一种容错实时计算机体系结构的研究与实现

陈 筠，桑 楠，熊光泽

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】为满足对安全关键领域日益增长的可靠性需求，通过对容错关键技术和多处理器系统的深入研究，提出了一种基于松耦合多处理器体系结构的双机容错实时嵌入式系统设计方案。该方案无缝整合了计算机硬件级、操作系统级、应用级的容错技术，以达到从整体上提高系统可靠性的目的。然后，利用马尔科夫状态图法对该系统进行了可靠性分析和数值模拟，结果表明该设计方案能显著地从整体上提高系统的可靠性水平。

关 键 词 双机热备份; 容错; 实时嵌入式系统; 可靠性
中图分类号 TP302.8 文献标识码 A

Design and Implementation of a Fault-Tolerance Real-Time Computer Architecture

CHEN Jun，SANG Nan，XIONG Guang-ze

(School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Based on fault-tolerance technique and multi-processors system, a fault-tolerance real-time embedded dual system solusion is put forward in this paper. The proposed solusion is based upon the loosely coupled multiprocessors architecture. this architecture seamlessly. integrates the fault-tolerance design techniques of hardware level, operating system level, and application level The system reliability is analyzed by the Markov state diagram The results show that the design scheme can enhance the system reliability remarkably.

Key words duplicated hot backup; fault–tolerant; real-time embedded system; reliability

随着计算机技术的日益成熟，以及计算机硬件成本的迅速降低，各种结构复杂、功能强大的实时计算机系统被广泛应用于航空航天器、武器装备、核电监控装置和医疗设备等安全关键系统(Safety-Critical System)中[1]。确保这些计算机系统的可靠性(Reliability)成为人们日益关注的问题[2-3]。

双机热备份设计方案可切实提高系统的可靠性。但它主要针对硬件错误，对于软件错误却无能为力。目前，由于硬件制造技术水平的提高和硬件容错技术的成熟，软件错误成为导致系统失效的主要原因[4]。据调查，在具有硬件容错能力的计算机系统中，其失效65%来自软件[5]。

早期的实时计算机系统为特定的应用设计专用的硬件和软件，其最大的缺点是软硬件的耦合度大，不利于系统可靠性设计，特别是软件错误容忍设计。随着实时操作系统技术的日益发展成熟，实时软件被分离成为实时操作系统和实时多任务软件两部分，实时操作系统实现对硬件的管理，使得实时多任务应用软件与底层硬件无关。这种分层的实时计算机体系结构为提出新的实时计算机容错体系结构提供了契机。

1 双机容错实时系统的体系结构

双机容错实时系统体系结构是在考虑双机比较系统[6]的基础上，结合松耦合多处理机体系结构，在实现系统隔离的同时，在不同的处理机间通过通道互连实现通信，为在硬件容错中结合软件容错提供可能[7]。

双机系统的运行状态定义为：(1) 如果A机与B机均正常运行，则将A机作为主系统，B机作为备份使用，A机的运行结果作为系统输出，A机运行到检测点，向B机发送日志，B机更新日志列表。(2) 如

收稿日期：2005 – 09 – 21
基金项目：信息产业部十五预研基金资助项目(41315040106)
作者简介：陈 筠(1980 –)，女，硕士，主要从事嵌入式容错系统方面的研究.

果A机正常而B机故障，亦将A机的运行结果作为系统输出，同时将B机的运行故障状态报告A机，并向B机进行复位控制操作。(3) 如果A机故障，B机正常，则进行开关切换操作，B机进行系统备份任务重调度，B机运行结果作为系统输出，向A机进行复位控制操作，并在检测点更新A机日志，保持需要备份的任务的状态一致[8]。

双机容错实时系统体系结构结合嵌入式实时系统的体系结构，采用层次结构和模块结构相结合的思想，无缝整合计算机硬件、操作系统、应用软件等三级容错设计，克服了软、硬件分离和脱节的问题，可提高系统的灵活性和可移植性。

2　双机容错实时系统的设计

双机容错实时系统体系结构的每一层均可看作是一个相对独立的子系统，层中包含不同的功能模块，结构如图1所示。图中分别加入了容错通信模块(Multiprocessor Communication for Fault-Tolerance，MCFT)、实时系统(Real-Time Operating System，RTOS)系统级容错组件、任务级大动态冗余组件。

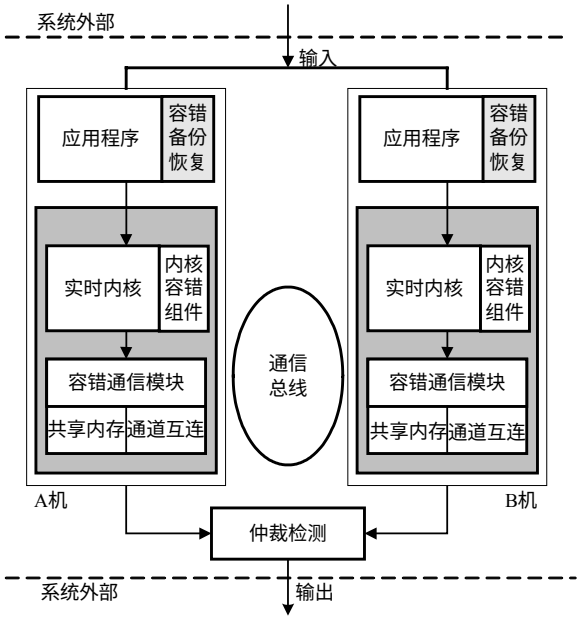

图1　双机容错实时系统体系结构

第一层中加入MCFT模块，作为板级支持包(Board Support Package，BSP)的一部分，也是硬件平台的抽象层，可为操作系统提供统一的界面，提高系统的可移植性。有容错需求的任务，通过MCFT所提供的功能传递日志，保持主系统和备份系统关键任务的状态和数据一致。MCFT屏蔽了底层通信的具体实现细节，使系统的实现与连接介质无关。

为保证实时系统从硬件故障和永久软件故障恢复，采用系统切换方法，在第二层中加入RTOS系统级容错组件，包括系统内核级容错支持组件、主/备用机切换支持组件和系统自诊断组件。

任务级动态冗余[9]模块被用于嵌入式实时系统，可使实时系统从暂时软件故障恢复。这也是当软件发生错误时保证系统实时性的重要措施。

2.1　故障检测

系统容错以故障检测技术为基础，以各种冗余技术为手段。对于实时系统来说，为提高故障判别的成功率，故障检测应该及时准确地定位故障并尽量减小系统开销。

在系统中，故障检测按层次模型进行，其目的是实现信息隐藏，避免故障跨层次传播。采用自诊断的方法诊断系统级的故障，用任务级的检测诊断应用级的故障。

2.1.1　系统自诊断

系统自诊断划分为系统启动自检测阶段和周期自检测阶段。自动启动诊断的因素有主/备用机定时切换和主用机发生故障。周期自检测阶段根据系统需求，周期性检测外设和通信口。每个阶段对应设备的几种功能块，包括CPU的自诊断、中断响应自诊断、串口自诊断、定时器自诊断、离散量自诊断和RAM自诊断等。

由于结果比较是实时系统中任何事务处理都需要经历的步骤，因此把任务级的故障检测放到结果判别部分进行。

2.1.2　任务级动态冗余

任务级动态冗余方法是实时系统中瞬间故障的恢复方法之一。在实时多任务的环境下，充分利用操作系统提供的功能，为各个基本任务建立后备任务作为冗余，对后备任务进行容错调度，从而起到类似于重试或回溯的作用，并利用检查点技术和传递日志法保持主系统和备份系统状态的一致性，实现错误恢复。

根据应用程序的要求，结合任务实时性，采用以下的模型定义。

(1) 把应用程序P分解成多个任务T，$P=\{T_1, T_2, \cdots, T_n\}$，任务以过程的形式出现。

(2) 当$i>j$时，任务优先级 $p_{T_i} > p_{T_j}$，任务可以根据要求及时占有处理器，实现实时处理。在每个任务的最后设置检查点，传递日志。

(3) 为各基本任务准备一个后备任务$P'=\{T_1', T_2', \cdots, T_n'\}$存放在内存中。一般情况下，后备任务不建立，不占有系统资源，仅在需要时才激活使用。后备任务的优先级比相应的优先级要高。一旦建立就抢占执行，是某种意义上的重试或程序卷回。

(4) 为实现恢复功能的后备任务，可以与原有任务完全一样，也可以是替换算法。以下任务级动态冗余替换算法，能为各个任务产生容错调度，从而实现任务冗余。

Step1： 建立任务T_1，T_2，…，T_n；

Step2： while N=1；N <=$N_{\max}$；

N=N+1；

运行任务T_i；

检测T_i的结果；

IF 结果通过 THEN 输出结果，删除任务T_i；

ELSE 激活任务T_i'；break；

END

Step3： N>$N_{\max}$ 系统报警

当后备任务执行了$N_{\max}$次之后还通不过检测，就认为系统出现永久故障，系统报警。$N_{\max}$是个阀门值，是由实时要求所决定的。

2.2 主/备份切换

仲裁检测电路中为主/备用机设置了“看门狗”监视器。当主/备用机处于正常工作状态，运行于CPU上的某一任务周期性地对“看门狗”施加复位信号，“看门狗”计数器就不可能产生溢出触发信号；当CPU出现故障时，“看门狗”会输出一个离散触发信号并发出报警，此时系统进行自动切换，让备用系统机工作。

3 利用马尔科夫状态图进行的可靠性分析

3.1 错误模型

双机容错实时系统的错误模型定义如下：

(1) 系统错误的到达过程是一个泊松流(Poisson Process)，相继错误到达时间间隔服从负指数分布$T_f = \mathrm{e}^{-\lambda t}$。根据泊松分布的平稳增量性质，可知$\boldsymbol{P}\{N(\Delta t)>=2\}=o(\Delta t)$，即在间隔时间$\Delta t$充分小时，系统连续发生多次错误的可能性为$\Delta t$的高阶无穷小。

(2) 错误可分为硬件错误和软件错误，软件错误包括操作系统和任务发生的错误。另外，硬件错误可分为暂态硬件错误和永久硬件错误；软件错误可分为本机可恢复的错误和需要备份系统恢复块恢复的错误。

(3) 故障的发生是不相关的，部件的失效率λ和维修率μ是常数。

(4) 故障不传播。

3.2 利用马尔科夫状态图法评估可靠性

可靠性是指一个系统在一定的环境下和给定的时间内能按预定的要求完成一定功能的概率。

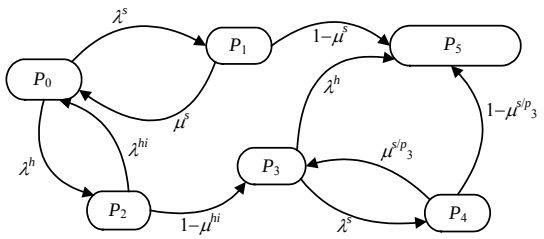

图2 采用双机容错方式下的马尔科夫状态转移图

图2是利用上述假设构造出的双机容错实时嵌入式系统的马尔科夫状态。系统运行过程中的6个状态定义为：

状态P_0：双机都正常。

状态P_1：系统处于软件容错状态。

状态P_2：系统处于硬件容错状态。

状态P_3：硬件系统发生永久失效，系统运行在单机系统中。

状态P_4：系统处于单机软件容错状态。

状态P_5：整个系统失效。

由图2可以得到马尔科夫状态微分方程：

$$\tilde{p}_n = \tilde{\boldsymbol{p}}_0 * \boldsymbol{p}^n \tag{1}$$

$$\boldsymbol{P} = \begin{bmatrix} 1-\lambda^s-\lambda^h & \lambda^s & \lambda^h & 0 & 0 & 0 \\ \mu^s & 0 & 1-\mu^s & 0 & 0 & 0 \\ \mu^{hi} & 0 & 0 & 1-\mu^{hi} & 0 & 0 \\ 0 & 0 & 0 & 1-\lambda^s-\lambda^h & \lambda^s & \lambda^h \\ 0 & 0 & 0 & \mu^{s/p_3} & 0 & 1-\mu^{s/p_3} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{\boldsymbol{p}}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2}$$

式中 $\boldsymbol{P}$ 为状态转移概率矩阵，矩阵方程(方程组)称为查普曼-柯尔莫戈罗夫(Chapman-Kolmoqorov)方程，由此可以解出系统处于任意状态的概率。

状态P_{66}表示系统失效，所以系统的可靠度为：

$$R(t)=1-P_{66}(t) \tag{3}$$

在计算该系统的可靠度时，将状态5作为吸收状态。对式(3)求该微分方程的数值解，不同参数下，系统的可靠度值(精度为10^{-10})如表1所示。

表1　双机容错系统可靠度 ($\mu^s = \mu^{s/p_3} = \mu^{hi} = 0.9$)

t	λ			
	λ^s =0.010 0	λ^s =0.001 0	λ^s =0.010 0	λ^s =0.005 0
	λ^h =0.001 0	λ^h =0.001 0	λ^h =0.000 1	λ^h =0.000 5
1	1	1	1	1
10	0.999 988 921 6	0.999 995 839 7	0.999 997 347 4	0.999 997 202 9
50	0.999 571 988 4	0.999 861 772 1	0.999 872 945 4	0.999 890 140 1
100	0.998 260 737 4	0.999 438 989 7	0.999 465 911 1	0.999 545 559 9
500	0.965 383 618 5	0.987 647 319 6	0.987 907 918 3	0.989 718 131 8
1 000	0.896 584 602 5	0.958 587 771 2	0.959 277 817 6	0.964 945 164 4
5 000	0.413 504 493 7	0.641 284 447 9	0.644 433 196 4	0.675 138 136 3

3.3　可靠性对比

用马尔科夫状态图法对采用双机热备份方式和采用恢复块方式的单机容错系统进行可靠性分析。在系统软件失效率 $\lambda^s = 0.005$，以及硬件失效率 $\lambda^h = 0.001$ 和维修率 $\mu = 0.9$ 的相同条件下，在区间[0,1 000]上进行可靠性对比，结果如图3所示。

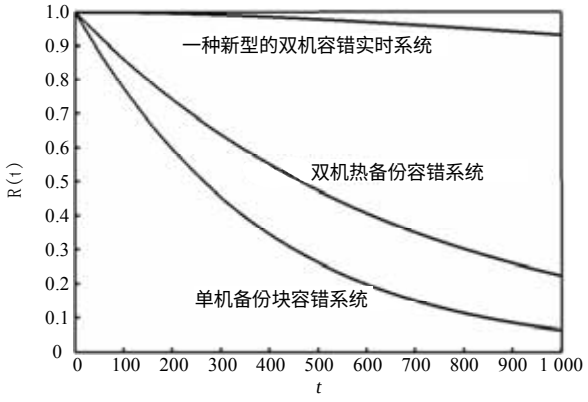

图3　三种容错方式下可靠度随时间变化曲线

双机热备份系统[10]由两个能完成相同功能的计算机模块并行执行相同的计算，双机不能通信，根据A机和B机周期向仲裁检测电路发送的自检信号判断A机系统和B机系统的运行状况。

单机备份块容错系统中主模块的运行结构由验收测试检验，若结果通过测试结果，则输出；否则运行备份模块。恢复块在无错和出错情况下的响应时间差异很大。应用于实时系统时，恢复块必须与时间冗余相结合。结果显示，本文提出的双机容错实时系统比采用单纯硬件容错的双机热备份系统和采用单纯软件容错的单机备份块容错系统的可靠性都有很大的提高，而且随着时间的增长，可靠性更为明显。

4　结 束 语

随着实时系统在安全领域内越来越多的应用，可靠性已经成为衡量系统优劣的关键因素之一。传统的双机热备份容错系统只能满足系统某一方面的容错需求。为了在硬件(或软件)出现暂时或(永久)故障的情况下，系统仍能在规定的时限范围内完成运算，并输出正确的结果，本文提出了一个软、硬件结合的完整的解决方案，该方案在满足系统实时性的同时，从整体上提高系统的可靠性。数值模拟结果表明该系统具有极高的可靠性。

参 考 文 献

[1] TAL O, MOCOLLIN C, BENDELL A. Reliability demonstration for safety-critical systems[J]. IEEE Trans. on Reliability, 2001, 50(2): 194-203.

[2] 陈　宇. 实时异常处理技术的探讨[J]. 计算机工程, 2004, 30(21): 61-63.

[3] 吕　勇, 谢长生, 高三红. 实时测控计算机应用谢的可靠性保障技术[J]. 计算机应用, 2003, 23(6): 101-106.

[4]韩建军, 李庆华. 基于软件容错的动态实时调度算法[J]. 计算机研究与发展, 2005, 42(2): 315-321.

[5] KIM K. The distrubuted recovery block scheme in software fault tolerance[M]. [S. l.]: Wiley, 1995.

[6] 陈　宇. 高可靠容错实时系统的支撑技术研究[D]. 成都: 电子科技大学, 2004.

[7] 金士尧, 胡华平, 李宏亮. 具有容错结构的高可用计算机双系统研究[J]. 中国工程科学, 1999, 1(3): 46-50.

[8] 吴　娟, 马永强, 刘　影. 一种基于主备机快速切换的双机容错系统[J]. 计算机应用, 2005, 25(8): 1948-1951.

[9] KRISHNA C S K. On scheduling tasks with a quick recovery from failure[J]. IEEE Trans. Computer, 1986, C-35: 448-454.

[10] 李宏亮, 金士尧, 胡华平. 短事物、强实时双机容错系统的研究[J]. 计算机学报, 2003，26(2): 243-249.

编　辑　熊思亮