自恢复容错系统的建模与分析

郭成昊,刘凤玉
(南京理工大学计算机科学与技术系,南京 210094)

[摘要] 容错系统不仅会产生硬件故障,由于连续长时间的运行,系统的性能也会逐渐下降或失效,即老化现象。综合考虑容错系统中的硬件故障和老化现象,提出了将传统的冗余技术和软件抗衰技术相结合的策略,并给出了该系统的非马尔可夫随机 Petri 网模型,随之对基于该模型的系统进行了定量分析。

[关键词] 容错系统;软件抗衰;软件老化;冗余策略;非马尔可夫随机 Petri 网
[中图分类号] TP302 [文献标识码] A [文章编号] 1009 - 1742(2007)10 - 0075 - 05

1 引言

在容错系统中,人们一般只注意到系统硬件部分的错误,硬件可靠性研究已有成熟的模型和分析方法[1]。而大量实验表明,由于异常的软件行为导致系统性能崩溃的可能性要远远高于硬件故障对系统产生的影响。这就需要在容错系统中考虑软件的可靠性和可用性。因此软件可靠性已经成为制约容错系统可靠性的一个重要因素。

一般来说,投入运行的软件系统由于种种原因(设计阶段考虑的不周全,环境因素等)都会存在一些缺陷,当这些应用软件运行一段时间以后,其中存在的缺陷或客户的不当操作会使系统性能下降,如响应时间或负载的增加。在性能持续衰退期间,如果不采取适当的干预,最终会导致整个系统的崩溃。这种现象叫做软件衰老(software aging),它目前已在许多广泛使用的系统中出现,如 Netscape, Patriot 系统[2],目前,容错系统由于其引发系统崩溃因素的多样性,而越来越受到关注。

Huang 等提出了自愈(rejuvenation)技术来处理软件衰老对系统造成的影响[3]。该方法是一种预防性的软件容错策略,它包括周期性停止系统运行,清理内部状态,重新启动。研究软件衰老的方法主要有基于检测和基于模型的方法。基于检测的方法是通过对系统运行时性能参数的样本值进行统计分析来预报软件衰老的时间及其对系统造成的影响[4]。基于模型的方法主要是通过建立系统的数学模型,分析它的不同状态来评价软件衰老的程度和自愈性能[5]。

首先分析了出现衰老现象的容错系统状态,接着利用非马尔可夫随机 Petri 网(NMSPNs)来对系统进行建模,并基于马尔可夫再生理论对该模型进行分析求解,定量地评价系统性能。

2 自愈系统状态分析

Huang 等提出了一个基于模型的方法来分析自愈系统的性能[3],模型如图 1 所示。

图 1 中箭头表示自愈系统的各个状态之间转变的方向,圆圈表示自愈系统的 4 个状态:

状态 1 系统正常状态;
状态 2 系统可能出错状态;
状态 3 系统出错状态;
状态 4 系统自愈状态。

系统在开始运行时处于状态 1,运行一段时间以后,系统进入状态 2,该状态表明系统仍然能够提供服务,但出现错误的概率不为零。在进入状态 2

[收稿日期] 2006 - 04 - 21; 修回日期 2006 - 07 - 08
[基金项目] 国家自然科学基金资助项目(60273035)
[作者简介] 郭成昊(1981 -),男,江苏南京市人,南京理工大学博士研究生,主要研究领域为软件自愈与抗衰体系结构

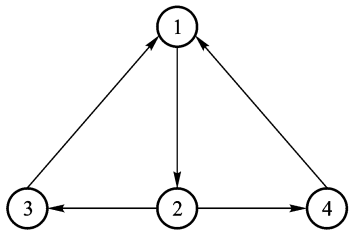

图 **1**　自愈系统的状态图

Fig. 1　The state graph of rejuvenation system

后,如果系统进入状态 3,软件失效,花费一定的时间和成本修复后,软件重新回到状态 1。否则进入状态 4,执行自愈操作后,系统将回到状态 1。自愈系统的状态变化周而复始地进行下去。

3　出现衰老现象容错系统状态分析

容错系统通过冗余容错策略来处理由于硬件问题造成整个系统的崩溃。

在容错系统中导致系统崩溃的原因除了硬件问题外还包括软件问题,在硬件出错之前软件的问题也可能使容错系统的性能衰退甚至使整个系统失效。系统的状态变化如图 2 所示。

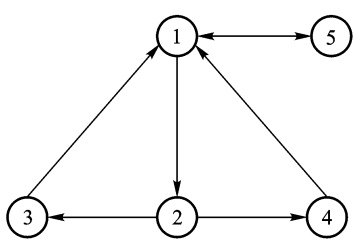

图 **2**　容错系统的状态图

Fig. 2　The state graph of fault-tolerant system

图 2 中箭头表示自愈系统的各个状态之间转变的方向,圆圈表示容错系统的 5 个状态:

状态 1　系统正常状态;

状态 2　系统可能出错状态;

状态 3　系统出错状态;

状态 4　系统自愈状态;

状态 5　系统硬件故障状态。

在系统进入状态 2 后,如果系统进入状态 5,则表示出现了硬件故障,马上执行冗余容错策略,回到状态 1,否则系统的状态变化如前所述。

4　容错系统非马尔可夫随机 **Petri** 网建模和分析

4.1　非马尔可夫随机 **Petri** 网(**NMSPNs**)

随机 Petri 网(SPNs, stochastic Petri nets)是对复杂系统进行建模的一种强有效的形式化工具[6]。它能够形象地描述系统的各种特性(如并行特性、分布式特性、异步特性、自适应性等),从而很好地刻画系统的动态行为、分析系统的性能。

在 SPN 中,变迁的实施联系一个服从指数分布的随机变量,这样模型就可以转化为具有马尔可夫特性的随机过程进行分析。但是在自恢复容错系统中要进行定期的自愈操作,这就导致了模型中变迁的实施时间不服从指数分布,从而使得研究非马尔可夫模型成为当务之急。

一个 NMSPNs 包含 3 种类型的变迁:

- 瞬时变迁;
- 实施时间服从指数分布的变迁;
- 实施时间服从一般分布的变迁。

一个 NMSPNs 具有如下定义:

$$\Sigma = (P, T, F, \sigma),$$

其中 (P, T, F) 是一个网,P 元素是位置,T 元素是变迁,$F \subseteq (P \times T) \cup (T \times P)$ 是弧的集合;σ 表示网中的时间元素,服从一般分布。

4.2　非马尔可夫随机 **Petri** 网分析方法

利用马尔可夫再生理论(Markov regenerative theory)[7]对非马尔可夫随机 Petri 网进行分析。

马尔可夫再生理论的基本思想是在系统到达某一特殊状态 M_n 时对系统进行采样,得到的点称为再生点,在其上系统的行为表现出无记忆性,从而形成离散时间马尔可夫链(DTMC, discrete-time Markov chain)这样简化了问题的复杂性,使分析 NMSPNs 的状态变化成为可能。

考虑一个 NMSPNs 模型,对应的随机过程记为 $M(t) = \{M_t, t \geqslant 0\}$,其中 M_t 表示 NMSPN 在 t 时刻所处的可达状态,状态空间为 Ω。选取再生点 τ_n,选取不存在可实施的第三类变迁的状态作为再生状态 v_n,由 v_n 构成的状态集 $\Omega' = \{v_n : (\tau_n, M_n)\} \subseteq \Omega$ 就构成了离散时间马尔可夫链。随机过程 $M(t) = \{M_t, t \geqslant 0\}$ 称为马尔可夫再生过程(MRGP, Markov regenerative process)。

$M(t) = \{M_t, t \geqslant 0\}$ 的状态转移概率为 $V_{ij}(t) = p(M(t) = j \mid M(\tau_0) = i)$;它是系统性能分析的基

础。$K(t)$表示再生点τ_n上系统的行为，

$K_{ij}(t)=p(M(\tau_1)=j,\tau_1\leqslant t|M(\tau_0)=i)$。

$E(t)$表示相邻两再生点之间系统的行为，

$E_{ij}(t)=p(M(t)=j,\tau_1\leqslant t|M(\tau_0)=i)$。

从以上定义得出[8]：

$$\sum_{j\in\Omega'}K_{ij}(t)+\sum_{j\in\Omega}E_{ij}(t)=1 \tag{1}$$

$$\boldsymbol{V}(t)=\boldsymbol{E}(t)+\boldsymbol{K}\boldsymbol{V}(t) \tag{2}$$

$$\pi_i=\sum_{k\in\Omega}v_k\alpha_{ki}\Big/\sum_{k\in\Omega}v_k\sum_{j\in\Omega}\alpha_{kj} \tag{3}$$

其中π_i为马尔可夫再生过程(MRGP)的稳定状态概率，v_k为离散时间马尔可夫链(DTMC)的稳定状态概率，且$\sum_{i\in\Omega'}v_i=1$，$\alpha_{ij}=\int_0^\infty E_{ij}(t)\,\mathrm{d}t$。

4.3 容错系统非马尔可夫随机 Petri 网模型

容错系统的 NMSPNs 模型如图 3 所示。

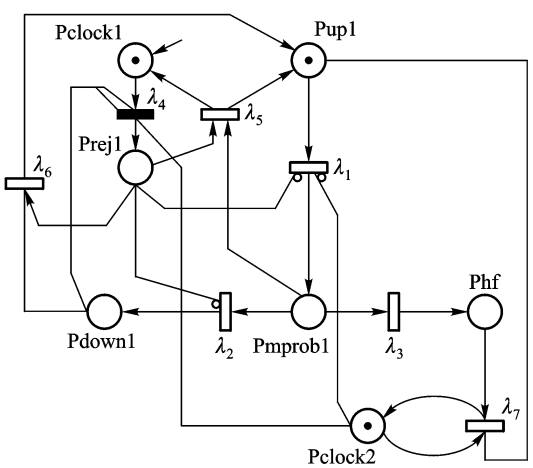

图 3 容错系统的 NMSPNs 模型

Fig. 3 NMSPNs model of fault - tolerant system

图 3 中，位置 Pup1 表示模块 1 处于正常工作状态，位置 Pclock1 表示模块 1 处于自愈准备状态，位置 Pclock2 表示模块 2 处于替代准备状态，位置 Pdown1 表示模块 1 处于失效状态。初始时刻，标记处于 Pup1 位置，当变迁λ_1触发，标记进入 Pmprob1 位置，表示模块 1 处于可能出错状态。当变迁λ_2触发，标记进入 Pdown1 位置，表示模块 1 失效，这时变迁λ_6触发，标记重新回到 Pup1。同时变迁λ_4禁止，表示当标记重新回到 Pup1 时，不执行自愈操作。模块 1 处于可能出错状态即标记到达 Pmprob1 位置时，如果执行自愈操作，则变迁λ_5触发，标记重新回到 Pup1。变迁λ_4用来模拟系统自愈周期，当系统处于自愈周期内，并且变迁λ_1没有触发，则变迁λ_4触发，位置 Pclock1 中的标记到达 Prej1 状态，变迁λ_1，λ_2，λ_6禁止，表示当执行自愈操作时，模块 1 中其他活动全部停止，随后变迁λ_5触发，标记重新回到 Pup1；当变迁λ_3触发，标记进入 Phf 位置，表示模块 1 出现硬件故障，当位置 Pclock2 中标记移动到变迁λ_7时，变迁λ_7触发，模块 2 替代模块 1 完成相应的任务，标记又回到 Pup1 位置，则整个系统重新回复正常状态。

4.4 容错系统的时间颜色 Petri 网模型分析

为了便于分析系统的性能，需要简化模型。考虑在系统进入 Phf 状态后，经过相应容错策略直接回到 Pup1，从而忽略模块 2 的一系列变迁状态。一个 6 元组 $E=$(Pup1, Pmprob1, Pdown1, Pclock1, Prej1, Phf)，其中每一个元素都是 NMSPNs 模型中相应的位置，且元素的取值为 1 或 0，如果一个标记在位置p_i中，$p_i=1$，否则为 0。这样得出一个 6 位 01 串，用来表示系统的变迁状态。状态变迁图如图 4 所示。

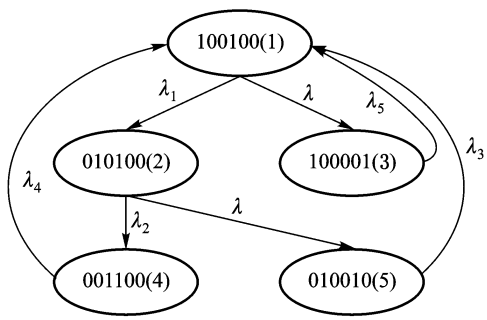

图 4 系统的状态变迁图

Fig. 4 State transformation graph of system

从图 4 中可以看出系统共包含 10 个状态，椭圆形表示系统的可达状态，从状态 i 到状态 j 的弧表示状态 i 到状态 j 可能的变迁。x_1, x_2, …, x_5 表示λ_1, λ_2, …, λ_5 的变迁平均实施速率，对应的变迁实施速率值如表 1 所示。

表 1 变迁平均实施速率参数选择

Table 1 Parameter of speed values of transition

变迁	λ_1	λ_2	λ_3	λ_4	λ_5
变迁平均实施速率	x_1	x_2	x_3	x_4	x_5
数值/h	0.1	0.2	5	6	2

由式(1)至式(3)，得出系统的稳态概率：

$$\pi_1=\{(1-\exp(-x_1x))/2x_1\}/\{(1-\exp(-x_1x))/2x_1+(\exp(-x_1x))/2x_3+[1/2-x_1(\exp(-x_2x))/2(x_1-x_2)-$$

$$x_2(\exp(-x_1x))/2(x_1-x_2)]/2x_4+x_1\cdot$$
$$[\exp(-x_2x)-\exp(-x_1x)]/2x_5(x_1-x_2)\},$$
$$\pi_3=\{(\exp(-x_1x))/2x_3\}/\{(1-\exp(-x_2x))/$$
$$2x_3+(\exp(-x_1x))/2x_3+[1/2-$$
$$x_1(\exp(-x_2x))/2(x_1-x_2)-$$
$$x_2(\exp(-x_1x))/2(x_1-x_2)]/2x_4+$$
$$x_1[\exp(-x_2x)-\exp(-x_1x)]/2x_5(x_1-x_2)\},$$
$$\pi_4=\{[1/2-x_1(\exp(-x_2x))/2(x_1-x_2)-$$
$$x_2(\exp(-x_1x))/2(x_1-x_2)]/2x_4\}/\{(1-$$
$$\exp(-x_2x))/2x_3+(\exp(-x_1x))/2x_3+$$
$$[1/2-x_1(\exp(-x_2x))/2(x_1-x_2)-$$
$$x_2(\exp(-x_1x))/2(x_1-x_2)]/2x_4+x_1\cdot$$
$$[\exp(-x_2x)-\exp(-x_1x)]/2x_5(x_1-x_2)\},$$
$$\pi_5=\{x_1[\exp(-x_2x)-\exp(-x_1x)]/$$
$$2x_5(x_1-x_2)\}/\{(1-\exp(-x_2x))/2x_3+$$
$$(\exp(-x_1x))/2x_3+[1/2-x_1\cdot$$
$$(\exp(-x_2x))/2(x_1-x_2)-x_2(\exp(-x_1x))/$$
$$2(x_1-x_2)]/2x_4+x_1[\exp(-x_2x)-$$
$$\exp(-x_1x)]/2x_5(x_1-x_2)\}。$$

系统的非稳定概率 $\pi=\pi_3+\pi_4+\pi_5$。

4.5 数值分析

从图 5 中看出,自恢复容错系统随着运行时间的增加,系统的非有效概率先是增加,随后下降并趋于稳定。

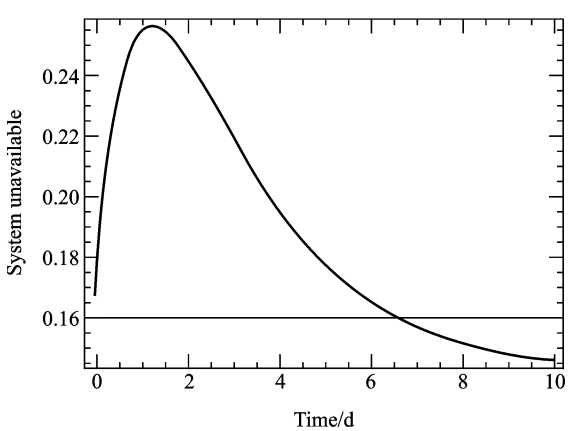

图 5 系统非稳定性与和时间的关系

Fig. 5 System unavailable versus time

图 6 表明了系统的非稳定性随着自愈率 x 增加即自愈时间间隔减少而增加。

5 结语

综合考虑了容错系统的硬件故障和老化现象,

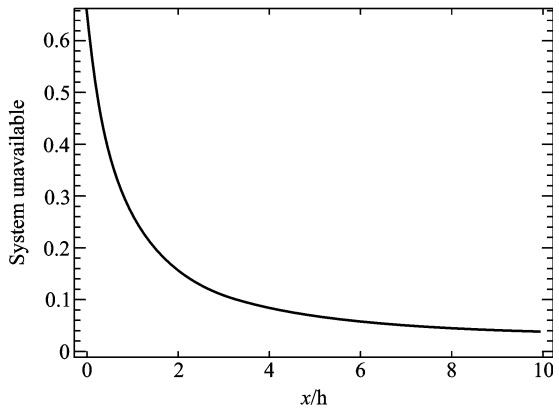

图 6 系统非稳定性与自愈率的关系

Fig. 6 The rate of rejuvenation versus system unavailable

提出了将传统的冗余策略和软件抗衰策略相结合的技术应用于容错系统中,并分析了该系统的状态变化,由于自恢复容错系统中存在定期的自愈操作,使用非马尔可夫随机 Petri 网对系统进行建模。在结合数据分析自恢复容错系统的行为后,得出了系统非稳定概率及系统非稳定性和自愈率之间的关系。

参考文献

[1] Castelli V, Harper R E, Heidelberger P, et al. Proactive management of software aging [J]. IBM J Research & Development, 2001, 45: 311 ~ 332

[2] Marshall E. Fatal error: how Patriot overlooked a scud [J]. Science, 1992, 255: 1347

[3] Huang Y, Kintala C, Kolettin N, et al. Software rejuvenation: analysis, module and applications [A]. Proc 25th Int'l Symp on Fault Tolerant Computing [C]. IEEE CS Press, 1995. 381 ~ 390

[4] Shereshevsky M, Cukic B, Crowel J, et al. Software aging and multifractality of memory resources [A]. Proc Int'l Conf on Dependable Systems and Networks [C]. IEEE CS Press, 2003. 721 ~ 730

[5] Garg S, Telek M, Puliafito A, et al. Analysis of software rejuvenation using Markov regenerative stochastic Petri net [A]. Proc 6th Int'l Symp on Software Reliab Eng [C]. IEEE CS Press, 1995. 24 ~ 27

[6] Balbo G. Introduction to stochastic Petri nets [A]. Brinksms E, Hermanns H, Katoen J-P (Eds). FMPA 2000, LNCS 2000 [C]. ©Springer-Verlag Berlin Heldelberg, 2001. 183 ~ 231

[7] German R. Iterative analysis of Markov regenerative models [J]. Journal of Performance Evaluation, 2001, 44: 51 ~ 72

[8] Choi H, Kulkarni V G, Trivedi K. Markov regenerative stochastic Petri nets [A], Iazeolla G, Lavenberg S S eds. Performance-93 [C]. 1993. 339 ~ 356

Modeling and Analysis of Fault-tolerant System With Rejuvenation

Guo Chenghao, Liu Fengyu
(*Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China*)

[**Abstract**] A fault-tolerant system experiences a crash due to hardware components' faults or progressive performance degradation as an "aging" phenomenon, because of running continuously for very long periods. This paper considers both hardware components' faults and "aging" phenomenon, proposes composing redundant structure and rejuvenation schedule in the fault-tolerant system, formalizes the system with Non-Markovian stochastic Petri nets, and evaluates quantitatively the performance of a system based on this model.

[**Key words**] fault-tolerant system; software rejuvenation; software aging; redundant strategy; Non-Markovian stochastic Petri nets

---

(上接第 64 页)

Optimization Strategy of MPEG－4 AAC Decoder on a Low-cost SoC

Gao Gugang, Shi Longxing, Pu Hanlai, Zhou Fan
(*National ASIC System Engineering Research Center, Southeast University, Nanjing 210096, China*)

[**Abstract**] This paper proposes software optimization strategies using a low-cost SoC which include float-point to fix-point conversion scheme based on statistical analysis and performance oriented customizing scheme for on-chip memory's capacity, and presents optimization methodology based on these strategies for computation intensive applications. the MPEG－4 AAC decoding in real-time is implemented as a case study to illustrate the efficiency of the proposed optimization strategy in both performance and cost. The strategy and methodology also can be used to optimize other DSP applications.

[**Key words**] software optimization; SoC; FFC; on-chip memory; AAC