these obstacles, we consider a *path-space* approach which was not only treated in queuing theory [12] but also proven useful in solving other reliability modeling problems [13]. Our goal is to derive tight upper and lower bounds for the probability of a system failure in terms of two simple parameters; (i) the probability of k ($k > 0$) interruptions during the operating period T—for instance this can be estimated by using field data or a certain analytic model can be built like our previous work [2] evaluating the susceptibility of controller computers against EMI inducing upsets—and (ii) the probability of successful recovery (before the hard deadline) given an interruption. For the first parameter, computing the probability of k events during a time period is straightforward, and there are analytical formulas for some of the more popular probability distributions [14]. For the second parameter, using the probability of successful recovery has three advantages; (i) it is mathematically more tractable than the density function for the recovery time that is required by the Chapman–Kolmogorov equations, (ii) it is experimentally and statistically less demanding to obtain the binomial parameters of failures than to do curve fitting for a density function, and (iii) it permits model reduction because it reduces encountered complexity, that is, multi-state recovery models to a single state with jump probabilities to successful recovery or unsuccessful recovery. Despite all of these simplifications, it is shown by proper examples that this approach yields tight bounds for a wide variety of models. It is especially suitable for the stiff models of highly reliable systems.

The recovery/repair procedure begins at the start of an interruption. There can be a time lag between the occurrence of an interruption and the beginning of the actual repair, but this time lag is included in the repair-time distribution. Henceforth, in the models recovery begins when the system enters a "down" state, often called the *recovery/repair* state. The probability distribution for the recovery time is also fixed for a given model. That is, it is assumed that recovery is either an automated procedure or done by a repair crew that does not become either more proficient or fatigued. These properties of the recovery/repair procedures imply that the time to recovery depends only on the time since entering the recovery/repair state. Hence, recovery/repair is captured by a semi-Markov process for all the models described below, even if the distribution for system malfunctions is dependent on the global time.

2. Semi-Markov model for hard deadlines

This section first discusses the assumptions needed for a semi-Markov formulation in reliability modeling of a hard real-time system, and presents a semi-Markov model itself (embedded with the deadline information) and its Chapman–Kolmogorov equations to derive state probabilities. It then describes a path-space approach and derives the upper and lower bounds for the probability of system failure due to a lengthy interruption.

2.1. The semi-Markov formulation

We begin with a semi-Markov model. (A fixed deadline can be modeled as a semi-Markov transition with zero variance.) Later sections extend the results to the models with global-time dependencies. The assumptions for the model are as follows:

- recovery is as-good-as-new;
- the recovery distribution depends on elapsed time since malfunction;
- the deadline is a some fixed time;
- all of the processes (malfunctions, recoveries, and deadlines) are independent of one another.

Let $f(t)$, $g(t)$, and $d(t)$ be the density functions for the arrival of the malfunction, the density for recovery, and the density for the deadline, respectively. With the four assumptions above, the model is declared to be semi-Markov with three states as given in Fig. 1. The first assumption is appropriate in case of either perfect replacements or the repair of high-quality equipment (such as electronic components) which has a constant, or nearly constant, failure rate. Obviously, this assumption is the place to generalize the model in order to handle a wider variety of systems. This is done in later sections, but it requires global-time dependent models. The second assumption says that the repair procedure begins when a breakdown occurs, and that the repair-time distribution remains the same throughout the

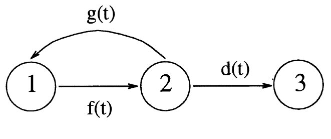

Fig. 1. The semi-Markov model for hard deadlines.