1999 年 12 月
第 1 卷第 3 期

中国工程科学
Engineering Science

Dec.1999
Vol.1 No.3

学术论文

具有容错结构的高可用计算机双系统研究

金士尧，胡华平，李宏亮
（国防科技大学计算机学院，长沙 410073）

[摘要] 为了确保高的可用性，在重大工程实践中往往采用具有容错结构的计算机双系统。双系统处理得到两个结果的异同性，历来是研究双系统容错的技术重点和难点。文章在双系统可用性分析、结果判错与选择、以及双工切换技术方面都有突破性的进展，并在重点工程中得以实现，取得显著效果。

[关键词] 可用性；双工系统；故障判别率；切入成功率

1 问题的提出

群机冗余是提高系统可靠性，实现系统高可用度的有效途径[1,2]。在工程实现中，常见的是同构型计算机双系统冗余。其工作方式有四种，即冷备、温备、热备和双工[3,4]。它们的拓扑结构如图 1 所示。

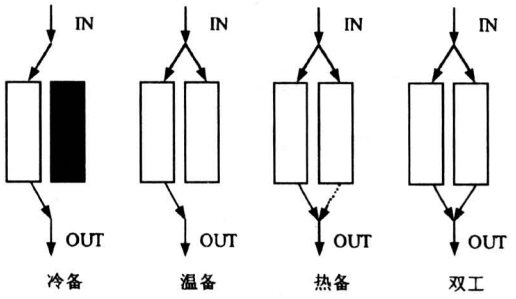

图 1 双系统工作状态

Fig.1 The working status of dual systems

（1）冷备 双系统冷备的工作方式是在工作机故障时，未加电的备份机经过启动后自动切入接替工作机工作，待故障修复完毕，工作机去电变成备用机。

（2）温备 双系统温备的工作方式是双系统同时加电，且一机工作，另一机处于等待或故障诊断状态。一旦工作机故障时，系统将进行自动切换，由温备机接替工作机工作，待故障修复完毕后，工作机变成备份机。

（3）热备 双系统热备的工作方式是双机同时加电，且均处于工作状态，只是热备机的处理结果不输出。一旦值班工作机出现故障时，更换值班机，进行结果切换，即原热备机的处理结果输出。

（4）双工 双系统双工的工作方式是双机同时加电和工作，处理的两个结果进行比较后选择合适（正确）的输出。

实际上，双系统热备和双工都是双机工作同时处理，不过前者由值班机输出结果，非值班机的结果备用，输出结果带有盲目性；而后者结果经过比较，选择合适（正确）的输出，应该说可用性更高。

所以，双工（包括热备）系统运行的工作状态包括双工、温备和冷备状态，其工作状态转换如图 2 所示。

由此可见，具有容错结构的同构型计算机双系统的高可靠性和高可用度，除了来自系统软硬件本身的可靠性可用度外，还严重地依赖下列因素：系统发现故障的能力，包括监测时间和成功率；系统状态的转换能力，包括转换时间和成功率。

[收稿日期] 1999－07－01；修回日期 1999－09－09

[作者简介] 金士尧（1937－），男，江苏苏州市人，国防科技大学教授，博士生导师

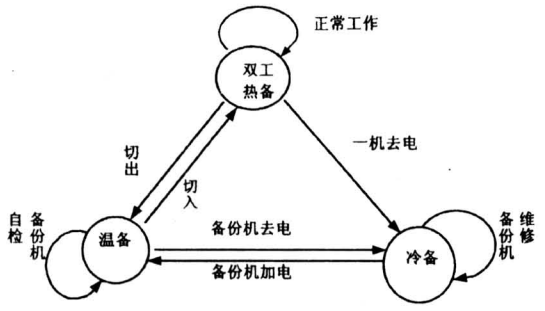

图 2　双工系统状态转换图

Fig.2　The status transition diagram of dual systems

实际上，双系统发现故障的能力不可能达到 100%，它的状态转换也不会 100%成功。也就是说，应该研究具有一定故障判别成功率和一定状态切换成功率的双系统。这是实际工程中提出的现实问题。

2　同构型双系统的可用度分析

在经典的双系统可用度分析中，都是假定故障判别成功率为 100%，且故障诊断的时间允许忽略不计，同时假定切换（包括切入、切出）成功率为 100%，且切换时间可以忽略不计[5,6]。在上述条件下，双系统中只要有一台机器工作，即认为双系统可用。其状态转移图见图 3a。而现实工程中，必须考虑双机处理结果不同时的故障判别方法，以及故障无法判定时的故障诊断技术。即要考虑故障判别成功率和故障诊断时间，它的状态转移图见图 3b[7]。

在工程实践中，为了更精确地分析双系统的可用度，还应考虑双机的切换问题。其中包括切入成功率，与此相关的切入时间和再次切入的时间。图 3c 既考虑故障判别问题，又考虑切入过程的双系统状态综合转移图[8]。

由此可见，双系统可靠使用的状态为 0 状态和 1 状态，2 状态是不能使用的故障状态，而状态 0′、1′和 1″都是双系统的过渡状态，它们不是有效工作状态。

从双系统状态转换图 3c 中，可以计算出考虑故障判别率和切换成功率情况下的双系统稳态可用度

$$A_{\text{double}}^{(\infty)} = \left(1 + \frac{2\lambda}{\mu}\right) \Big/ \left(1 + 2\frac{\lambda(1-D)}{\alpha'} + 2\frac{\lambda}{\alpha} + 2\frac{\lambda}{\mu} + 2\frac{\lambda^2}{\mu^2} + \frac{2(1-C)\lambda}{\beta}\right)。$$

当 $D=1$，$\alpha=\alpha'=\infty$，相对应双系统状态转移图 3b 情况，即考虑故障判别率和诊断时间的情况，则：

$$A_{\text{double}}^{(\infty)} \Big|_{\substack{D=1 \\ \alpha=\infty,\alpha'=\infty}} = \left(1 + \frac{2\lambda}{\mu}\right) \Big/ \left(1 + \frac{2\lambda}{\mu} + \frac{2\lambda^2}{\mu^2} + \frac{2(1-C)\lambda}{\beta}\right)。$$

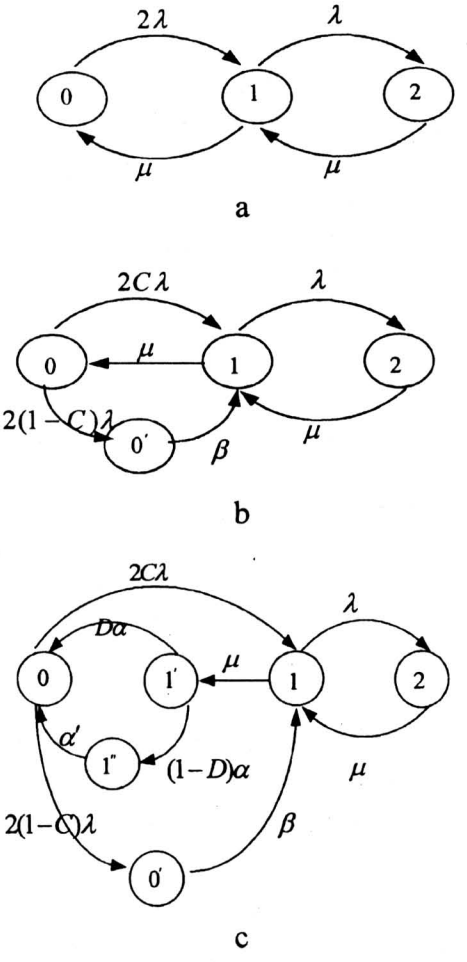

图 3　双系统状态转移图

Fig.3　Status transition diagram of dual systems

0－双系统双机正常工作状态；1－双系统一机正常工作，另一机维修状态；2－双系统两机均不能正常工作状态；0′－双系统虽然双工，但有故障而无法判定状态故障机；1′－双系统一机维修成功，工作机等待切入状态；1″－双系统一次切入不成功状态；λ－平均失效率为平均无故障时间（MTBF）的倒数；β－故障诊断率为平均诊断时间的倒数；μ－平均维修率为平均维修时间（MTTR）倒数；α－切入失效率为平均切入时间的倒数；α′－再次切入失效率为再次切入时间的倒数（重启双工时间的倒数）；C－故障判别率；D－切入成功率

经典的双系统可用度计算，就是在 $D=1$，$\alpha=\alpha'=\infty$，$C=1$，$\beta=\infty$的特殊条件下的可用度。

这是双系统可用度的极限：

$$A_{\text{double}}^{(\infty)} \mid_{\substack{C=D=1 \\ \alpha=\infty,\alpha'=\infty}} = \left(1+\frac{2\lambda}{\mu}\right)/\left(1+\frac{2\lambda}{\mu}+\frac{2\lambda^2}{\mu^2}\right)。$$

采用现代典型的工作站和服务器的各种可靠性数据代入计算，即设定：

MTBF $\doteq$ 20 000 h $\quad \lambda = 0.000\ 05\ \text{h}^{-1}$

MTTR = 10 h $\quad \mu = 0.1\ \text{h}^{-1}$

C = 0～1 $\quad \beta = 1/(1\ \text{min}) = 60\ \text{h}^{-1}$

D = 0～1 $\quad \alpha = 1/(1\ \text{s}) = 3\ 600\ \text{h}^{-1}$

$\alpha' = 1/(1\ \text{h}) = 1\ \text{h}^{-1}$

则双工系统的可用度与故障判别成功率和切入成功率的关系曲线见图 4 和图 5。

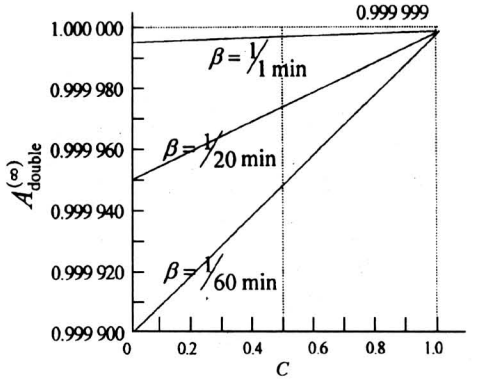

图 4 可用度与 C 的关系

Fig.4 The relationship between availability and C

如果工程需求的可用度为 99.993%。采用同构双系统，在典型值的计算中，可以获得 99.999 95% 的可用度。考虑故障判别成功率 $C=0.5$，且 $\beta=1/(30\ \text{min})$，则该双工系统的可用度约为 99.997%。进一步考虑双工切换的成功率 $D=1/(30\ \text{min})$，则该双工系统的可用度约为 99.997%。进一步考虑双工切换的成功率 $D=0.5$，$\alpha=1/(1\ \text{s})$，则它的可用度只有 99.993%。因此，在双系统工程中，必须重视实现故障成功率和切入成功率的有关技术设计。

3 结果判别与质量报告

双工系统的结果判别成功率取决于结果比较的错误判据。常规的双工系统结果判别原则是：

(1) 比较结果相等时，由值班机输出；

(2) 比较结果不相等时，判别两个机器在处理过程中是否有硬件故障和软件故障（例如处理超时、非法操作、越界等），由值班机选择结果优者输出。

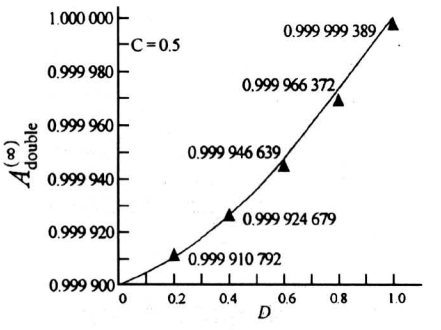

图 5 可用度与 D 的关系

ig.5 The relationship between availability and D

为了避免偶然性故障，有的常规双系统还增设了结果复执和标准程序的测试。结果复执能形成结果三取二的有利局面，然而复执不但耗时而且会造成双机失步的困难局面。

文章主张采用过程控制质量的原理。在双工系统处理结果的中间过程，分成若干自然子过程，每个子过程处理结果都附有一个处理子过程的质量报告，其格式如下：

序号	方法/模型	处理时间	自评报告	信息段

序号—子处理过程的编号；方法/模型—子处理的处理某一种方法/模型；处理时间—可以是正常时间与处理时间的差数；自评质量—自评质量是与标评质量（标准测试）相比的误差数；自评质量假定分为 A、B、C、D 四等

值得注意，质量报告的形成离不开定时的标准例题测试，它是构成质量评定的基准。在二次测试的间隔时间内的机器处理服务质量视同自评的基准质量。

这样，在运用了质量评估报告的前提下，双工系统的结果比较选取规则如下。

(1) 结果比较相等，任选一个结果，并在双工选择质量报告中评定为 A。

(2) 结果不相等，选择质评报告优者，但双工选择质量评定不能超过 B，且有：选取质量 A 结果输出，选择质评为 B；选取质量 B 结果输出，选择质评为 C。

(3) 判两机的硬故障和软故障存在否？存在，则将选择质评报告的质评定为 D。

(4) 对双机的两个结果比较由值班机负责选取，选择质评优者的结果输出。

整个双工结果比较有三个阶段，即结果匹配、

结果比较和结果选取。具体流程见图 6。

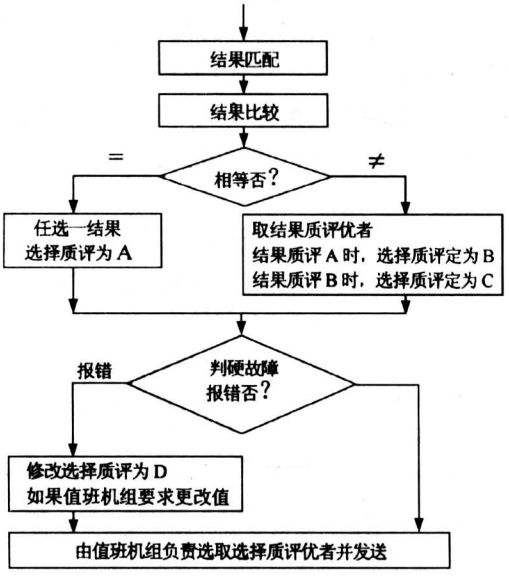

图 6　双工结果比较流程图

Fig.6　Flow chart of results comparing

4　双工系统的状态切换技术

计算机双系统的状态切换主要是指双工系统变成单工系统或单工系统切换成双工系统。前者称为切出，后者称为切入。切出过程比较简单，且没有双机的同步问题。而切入过程比较复杂，切入命令对工作机来讲，相当于一个系统宏中断，切入现场要迁移到切入机中；还存在一个切入同步问题，即当双机的切入准备都完毕时，才发出双工整步命令。系统进入双工状态。切入全过程见图 7。

其中有一个选择切入最佳时机问题。所谓切入时机最佳就是迁移的切入现场最小。对于服务流水处理的机器而言，切入现场最小是流水服务的初态。为了保障系统对外服务不断流，有效的方法是采用输入双缓冲技术。

当切入命令来临时，无论工作机与切入机均改换输入缓冲，缓冲新的输入请求服务，但不执行；与此同时，切入机继续处理原缓冲区内的剩余请求服务，直到原缓冲空，执行现场迁移；等到双工整步命令，两台机器才开始处理新的缓冲区中的服务请求。双缓冲技术的原理图见图 8，设：

切入时刻 t_0
选择最佳切入时刻 t_1
迁移现场时刻 t_2

$t_1 - t_0 = t_a$ 选择最佳切入时延，
$t_2 - t_1 = t_b$ 迁移现场时延。

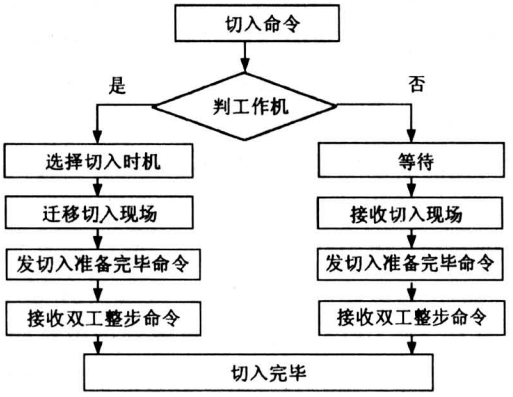

图 7　切入过程流程图

Fig.7　The flow chart of cut－in process

其中，选择最佳切入时延 t_a 与切入时刻以前进入系统而尚未处理的请求服务数量 n 和平均服务时间 t_s 有关。取平均值

$$\bar{t}_a = \bar{n} \cdot \bar{t}_s$$

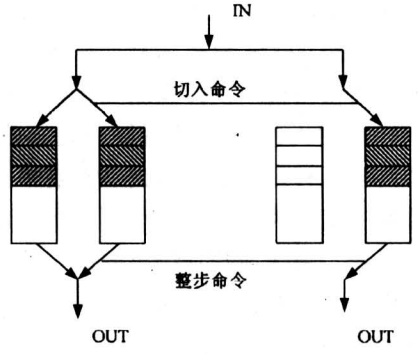

图 8　输入双缓冲技术

Fig.8　The technology of double input buffers

迁移现场时延 t_b，与迁移的现场和大小以及迁移现场的单位传输时间 t_c 有关。通常现场 f 中包含两部分，一部分现场是时间函数 f_t，另一部分是非时间函数 $f_{\bar{t}}$，与时间无关的现场可以在切入时刻之前先期准备完毕。所以：

$$\bar{t}_b = f \cdot \bar{t}_c = (f_t + f_{\bar{t}}) \cdot \bar{t}_c \xrightarrow{f_{\bar{t}}=0} f_t \cdot \bar{t}_c。$$

在忽视双工整步命令时延的条件下，整个切入过程的时间为：

$$t_{cutin} = \bar{t}_a + \bar{t}_b = \bar{n} \cdot \bar{t}_s + f_t \cdot \bar{t}_c。$$

5　结束语

文章是有关计算机双工系统方面的理论与技术

成果。它突破了经典的双工系统理论分析，提出了故障判别成功率，故障诊断时间，切入成功率，切入时间和再次切入时间等新的概念。丰富了双工分析的内涵，使理论分析更接近工程实践。用这些新概念可提高研制质量，具有重要的指导意义。

例如，故障判别成功率直接来源于双工设计中的故障判据。因为任何故障判据不可能覆盖机器可能出现的所有故障。也就是说，故障判别成功率在实际中不会是100%。复杂的故障判据虽然能提高故障覆盖范围，但实现判据的难度会增加，而且直接影响到可靠性。折衷的设计是合理的。文章的双工系统理论分析指出：在一定的故障判别成功率的条件下，缩短故障诊断时间将是有效提高双工系统可用度的关键。换言之，应该平衡选取故障判别成功率和故障诊断时间。

再例如，不可能设计出一个切入成功率达到100%的切入方案，其中，切入成功率与切入时间、包括再次切入时间相互关联。理论分析证明，应该平衡设计切入成功率和切入时间。

本文在实践双工系统故障判据和状态切换过程中，提出了两项新技术，即控制处理信息过程的质量报告和选择切入最佳时机的最小迁移现场的技术。前者是为了有效提高故障判别成功率，即在两机处理结果不相同的情况下，利用信息处理中间过程的质量差异，来增加判别结果的真假；而后者是直接影响切入成功率和切入时间的关键所在。目前切入过程最常见的技术是“断流归零”法。例如DEC的TruclusterOracle公司的数据库并行工作方式，以及网络通讯切换……。它们都在切入过程中中断服务处理，迁移基本现场，丢掉过程中的服务。而本文中提出的双输入缓冲方案，目的在于不让服务有断流，达到迁移现场最小的目的，以实现提高的切入成功率，缩短切入时间的研制要求。这两项工程实现技术都具有重要的参考价值。

参考文献

[1] 金士尧，王志英，胡华平. 强实时高可靠性的系统研究 [J]. 计算机工程与科学，1997，(A1)：1～6

[2] Laprie J C. Dependable computing：concepts，limits，challenges [A]. In：Special Issues FICS－5 [C]，Pasadena，CA，1995. 42～54

[3] Welks S R. Reliability modeling of Hardware/software systems [J]. IEEE Trans. on Reliability，1995，44 (3)：413～418

[4] 胡华平，金士尧，王维. 分布式实时系统的高可靠性研究与实现 [J]. 计算机研究与发展，1998，35 (9)：841～845

[5] Heimann D I. Dependability modeling for computer system [A]. In：Proceedings of Annual Reliability and Maintainability Symposium [C]. USA，1991. 120～128

[6] 曹晋华，程侃. 可靠性数学引论 [M]. 北京：科学出版社，1992. 219～256

[7] 胡华平，金士尧，王维. 分布式实时系统可靠性模型 [J]. 计算机学报，1997. 29 (增刊)：71～76

[8] 胡华平，肖晓强，金士尧. 考虑切换的强实时双机系统的可靠性研究与实现 [J]. 计算机学报，1999，22 (10)：1080～1084

Study of a High-Availability Dual Computer System with Fault Tolerant Architecture

Jin Shiyao，Hu Huaping，Li Hongliang

(*School of Computer Science*，*National University of Defense Technology*，*Changsha* 410073，*China*)

[**Abstract**] To ensure the high availability，the duplex-computer system with fault tolerance architecture is used in crucial projects. The similarites and differences of the two results from the duplex system have been the key difficulty of this domain. In this paper，there are great progresses in the analysis of availability，the judgement over of two results and the switch over two systems. The techniques have been realized in an important project with significant effect.

[**Key words**] availability；duplex system；fault judgement ratio；switch success ratio