第 36 卷 第 8 期 计 算 机 工 程 2010 年 4 月
Vol.36 No.8 Computer Engineering April 2010
·开发研究与设计技术· 文章编号：1000—3428(2010)08—0267—03 文献标识码：A 中图分类号：TP338.7

实时并行系统的性能及其测试方法

郝丽蕊[1]，薛弘晔[2]，覃 科[1]
(1. 桂林航天工业高等专科学校计算机系，桂林 541004；2. 西安科技大学计算机系，西安 710065)

摘 要：实时并行系统是并行/分布系统的重要组成部分，它要求系统计算时间具有可预测性，即符合应用要求的硬性响应时间约束。根据该特点，分析影响实时并行系统性能的各种因素，设计并实现一种简单的实时并行系统评测方案，并在实验室对 C3I 实时集群系统进行评测，从而验证该方案的有效性。

关键词：实时并行系统；测试；性能；可预测性；延迟时间

Performance and Test Method for Real-time Parallel System

HAO Li-rui[1], XUE Hong-ye[2], QIN Ke[1]
(1. Department of Computer Science and Technology, Guilin College of Aerospace Technology, Guilin 541004;
2. Department of Computer, Xi'an University of Science and Technology, Xi'an 710065)

【Abstract】 Real-time parallel system is important part of parallel/distributed system, which requires predictability for system computing time, and fits the response time constrait. According to this character, the real-time functions of real-time parallel system are analyzed, and a simple test method of real-time cluster computing is designed and realized. The test process and results of real-time cluster computing in laboratory are given, which proves it is an effective test method.

【Key words】 real-time parallel system; test; performance; predictability; latency time

1 概述

随着实时并行/分布系统在军事指挥控制、机器人、自动运载器等领域的应用日益广泛以及需要处理的信息量和处理结果期望值的提高，实时并行系统的性能和安全性受到越来越多的关注。但在测试基准领域，类似 SPEC,AIM,LINPACK, Whetstone,Dhrystone 这样的标准化基准被用于衡量通用分时计算环境的性能，并不能真正反映实时系统本质特征。

本文分析实时并行系统的本质特征，设计实时并行/分布环境性能评测方法，并通过它对现有的实时集群系统实施相应的技术分析和测试，开发这一方法，分析实时并行系统的本质特征动机，从而帮助系统结构设计者和软件开发人员更好地理解影响并行系统中的实时应用的因素，为实时集群设计者和使用者提供一种有效的评测方法。

2 实时并行系统的时效特征

实时系统一个重要特征就是系统对硬性计算截止期约束，超过截止期的计算将使系统效用(utility)无意义或下降[1]。在低层的集中式的实时控制系统中，硬性截止期的满足主要通过确定的中断服务和系统服务的响应时间来实现。在分布/并行实时应用系统中，对计算截止期的满足程度的预测就要复杂得多，低级中断服务和系统服务的响应时间仅是所有因素中权重不高的一部分。影响分布/并行实时系统响应时间的因素主要有操作系统任务调度和应用程序 2 个方面因素。

2.1 操作系统因素对响应时间的影响

假设调度器调度任务的时间开销是 C_{sched}，调度器周期为 T_{sched}，中断周期为 $T_{interrupt}$，中断处理时间为 $C_{interrupt}$，考虑操作系统因素影响的并行实时系统响应时间 R_i 为

$$R_i = C_i + \max C_k + \left\lfloor \frac{R_i}{T_{sched}} \right\rfloor C_{sched} + \sum_{v=1,2,\cdots} \left\lfloor \frac{R_i}{T_{\text{int}\,errupt_v}} \right\rfloor C_{\text{int}\,errupt_v} \quad (1)$$

式(1)中的项目依次为：任务调用 i 的计算时间，任务调用 i 受其他 SCHED_FIFO 模式调用的阻塞时间，定时器驱动的调度器开销以及零散的中断服务 v 打断任务调用 i 执行的时间。

2.2 应用任务因素对并行实时系统响应时间的影响

为了简化软件系统结构，非周期性任务和零散任务都被转化为周期性任务。这样，总的任务计算时间开销 C 为

$$C = \sum(\max C_{comp_x} + \max C_{comm_x} + C_{syn_x}),\ x=1,2,\cdots \quad (2)$$

每一个时段经历计算 C_{comp}、通信 C_{comm}、路障同步(超步计算的触发开销)C_{syn}，式(1)、式(2)分别反映操作系统因素和应用任务因素对实时集群响应时间的影响，合并 2 个式子，就能够得到考虑了各种影响因素的响应时间：

$$R_i = \sum(\max C_{comp_{i,x}} + \max C_{comm_{i,x}} + C_{syn_{i,x}}) + \max C_k + \left\lfloor \frac{R_i}{T_{sched}} \right\rfloor C_{sched} + \sum_{v=1,2,\cdots} \left\lfloor \frac{R_i}{T_{interrupt_v}} \right\rfloor C_{interrupt_v} \quad (3)$$

综合并行计算性能评价指标(加速比、效率、可扩展加速比)、硬件体系结构(并行结构互连网络性能)、并行算法(粒度、消息量、同步路障)等方面的因素[2]，进一步分析影响响应时间的确定性和不确定性因素，并行性能主要受并行计算实现过程中的额外系统开销 W_o 的影响。W_o 包含网络设备反应时间(T_{comm} 的基本部分)、消息量带来的传输延迟(T_{comm} 的可变部分)、计算中处理器的空闲时间(T_{idle})、操作系统的额外管理开销(T_{os})，W_o 的影响因素如表 1 所示[3]。

基金项目：国家科技部创新基金资助项目(01c26226111003)
作者简介：郝丽蕊(1975－)，女，讲师、硕士，主研方向：网络与高性能计算技术；薛弘晔，副教授；覃 科，讲师、硕士
收稿日期：2009-12-10 **E-mail：**lirui_h@sina.com

表 1　影响 W_o 的因素

W_o	硬件		算法		系统
	响应时间	链路速率	并行粒度	消息量	任务调度
T_{comm}-Latency	√				
T_{comm}-链路延迟		√		√	
T_{idle}-空闲时间			√		√
T_{os} -系统开销					√

3　实时并行系统评测标尺

对于实时并行系统来说，硬件、软件和体系结构的工作机制，都是影响系统性能的因素。基于上面讨论的实时系统的本质，定义描述性能标尺。由于实时系统的复杂性，单纯的量化标尺很难反映一个实时系统的适应性，因此考虑量化尺度以外还增加一种方法：分级定性标尺[4]。根据实时并行系统特征定义评测标尺如表 2 所示。

表 2　实时集群系统评测标尺

	II	III	IV	评测标尺	
				定性分级	定量参数
I	软件	操作系统	切换时间		•
			中断时间		•
		中间件	计算粒度	•	
			负载策略	•	
			消息传递延迟		•

性能标准分为 2 组：定性标尺和定量标尺(形成一些可度量的参数)。分级定性标尺能够比较 2 个不同对象的相对特征(优或差)，但不能准确量化对象特性；量化尺度能够准确给出对象特性的定量参数，如处理器的 MIPS 参数。

4　测试方法及结果分析

4.1　操作系统性能测试

4.1.1　进程切换延迟的测试

进程之间的切换是衡量操作系统反应灵敏度的重要指标。由于操作系统结构和运行过程的复杂性，因此很难精确测量这些数据。本文根据 Rhealstone 测试框架和软件执行时间测试的有关思想，实现了对 Linux3.0 操作系统实时性能测试。测试的基本思想如下：

Step1 创建父进程。

Step2 取当前时钟 *T*1。

Step3 父进程循环执行进程切换原语 Schedule()*n* 次。

Step4 取当前时钟 *T*2。

Step5 创建子进程。

Step6 取当前时钟 *T*3。

Step7 父进程调用 Schedule()，将执行权切换给子进程；子进程再调用 Schedule()，将执行权切换给父进程，如此循环 *n* 次。

Step8 取当前时钟 *T*4。

进程切换时间：$T0=\{T4-T3-(T2-T1)\}/n$，测试结果见表 3。

表 3　测试结果　　μs

测试指标	1	2	3	4	5
进程切换时间	1 792	1 745	1 720	1 735	1 696
中断响应延迟	2 609	1 589	5 680	7 776	6 541

4.1.2　中断响应时间

该延迟是指从内核接收中断到中断处理例程第一条指令执行之间的时间。测试时，先将一随机负数加载到实时时钟计数器中。时钟计数到零时产生中断，并继续计数[5]。用中断处理例程的第 1 条指令读时钟值，即为中断延迟时间，测试结果见表 3。

从测试结果可以看出：Linux3.0 在其系列平台上运行时的几种主要实时性能指标是微秒量级的。所测结果是统计平均值，它远小于实时任务的执行时间(毫秒级)。如果实时软件系统有良好的结构，则操作系统的开销是完全可预测的。

4.2　点到点的通信性能

因为系统的实时性是以消息传递系统性能来保证的，所以有必要对消息传递系统的传输延迟进行测试。

用户数据包协议(UDP)支持广播发送和多点消息传送，在通信前不需要建立连接，系统开销较小。因此，在实时集群计算机中，采用用户数据包协议，简化了消息传递机制，增强了系统响应的实时性。但 UDP 数据包协议属于不可靠连接，需要进行测试分析其可靠性。

本文采用用户数据包协议对节点控制的实时集群系统进行网络数据传输可靠性和延迟测试。测试使用一个乒乓小型基准测试程序。进程“乒”发送一条消息，然后再从“乓”接收回来。“乒”测量执行发送/接收对所需要的时间，即来回时间。进程“乓”做的是反向操作。这种动作重复若干次得到统计结果，把最初的几次测量结果丢弃，以排除“热身”的影响。延迟时间 *T* 用平均来回时间的一半来计算。

(1)测试环境

硬件系统：集群计算机系统由高性能浪潮 NF300IA 构架服务器构成，通过千兆光纤以太网采用 UDP 协议传输。其配置规格为：2×IntelPIIIXeonIG 处理器、512 MB ECC 存储器、2×18 GB 高速硬盘、Intel PRO 1 000 Mb 网卡。

软件环境：由可裁剪的具有实时性的 REDLinux3.0 和 C 语言开发平台 GUN C。

(2)测试结果及分析

设对集群工作有效的数据包长度为 *L*，有效数据包数量为 *N*，加载无效的广播数据包数量为 *n*(模拟网络负载)，无效的广播数据包固定长度为 2 000 Byte。定义 *R* 为加载无效数据包数量与有效数据包数量之比，即 $R=n/N$(表征网络负载状况)；定义有效数据包丢失率为 *M*，有效数据包传输顺序正确率为 *S*，节点间数据传输延迟为 *T*。集群网络数据传输延迟测试是在不同网络链路负载下，测试不同长度的数据包在集群节点间的传输延迟。集群网络数据传输可靠性测试是在不同网络链路负载、不同包长度下对集群节点间数据包传输的丢失率和数据包传输的顺序正确率进行测试，统计测试结果如表 4 所示。

表 4　千兆以太网点对点数据传输延迟测试

数据帧长度 /Byte	延迟 *T* /ms	有效丢包率 *M*	有效包正确率 *S*/(%)	网络负载 *R*
1 000	0.299	0	100	0
2 000	0.317	0	100	0
3 000	0.330	0	100	50
4 000	0.347	0	100	50
5 000	0.397	0	100	100
3 000	0.341	0	100	100
2 000	0.345	0	100	50
1 000	0.357	0	100	100

由表 4 可知：(1)不同网络负载、不同包长度对节点间数据传输延迟影响不大，都在 0.3 ms 级。(2)不论集群系统在空载(R=0)时，还是在负载较重(R=100)时，不同包长度的数据包传输丢失率(M)均为 0，数据包传输顺序正确率(S)均为 100%。

由此表明用户数据包协议的传输可靠性能够在网络负载不重、不发生网络拥塞时得到保证，尤其在局域网中，其可靠性较高。

4.3 进程计算时间及负载平衡测试

实时并行计算中的所有进程全部消耗的时间由计算时间、通信时间和等待时间 3 个部分构成。用户根据计算和通信预测时间，可以适当选择问题粒度，获取期望性能。负载平衡可减少计算机之间的通行时间和所有进程之间的等待时间，即减少计算机间数据交换时的耗损。本文通过在实时集群中嵌入计算节点评估模块实现对节点计算时间监控和负载平衡监视。

计算节点评估模块由时统信号、计算进程、显示进程组成。计算节点评估模块工作原理：在集群中的计算节点，开始计算和计算完毕各取一次系统的时间，并将同步周期一同发给主控制台，主控将其写入指定的共享内存。集群负载平衡工具根据发来的数据，计算出计算节点的计算周期，用它来和集群的时统周期相比较，并在界面显示出来。

集群负载平衡工具实现如下：

Step1 创建或打开共享内存(首次为创建)。

Step2 创建或打开消息队列(首次为创建)。

Step3 接收时统信号。

Step4 计算时统周期并且求和。

Step5 取本机时间和时统信号基准时间存入共享内存。

Step6 判断：若时统周期和小于 1 000 转 Step7；否则转 Step3。

Step7 发消息给计算进程。

Step8 显示上周期数据，转 Step3。

计算节点评估模块的主要功能是：(1)评估各个计算节点在一个时间周期内，不同的调度机制(FIFO,OTHER)，相同或不同的实时优先级下，各个计算节点的事件时间关系，并且以字符的形式将各节点的执行时间显示出来。(2)将运行过程中的节点计算时间记录在文件中，软件开发人员可以分析各种运行状态(数据仿真、状态仿真、容错仿真等)下计算完成时间分布，从而评估应用要求的计算截至期的符合程度。(3)通过集群系统负载监视工具来实现负载平衡监视。

图 1 是主控在某一时刻捕捉的集群计算节点负载平衡情况。其中，横轴代表时统周期，纵轴为各计算进程，显示的是各进程在一个计算周期内实际的任务执行时间，为了更便于了解真实情况，这里采用节点完成每一周期计算任务的时间与时统周期相对比的方式来刻画节点的负载情况，而不用 CPU 的利用率这一模糊的概念描述负载情况。

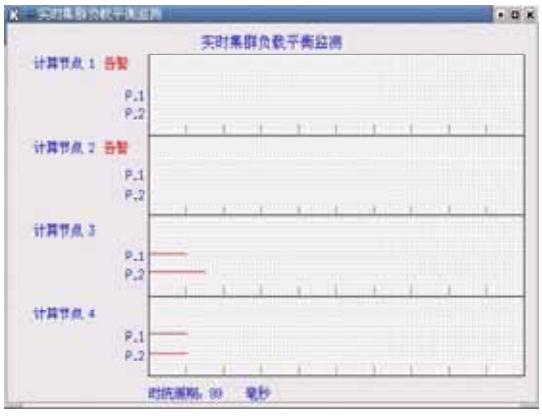

图 1　集群系统负载监视界面

完成任务的时间与时统周期的比值越小说明节点的计算裕量越大，反之越小，当超过“ 1 ”时，说明节点已不能正常工作。图中 4 个计算节点完成每一周期计算任务的时间与时统周期的比值小于 0.3，即计算节点仍有较大的计算裕量。各节点之间计算时间几乎均等，说明负载平衡。

5　并行环境应用的评测

并行环境应用的评测主要是评测并行环境对应用的适应性。应用级测试程序框架由 C3I 应用程序、应用环境仿真数据/事件发生器、运行结果记录和评测程序等 3 个部件构成，如图 2 所示。

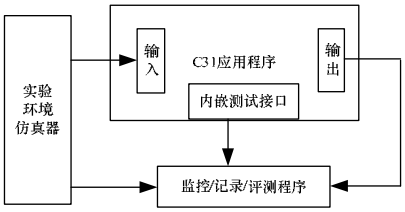

图 2　应用级测试程序框架

从负荷的角度出发，将应用的测试分为恒定负荷和可伸缩负荷 2 种情况。恒定负荷测试产生用于性能对比的量化参数，可伸缩测试验证系统对特定应用的适应性。C3I 应用程序是一套完整的经过应用验证的多时段、多源信息处理程序包。为了测试的目的，在 C3I 相关程序模块中嵌入了测试接口，通过这个接口向评测程序提供应用程序的运行状态。根据评测目的，实验仿真数据/事件发生器模拟实际外部传感器的工作情况，产生相应的数据和事件。在判定并行环境某一个给定应用的适应性时，测试者通过配置试验发生器，使得计算在时间和负荷 2 个方面接近实际应用，通过分析该程序在测试环境中的运行结果，评测并行环境对应用的适应性。监视/记录/评测模块给出测试结果。

在评测并行环境的服务质量时，实验环境仿真器程序模拟产生各种故障、超时和丢失，由评测程序检验在异常情况下系统的运行情况。

6　结束语

本文以一个实时集群系统平台以及 C3I 应用平台环境作为研究实时并行评测的实体，以应用对象特征为出发点，根据实时集群系统行为特征展开对实时集群系统的评测。本文测试方法已成功用于指控集群计算机系统(国内第 1 套投入实用的实时集群系统)的开发和验收阶段测试工作，证明了其有效性。

参考文献

[1] 林　闯. 计算机网络和计算机系统的性能评价[M]. 北京: 清华大学出版社, 2001.

[2] 郑纬民, 石　威. 高性能集群计算[M]. 北京: 电子工业出版社, 2001.

[3] 杜　鸿, 向建军, 白　欣. 一种实时集群综合平台[C]//2004 年全国博士生学术论坛会议论文集. 长沙: [出版者不详], 2004.

[4] Wolfgang A. Measuring the Performance of Real-time Systems[Z]. (1997-02-14). http://www.soe.uoguelph.ca/webfiles/rmuresan/MeasuringPerformanceOfRealTimeSystems.pdf.

[5] 毛羽刚, 金士尧, 凌云翔. DEC UNIX 操作系统的实时特性及评测[J]. 国防科技大学学报, 1999, 21(4): 71-74.

编辑　陈　文