基于微包协议的三模冗余容错计算机无缝重构算法

张伟功 朱晓燕 关 永 周继芹 尚媛园
(首都师范大学信息工程学院 北京 100048)

摘 要 提出了一种基于微包传输协议的系统重构算法,制定了系统重构命令协议,设计了关键数据区的一致性管理算法。不需增加额外的硬件资源和系统开销,利用三机之间的数据交换通道,即可实现三模冗余容错计算机的无缝重构,保证系统在故障恢复重构时工作的连续性。同时,完善的重构命令协议使得重构软件的开发可以采用面向对象思想构成标准化的中间件,具有很强的扩展能力,为构建标准化嵌入式三模冗余容错计算机内核奠定了良好的技术基础。

关键词 容错计算机,三模冗余,故障恢复,系统重构,微包协议
中图法分类号 TP302.8 **文献标识码** A

Seamless Reconstruction Method for TMR Computer Based on Micro-package Protocol

ZHANG Wei-gong ZHU Xiao-yan GUAN Yong ZHOU Ji-qin SHANG Yuan-yuan
(Information Engineering College, Capital Normal University, Beijing 100048, China)

Abstract This paper presented a system reconstruction algorithm based on micro-package protocol. It formulated the system reconstruction command protocol and designed a consistency management algorithm of the key data area. The seamless reconstruction of TMR fault-tolerant computer was implemented through data exchange channels among the three computers without any additional hardware resources and system overhead, which ensures the continuity of system operation during the period of fault recovery and system reconstruction. Meanwhile, due to the perfect reconstruction command protocol, the development of the reconstruction software can use object-oriented method to constitute standardized middleware, which enhances the system's expanding capability greatly. This will lay good foundation of constructing standardized kernel of embedded TMR fault-tolerant computer.

Keywords Fault-tolerant computer, TMR, Fault recovery, System reconstruction, Micro-package protocol

1 引言

三模冗余容错计算机(TMR)可以在单模故障时保证系统任务执行的正确性,可以有效地提高实时嵌入式系统的安全性与可靠性,在航天、航空、铁路控制等领域有着广泛的应用需求。当单个计算机模块出现故障后,TMR计算机会降级为双机工作模式,如果不进行修复,系统将无法对后续的故障进行容错,安全性与可靠性都会随之降低。引起系统降级的故障可能是瞬态故障,也可能是永久故障。瞬态故障经过一段时间或重新复位后可以自动修复,永久故障则需要经过人工修复。不论哪种情况,如果修复后的机器能够重新进入系统,使系统恢复为三模冗余工作模式,将会大大提高TMR计算机在长时间连续工作环境下的安全性与可靠性。

系统重构就是三模冗余容错计算机从故障状态恢复为三模运行状态的过程,主要包括故障修复、重构识别、工作现场恢复、重新同步等环节。工作现场恢复包括机器状态与内存数据区的恢复,是实现系统重构的基础。根据恢复策略的不同,现场恢复可分为向后回卷和前向恢复两种方式。

向后回卷[6]是在系统任务中设置一些检查点,将系统的关键状态保存下来。故障恢复时,用这些保存的关键状态数据使3个计算机模块均回退到相同的状态,重新开始运行。这种恢复方式三机之间交换数据量少,但需要耗费大量的系统时间开销,会使系统运行中断。

前向恢复[3-7]是通过将正常运行的机器的当前状态及内存数据拷贝到故障机器上,使故障机器与正常机器的状态一致后,从当前点继续运行。这种恢复方式需要在三机之间交换大量的数据,重构时间与三机数据交换速率及数据交换量密切相关。常见的恢复方法是通过串行数据通道进行集中数据恢复[3-5],恢复过程中暂时停止系统的运行。恢复数据量大时,可使系统中断运行时间达到数分钟[4]。文献[1]给出了一

到稿日期:2008-06-02 返修日期:2008-12-12 本文受国家自然科学基金(60873006),北京市自然科学基金资助项目(4062009和4082009),北京市教委重点项目(KZ200710028014)资助。

张伟功(1967—),男,博士,研究员,主要研究方向为嵌入式计算机系统结构与容错技术、系统集成与VLSI设计方法学,E-mail:zwg771@yahoo.com.cn;**朱晓燕**(1967—),女,高级工程师,主要研究方向为嵌入式计算机系统结构与可靠性技术;**关 永**(1966—),男,博士,教授,主要研究方向为智能信息处理与嵌入式系统设计;**周继芹**(1979—),女,硕士,主要研究方向为计算机容错技术;**尚媛园**(1977—),女,副研究员,主要研究方向为嵌入式系统设计。

种基于存储器双机窃取拷贝的恢复方案，可以不中断系统运行，快速实现大量内存数据的传送，但需要复杂的硬件支持，更适合在双机系统中实现。文献[6]是一种部分恢复方案，在数据/输出表决时一旦检查到故障状态，立即对故障机器故障区域进行恢复，可实现对瞬态故障的状态恢复，但不适用于模块级恢复。文献[2]提出的阶梯型恢复方法以进程为单位逐步恢复系统到三模冗余状态。恢复过程中，系统采用双机与三模混合运行模工，管理复杂，比较适合在三模冗余容错服务器中应用。

本文针对一般的嵌入式三模冗余容错计算机研究系统重构策略与重构恢复算法，提出一种采用微包协议结构的重构恢复算法和关键数据区一致性管理算法，可以在重构过程中保持系统的连续工作。第 2 节简单介绍了目标系统组成，第 3 节描述系统重构策略，第 4 节给出了重构恢复算法，第 5 节对关键数据区的一致性管理算法进行了研究，最后对研究工作进行总结。

2 系统组成

不失一般性，本文采用一个以 80X86 处理器组成的三模冗余容错计算机作为目标系统，系统结构如图 1 所示。系统由 3 个同构计算机模块组成，CPU 主频为 100MHz，内存 64MB，包括一个 32 位精度 1μs 的时基定时器和若干开关量输出。每个计算机模块中都包含三机同步模块、串行数据交换模块和电源管理等 3 个容错管理相关功能模块。三机同步模块在系统执行过程中用来维持三机的同步。电源管理模块在正常计算机模块的控制下可对故障机器进行开关电控制。3 个计算机模块的串行数据交换通道构成一个三机两两互连的同步串行通信总线系统，通信速率为 8Mbps，用来实现三机状态与数据交换，是三机进行数据表决与故障重构的基础。

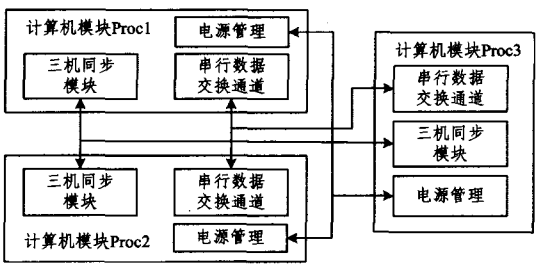


图 1 目标系统组成框图

3 重构策略

当三模冗余容错计算机中某一计算机模块不能正确同步、连续多次表决数据均不正确或软件看门狗溢出时，都认为它发生了严重故障，就要进行系统重试，对故障机器进行重构。重试时，两个正常计算机模块先将故障机器从系统中切除，使系统降级为双机工作模式，然后通过电源管理模块将故障机器关电，再加电。故障机器再加电后，若自检合格，它向其机器发送一个带有重构请求状态的同步请求，使系统从故障重试过程进入到重构识别过程。若故障机器加电后不能正常运行，在具备条件时，可对其进行人工修复后，再使其加电向其他机器发送重构请求。若不能进行人工修复，则保持系统在双机运行模式。

系统中各计算机模块只有在进入周期同步过程以后，才会对附加在同步请求中的重构请求状态进行判断和识别。如果一个同步请求带有重构请求状态，则认为发出该同步请求的那个计算机模块需要被重构。这种识别方式不仅可以实现对故障重试机器的重构，也可以适应对人工修复后重新投入运行的模块的重构。

识别到重构请求以后，系统将会进入工作现场恢复状态，但并不立即开始现场恢复过程。正常机器只是设置一个重构标志，即开始运行正常的任务程序，待恢复机器则保持在重构等待状态。进入空闲时间后，正常机器才会将它们的关键数据区与关键状态通过三机数据交换通道发送给要恢复的故障机器，故障机器对接收数据进行正确性校验后，用它来恢复自身的执行现场。当系统空闲时间结束时，无论重构恢复是否全部完成，系统都会退出重构恢复过程，进入正常的任务执行过程。再次进入空闲时间后，才会重新启动现场恢复的数据传送过程。

现场恢复完成后，故障机器重新进入一个正常的同步等待状态，等待与其他机器一起开始三模冗余的正常执行过程。

4 基于微包协议的重构恢复算法

系统重构的核心问题是将正常机器的内存数据与机器状态复制到故障机器上，使其能够恢复到与正常机器相同的状态。为了有效地利用系统任务的空闲时间，并能够提供较强的系统扩展能力，本文将重构恢复数据分成适当的小包，结合不同状态重构的需要，制定一套具有扩展能力的重构恢复命令包协议，然后在此基础上研究一种无缝重构恢复算法。

4.1 重构恢复命令包协议

协议制定时，首先根据三机数据交换速率确定数据包的大小，使一个数据包可以在 0.1ms－0.5ms 的时间内完成传送。当剩余的空闲时间不足时，禁止重构数据的传送。这样就可以使得重构数据传送时间得到较为精确的控制，尽可能地减小重构恢复模式下数据传送对系统任务的影响，同时又能够最大限度地利用系统的空闲时间进行重构恢复。考虑目标系统三机数据交换速率为 8Mbps，我们将重构恢复命令包的最大长度规定为 128 字节，包括接收端协议解析时间的延迟，单个命令包的最大传输时间约为 0.2ms。要求剩余空闲时间大于 0.5ms 时才进行命令包的传送。

为了提高数据交换通道的利用率，数据包采用如图 2 所示的可变长格式，最长不超过 128 字节，包括类型字、长度字、目标地址、数据和校验字节 5 个部分。类型字表示数据包的类型，它是一个重构命令控制字，使得数据包可以用来恢复故障机器的不同状态或数据区域；长度字表示数据包从类型字开始到校验字结束的总长度；目标地址域在内存数据包中表示数据存储的起始地址，在其他类型包中可以用来表示短信息；数据只在数据包中以可变长形式出现，用来传送内存数据，在其他类型包中该字段长度为 0；校验字是数据包中除校验字以外所有数据的字节累加和，用来在接收端对数据包传输的正确性进行校验。

类型字	长度字	目标地址	数据	校验字
2字节	2字节	4字节	0~119字节	1字节

类型字定义：
AC01：时基数据恢复命令包
AC02：数字量输出状态恢复命令包
AC05：模拟量恢复命令包
AC10：内存数据恢复包
……
CA55：重构结束命令包

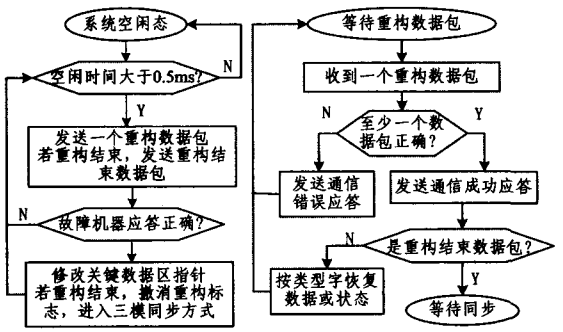
图 2 重构命令包格式

根据目标机器的实际需要，目前定义了图 2 所示的时基数据恢复、数字量输出状态恢复、模拟量恢复、内存数据恢复、重构结束命令包数据格式。

有了如上的协议包规定，重构软件可以分成标准框架和诸多相应的命令程序段两部分。标准框架完成协议包的接收与命令解析，根据不同的命令通过命令转移表调用不同的恢复子程序，完成相应的恢复工作。当扩展系统时，只需要增加相应的命令包定义，并扩展相应的子程序即可。这使得系统重构软件可以采用面向对象方法以标准构件库的形式进行开发，可以极大地提高系统的标准化程度。

4.2 无缝重构恢复算法

故障机器完成上电自检后，通过同步模块向两个正常机器发出带有重构请求标志的同步请求序列。正常机器在同步过程中，识别故障机器的重构请求标志。一旦识别到重构请求后，向故障机器发送带有重构允许的同步应答字，使系统完成重构识别过程，进入重构恢复状态。图 3 给出了重构现场恢复算法的流程图。下面从正常机器与故障机器两个方面对算法进行描述。



(a) 正常机器重构工作流程　　(b) 故障机器重构工作流程

图 3　系统重构数据恢复流程

对于两个正常计算机模块，它们在识别到重构请求后，设置一个重构标志 RCF，继续执行正常的系统任务。待系统任务执行完成进入空闲状态后，若 RCF 标志有效，说明有故障机器需要重构，进入重构流程，按以下步骤执行系统重构过程。

(1)若剩余的空闲时间小于 0.5ms，不进行重构恢复，返回系统空闲态；若剩余时间大于 0.5ms，执行步骤(2)。

(2)向故障机器发送一个重构命令包(可以是内存数据包，也可以是状态恢复命令包)，然后等待故障机器的应答。

(3)若故障机器的应答不正确，返回步骤(1)，执行下一次重构命令包发送过程。若故障机器返回正确应答，转步骤(4)。

(4)成功完成一个重构命令包的传送后，将关键数据区及重构状态指针进行相应的修正，转步骤(1)。

如果所有需要重构的数据均已传送完成，则在步骤(2)构造一个重构结束命令包发送给故障机器，同时在步骤(4)清除 RCF 标志，将本机同步模式改为三机同步模式，返回系统空闲态，等待下次同步时完成系统的重构恢复。

故障机器收到正常机器的重构允许应答同步字后，进入重构等待状态，等待正常机器进入空闲态后发送的重构命令包。重构恢复处理过程如下：

(1)故障机器进入重构状态后，一直等待接收其他两个正常机器的重构命令包；

(2)收到一个重构命令包后，故障机器按照协议约定，对命令包进行正确性校验。若两个机器的重构命令包均不正确，向两个正常机器回送通信故障应答字后，转步骤(1)，继续等待新的重构命令包。若至少一个机器的重构命令包是正确的，则向正常机器回送通信成功应答字，转步骤(3)；

(3)若收到的是重构结束命令包，设置三机同步工作模式，转步骤(5)，否则转步骤(4)；

(4)解析重构命令包的类型字，根据重构命令转移表调用相应的子程序，恢复相应的机器状态或内存数据区。然后转步骤(1)，继续等待新的重构命令包；

(5)退出重构状态，发送本机的同步请求，进入同步等待状态。

上面给出的基于微包协议的重构恢复算法，利用系统任务的空闲时间进行数据恢复，是一种后台恢复模式。如果需要进行集中恢复，可以在系统进入空闲状态以后，让重构程序停止系统时间片定时器，使系统的剩余空闲时间保持不变，待重构恢复完成后，再重新启动时间片定时器，从而实现集中重构。

5　关键数据区管理

关键数据区是指系统重构时需要在故障机器上恢复的那些内存区域，一般由全局变量、静态数据及任务堆栈等重要数据构成，可以是一个连续的内存区域，也可以由多个内存数据块组成。通过仔细选择关键数据区，可以大大减少系统重构时的数据传送量，降低对三机数据交换速率的要求，有效减少系统重构时间。

在上面讨论的后台数据恢复算法中，如果关键数据区中某些数据在传送后发生变化的话，将导致重构恢复的数据不正确。如何对关键数据区进行有效管理，保持正常机器与故障机器数据的一致性，是前述重构算法必须研究的重要问题。

本文采用单向链表方式按更新频度对关键数据块按队列进行管理。在重构程序中，为每个关键数据块设置一个包括数据块地址范围、恢复标志(RF)及更新标志(UF)等内容的数据块表项，并将它们按加入的先后顺序进行排队。数据恢复时，从队首到队尾依次恢复所有的数据块。初始时，所有表项的 RF 均为 0，UF 均为 1。

开始恢复某一数据块时，在其表项中将 RF 置为 1，将 UF 置为 0。恢复完成后，再将 RF 清为 0。恢复过程中，若发现其 UF 变为 1，则认为该数据块已被更新，立即停止对它的恢复。清除 RF 标志，将其移至队尾，再去恢复下一数据块。

某一数据块被更新时，将其 UF 置为 1。若它的 RF 不为 1，将其移至队列末尾；若 RF 为 1，说明该数据块正在被恢复，暂不改变它在队列中的位置。恢复程序会在检查 UF 标志时再将它移至队列末尾。

发送重构结束命令包之前，恢复程序需要重新检查队列中所有数据块表项中的 UF 标志。只有当所有数据块的 UF 均为 0 时，才能认为数据恢复全部完成。若某一数据块的 UF 标志为 1，需将其移至队尾，重新开始对它的恢复。

上述的关键数据区管理算法将频繁更新的数据块放到队尾，推后它们的恢复时间，可以有效地减少对频繁更新的数据块的传输次数，从而有效地节省重构耗费的时间，也可以保证

重构时故障机器拷贝数据与正常机器的一致性。

结束语 采用上述重构算法，在目标机上将任务周期设为100ms，重构数据量50kB(其中8kB数据在每个任务周期都会变化，另有10kB数据每3个周期变化一次，其他数据每8周期变化一次)，任务空闲时间约为18%。程序执行中，通过软件方式在某一机器上注入故障，使其不能正常同步，导致系统通过开关电方式对它进行故障重试，从而对系统重构过程进行测试。测试结果如表1所列。

表1 系统重构测试结果

系统恢复时间	5.1s
重构恢复时间	607ms
时基恢复精度	<4us

表中系统恢复时间包括故障机器的关电恢复与上电启动时间，其中关电恢复时间为3s，上电复位时间为1.2，其他时间为自检、重构识别、重新同步及重构恢复的时间总和。重构恢复时间是从第一个重构命令包开始传送到最后一个重构命令包传送结束的时间。时基恢复精度是系统重构完成后，三机时基定时器的误差，主要是由于通信延迟及三机同步误差引入的。

从测试情况可以看出，本文提出的重构算法可以利用系统空闲时间与串行数据通道，在不中断系统工作的情况下，对三模冗余容错计算机进行无缝重构。这种重构算法，不需要增加特别的硬件支持，也不会增加系统的软件时间开销。同时，完善的重构命令协议使得重构软件的开发可以采用面向对象思想构成标准化的中间件，具有很强的扩展能力，对构建标准化嵌入三模冗余容错计算机内核奠定了良好的技术基础。

参 考 文 献

[1] Nakamikawa T, Morita Y, Yamaguchi S. High Performance Fault Tolerant Computer and Its Fault Recovery[C]// Fault-Tolerant Systems 1997 Proceedings, Pacific Rim International Symposium on. 1997:2-6

[2] 李海山，欧中红，杨升春，等. 基于COTS的容错服务器及其故障恢复技术[J]. 计算机工程，2007,33(8):253-255

[3] 陈文赛. 一种高可靠、高安全性系统—三取二计算机系统[J]. 现代雷达，2004,26(6):19-21,32

[4] 刘天田，袁由光，杨升春，等. 一种TMR容错服务器永久故障恢复机制的研究与实现[J]. 舰船电子工程，2005,25(1):56-58,130

[5] 郭浩翔，袁由光. 一种三模冗余容错服务器的容错机制[J]. 舰船电子工程，2003,1:22-24,34

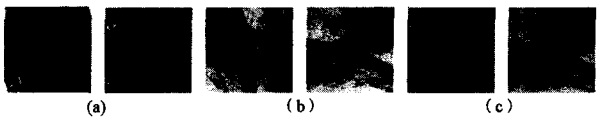
[6] Yu Shu-Yi, McCluskey E J. On-line Testing and Recovery in TMR Systems for Real-Time Applictions[C]// Test Conference Proceedings, International. 2001:240-249

[7] 黎忠文. 嵌入式实时系统容错集成技术的研究[J]. 计算机科学，2006,33(5):277-281

(上接第281页)

4.3 基于SIFT特征点的图像配准实验

图6(a)、(b)分别为待配准的两组图像，图6(c)中给出了其基于SIFT的配准结果。表1为配准的数据及与文献[2]中提出的PLPFFT算法的比较。可以看出，基于SIFT的图像配准方法精度更高，优于现有的PLPFFT算法。



(a) (b) (c)

图6 图像配准实验

表1 图6图像的配准实验数据及比较

Images	Correct (Scale,θ)	Computed (Scale,θ)	
		PLPFFT	SIFT
(a)	(0.6536,15°)	(0.6479,14.9414°)	(0.6509,15.0182°)
(b)	(0.76,62°)	(0.7724,61.875°)	(0.7662,62.0233°)

4.4 基于SIFT特征点的图像拼接

基于特征点的图像拼接是图像配准的一个重要应用[9]。图7是两幅256×256具有平移和旋转关系的花朵图像的拼接结果。两幅图像间的实际仿射参数为$\sigma=1,\theta=45°$。SIFT仿射参数计算结果为$\sigma=1.005,\theta=44.9612°$。从图中可以观察出，基于SIFT的算法实现了良好的图像拼接效果。

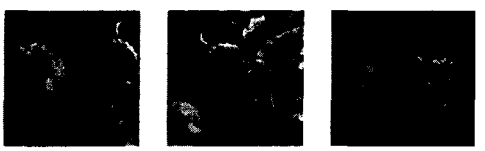


图7 图像拼接效果

结束语 本文研究了3类主要的基于图像特征点的仿射参数计算方案。介绍了SUSAN，Harris和SIFT 3种特征点提取方法，并给出了一种新的基于Zernike矩的特征点匹配算法。分别利用3种特征点进行仿射参数的计算，并针对其性能进行了比对和分析。SIFT是一种比较新的特征点提取算法，目前国内对其研究还相对较少。综合本文的验证与分析，我们认为SIFT特征点是性能最优越的一种图像局部不变特征点。SIFT特征点具有非常好的独特性、匹配精度高，且具有较快的计算速度，具有较为广阔的应用前景(如模式识别、图像理解等)，运用在对实时性要求较高的相关应用中。

参 考 文 献

[1] Zitová B, Flusser J. Image registration methods: a survey[J]. Image and Vision Computing, 2003,21(11):977-1000

[2] Liu Hanzhou, Guo Baolong, Feng Zongzhe. Pseudo-Log-Polar Fourier Transform for Image Registration[J]. IEEE Signal Processing Letters, 2006,13(1):17-20

[3] 罗述谦，周果宏. 医学图像处理与分析[M]. 北京：科学出版社，2003:140-201

[4] Smith S M, Brady J M. SUSAN-a new approach to low level image processing[J]. International Journal of Computer Vision, 1997,23(1):45-78

[5] Harris C, Stephens M. A combined corner and edge detector[A]// Proceedings of the 4th Alvey Vision Conference[C]. Plessey, United Kingdom, 1988:147-151

[6] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110

[7] Khotanzad A, Hong Y H. Invariant image recognition by Zernike moments[J]. IEEE Transactions on PAMI, 1990,12(5):489-497

[8] 卢力，王勇涛，田金文，等. 基于SUSAN算法的遥感图像去云[J]. 通信学报，2006,27(8):160-164

[9] 仵建宁，郭宝龙，冯宗哲. 一种基于兴趣点特征匹配的图像镶嵌技术[J]. 光电子·激光，2006,17(6):733-737