广西科学院学报
Journal of Guangxi Academy of Sciences

2008,24(4):296~299
Vol. 24,No. 4 November 2008

基于 Petri 网的动态负载平衡双层调度模型研究
Based on Petri Net Model for Dynamic Load Balancing Double-decked Scheduling

杨夏妮[1],覃海生[2]
YANG Xia-ni[1],QIN Hai-sheng[2]

(1. 玉林师范学院数学与计算机科学系,广西玉林 537000;2. 广西大学计算机与电子信息学院,广西南宁 530004)

(1. Department of Mathematics & Computer Science, Yulin Normal University, Yulin, Guangxi, 537000, China; 2. School of Computer, Electronics and Information, Guangxi University, Nanning, Guangxi, 530004, China)

摘要:根据集中式和分布式动态负载平衡调度方式的优点,提出一种动态负载平衡双层调度模型(DLBDSM),并在 Petri 网上进行建模。该模型将分布式系统分成若干相对独立的任务调度组,调度组由 1 个调度服务器和 3 个工作站组成,组内采用集中式调度,组间采用分布式调度,顶层子系统和底层子系统分别由每个任务调度组的调度服务器和工作站组成。与现有的动态负载平衡调度模型对比,DLBDSM 模型具有易实现、易管理和实时性等优点,并能有效地减少任务迁移所带来的系统开销。

关键词:调度模型 动态负载平衡 分布式系统 Petri 网
中图法分类号:TP311 **文献标识码**:A **文章编号**:1002-7378(2008)04-0296-04

Abstract: According to the behavior of the centralized and the distributed dynamic load balancing, the dynamic load balancing double-decked scheduling model (DLBDSM) is proposed and modeled by the theory of Petri net. The model divided the distributed system into several relative independent task scheduling groups, which each task scheduling group was made up by a scheduling server and three workstations. The centralized scheduling was used within the group and the distributed scheduling was used among groups. The top subsystem and the bottom subsystem is made up by scheduling server and workstations of each task scheduling group respectively. Compared with the existed dynamic load balancing scheduling model, DLBDSM has several advantages of manageability such as easy-to achieve and real-time, and decreases the system expense of task migrations effectively.

Key words: scheduling model, dynamic load balancing, distributed system, Petri net

自上世纪 80 年代以来,计算机网络以前所未有的速度向前发展,Internet 也渗透到社会生活的各个领域中,使我们充分享受到了网络所带给我们的高效、方便与快捷。随着用户数量的不断增多,传统的网络体系结构日趋接近其处理能力的物理极限。到了 20 世纪 90 年代,计算机系统开始由集中式向分布式的发展。对于分布式系统,由于各结点处理能力存在差异,当系统运行一段时间后,某些结点分配的任务很多(称为重载),而另外一些结点却是空闲的(称为轻载)。为了克服这种现象发生,需要采用一个有效的负载平衡调度系统[1,2]。

通常,负载平衡分为静态负载平衡和动态负载平衡。只是利用系统负载的平均信息,而忽视系统当前的负载状况的方法被称为静态负载平衡。根据系统当前的负载状况来调整任务划分的方法被称为动

收稿日期:2008-01-26
修回日期:2008-09-20
作者简介:杨夏妮(1980-),女,硕士研究生,主要从事计算机网络与并行分布式计算研究。

态负载平衡。它更能反映分布式系统的实际情况,提高系统性能,因此本文只考虑动态负载平衡调度。

一般来说,动态负载平衡调度可以分为集中式调度和分布式调度两大类。集中式调度是由一个调度服务器负责搜集系统负载信息,并由它来决定负载平衡调度方案,它的主要优点在于实现比较简单,但是在结点数较多的大规模并行分布系统中,由于各结点与调度服务器的通讯成为瓶颈,所以调度开销比较大[3~6]。所以,除非结点数目较少(典型的系统有 4 个结点),或者在底层硬件系统中采取比如超级集线器这样的一些特殊实现措施[7],否则在分布存储的并行系统中不大采用集中式平衡调度方法。分布式调度是根据局部范围内的一些负载信息来进行负载平衡操作和调度的。每台计算机定期把它的负载信息广播给其它计算机,去更新那些局部维护的负载向量。它的最大优点在于具有良好的可扩放性。根据集中式和分布式调度的这些特点,本文提出一种动态负载平衡双层调度模型(DLBDSMD),将分布式计算机系统分成若干相对独立的任务调度组,调度组由 1 个调度服务器和 3 个工作站组成,组内采用集中式调度,调度组之间采用分布式调度,顶层子系统和底层子系统分别由每个任务调度组的调度服务器和工作站组成。

Petri 网作为系统建模的工具,具有图形直观、能描述冲突和真并发,且能以状态分布表示,对并行性、不确定性、异步和分布式系统具有较强的描述和分析能力,被认为是迄今研究系统性能的最有力的工具[8]。因此,本文利用 Petri 网对模型进行建模。

1 动态负载平衡双层调度模型描述

在负载平衡调度的一般模型[9]中,根据负载平衡调度算法的不同,图的构成原则各不相同(如图 1 所示),我们假设结点表示工作站,两结点间的边表示两台工作站是物理邻接的,整个系统的拓扑结构是由若干相对独立的任务调度组组成的双层结构,组内的工作站两两邻接。由于组内的工作站两两物理邻接,因此可以忽略并行计算任务之间的通信量。

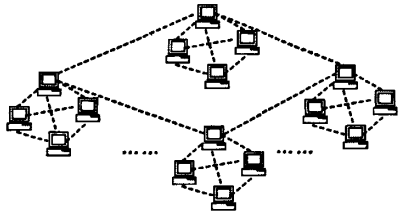

图 1 动态负载平衡双层调度模型

在负载平衡调度的一般模型[9]中,用户提供的任务类型按照对处理器、网络和 I/O 的资源要求可以分为静态文档请求和动态文档请求,任务提交方式有抢占式和非抢占式,对于请求分布并行计算任务,由于计算量较大,因此需要进行任务约束。我们将请求分布并行计算任务分为两种模式,一种是基本模式,即将一个大计算量的任务分解为一些相互独立的子任务,在一组工作站上并行执行,最后将子任务结果汇集为总的结果。另一种为协作模式,即在计算过程中,子任务间需要同步,交换中间结果,因而要求基本上同时开始,并以同样速度执行。

对于负载平衡调度一般模型的负载指标,我们选择目前大多数系统使用的 CPU 队列的长度作为评价一个结点负载的重要标准[10]。对于负载平衡调度一般模型的负载平衡策略,DLBDSM 的转移策略采用基于 CPU 队列长度的阈值转移策略,参与转移的结点要么是发送者,要么是接收者。设 CPU 队列长度的阈值为 T,当结点产生一个新任务使得负载大于或等于阈值 T 时,转移策略将该结点确定为发送者;当结点完成一个任务后使得负载小于阈值 T 时,转移策略将该结点确定为接收者。选择策略是在一个发送者启动时,选择一个最新产生的任务进行转移。位置策略与调度方式有很大的关联,由于任务调度组内采用集中式调度,并且结点较少,故采用顺序查找策略,轮流对每个结点进行查找。顶层则采用一种通过探询寻找任务转移伙伴的分散策略,而且探询结点的选择是基于在前面的探询中所收集的信息,这样可以大大减小探询的盲目性,从而提高探询效率。为了防止无休止地探询,设置了探询次数上限和探询次数,在探询接收结点次数达到探询次数上限时将停止探询。为了能够准确地查找到适合转移任务的结点,必须从系统中收集到足够多的信息,任务调度组内只需在某结点上状态发生某种程度的变化时,它就将自己的状态信息告知调度服务器即可,而顶层采用状态变化驱动和需求驱动相结合的策略,当一个结点成为发送者或接收者时,都将自身目前状态信息发送给其它结点,使其它结点及时更新状态向量,知道发送信息结点的最新负载情况;仅当一个结点成为发送者时才去主动探询接收结点的最新状态,以使其成为启动分载的一个合适候选者。

1.1 负载平衡双层调度模型的数据结构

保存负载平衡调度信息的数据结构包括:(1)任务队列 TQ(Task Queue),用来记录等待求解的用户任务,按照 FIFO 的原则添加任务,根据网络负载

的实时状况对这些任务进行统一调度和执行。(2)发送者表SL(Sender List),记录当前系统中重载结点的信息。(3)接收者表RL(Receiver List),记录当前系统中轻载结点的信息。(4)负载信息表LL(Load List),用于存放当前任务调度组中所有结点的负载信息,调度服务器根据该表存放的信息对任务队列TQ中的任务进行负载平衡调度,结点的当前负载状态划分如前所述,分别用1/0表示重载和轻载。

在开始时,顶层的每个结点假定其它所有结点都是接收者,即开始时RL中包含所有的结点,SL为空,而LL中的结点状态均为0,即轻载。

1.2 负载平衡双层调度模型的实现原理

DLBDSM系统包含两类结点,即调度服务器和工作站。由于它们参与负载平衡调度的作用各不相同,因此实现方法也不一样。

调度服务器用任务队列TQ、发送者表SL、接收者表RL和负载信息表LL来管理系统。任务队列TQ用来记录待求解的用户任务。发送者表SL用来记录顶层结点中的重载结点,时刻监视和更新结点的状况,将变化情况及时通知邻近的结点。接收者表RL用来记录顶层结点中的轻载结点,时刻监视和更新结点的状况,将变化情况及时通知邻近的结点。负载信息表LL用来记录同一任务调度组中各个结点的负载状态信息,时刻监视和更新负载信息表,根据负载信息表的实时负载状况统一调度任务队列中的任务。

工作站用任务队列TQ来管理系统。任务队列TQ用来记录待处理用户任务,主要负责执行用户提交的任务。它将任务调度服务器根据统一调度结果安排任务添加到任务队列TQ中;当接收到用户提交的任务或者完成某项任务而引起负载情况发生变化时,它将把负载变化情况及时地通知任务调度服务器。

2 动态负载平衡双层调度模型建立

基于Petri网的动态负载平衡双层调度模型如图2所示。假设DLBDSM系统的每一台计算机均可以接收来自系统用户的任务。对于这个模型我们做如下的约定:(1)新任务请求到达调度服务器的过程为泊松(Poisson)过程。(2)模型对新任务的请求不区分优先级,即请求获得处理的概率是相等的。(3)模型中包含多个服务,它们有不同的服务速率,服务速率是独立的、指数分布的。(4)所有的时间变迁只有在变迁的后向位置未达到缓冲容量的极限时才可以实施。

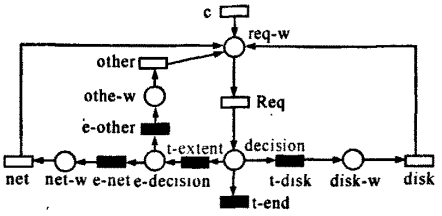

图2 基于Petri网的动态负载平衡双层调度模型

如图2,c是来自当前结点任务请求到来的时间变迁。req-w是等待资源处理进程处理的消息队列。req则是资源处理进程对消息的处理和协调过程。decision表示请求任务发生转换的位置,它瞬时保留到来的请求,根据t-disk,t-extent,t-end联系的实施概率进入不同的后续服务队列。t-disk变迁在当前结点状态处于轻载时触发,它将请求任务转入当前调度组的当前结点的任务队列中。t-extent变迁在当前结点处于重载时触发,它将请求任务转入非当前结点的任务队列中。经过e-decision来决定请求任务转入当前调度组的其它结点或者另一调度组执行。e-net变迁在当前调度组的其它结点中至少有一个结点处于轻载时触发,它将请求任务转入当前调度组与当前结点最接近并处于轻载的结点的任务队列中。e-other变迁在当前调度组的全部结点处于重载时触发,它将请求任务转入另一调度组与当前调度组最接近并处于轻载的结点的任务队列中。disk-w、net-w及other-w分别表示等待当前调度组当前结点、当前调度组与当前结点最接近并处于轻载的结点及等待另一调度组与当前调度组最接近并处于轻载的结点系统处理的任务队列。disk、net及other分别表示当前调度组的当前结点系统、当前调度组与当前结点最接近并处于轻载的结点系统及另一调度组与当前调度组最接近并处于轻载的结点系统对任务的处理过程。t-end变迁在不满足t-disk和t-extent触发条件情况下触发。

3 结束语

要在分布式系统上真正实现负载平衡是一个公认的NP问题,因此本文是针对前人的研究工作所作的改进,以提高系统性能。与现有的动态负载平衡调度模型相比,DLBDSM具有如下特点:

第一,DLBDSM不再是单一层次,而是将系统分成两个层次,综合应用了动态负载平衡调度的两种策略,既体现了集中式调度策略的易实现和管理性,又体现了分布式调度策略的实时性。

第二,DLBDSM 能有效地减少任务迁移所带来的系统开销。系统中采用分而治之的思想,各个任务调度组相对独立,有新任务到来先由任务调度组内部自行解决,当整个任务调度组处于重载状态时才将任务迁移到另一个任务调度组,这就使得迁移所引发的网络流量小。迁移时也采用了就近原则,使得迁移目的地系统与源端尽可能的保持同构性,从这方面看也节约了处理机的时间。

根据 DLBDSM 的特点,开发出相应的 DLBDSM 系统具有一定的实际价值,今后将进一步研究实现智能 DLBDSM 系统。

参考文献:

[1] Marc H, Willebeek-LeMair. Strategies for dynamic load balancing on highly parallel computers [J]. IEEE Transactions on Parallel and Distributed System, 1993, 4(9): 979-993.

[2] Hui Chi-chung, Samuel Chansons. Hydrodynamic load balancing [J]. IEEE Transactions on Parallel and Distributed System, 1999, 10(11): 1118-1137.

[3] Hesham El-rewini, Theodore G Lew is, Hesham HAli. Task scheduling. Englewood Cliffs[M]. New Jersey: PTR Prentice Hall, 1994.

[4] Joosen W, Pollet J. The efficient management of task clusters in a dynamic load balancer: proceedings of the International Conference' 94 on Parallel Distributed Systems[C]. Hsinchu, Taiwan, Dec, 1994: 19-21.

[5] Feng M D, Yuen C K. Dynamic load balancing on a distributed system: proceedings of the 6th Symposium on Parallel and Distributed Processing [C]. Dallas, Texas, Oct, 1994: 26-29.

[6] Chen Hua-ping, Lin Hong, Chen Guo-liang. Heuristic task scheduling in parallel distributed computing [J]. Computer Research and Development, 1997, 34 (Supplementary Issue): 74-78.

[7] Gene Eu Jan, Lin Ming-bo. Effective load balancing on highly parallel multicomputers based on superconcentrators: proceedings of the International Conference'94 on Parallel Distributed Systems [C]. Hsinchu, Taiwan, Dec, 1994: 19-21.

[8] 袁崇义. Petri 网原理与应用[M]. 北京:电子工业出版,2005.

[9] 李冬梅,施海虎,顾毓清. 基于规则的分层负载平衡调度模型[J]. 计算机科学,2003,30(10):16-21.

[10] 袁磊. 分布式数据库系统的动态负载平衡策略及算法设计[J]. 计算机工程与设计,2004, 25(8): 1375-1378.

(责任编辑:邓大玉)

(上接第 295 页)

4 结束语

本文针对传统模板匹配算法存在运算量大以及干扰物中有部分和目标相同时识别率差的问题,通过改进传统模板匹配算法,在 MATLAB 6.5 环境下,运用新方法对在简单背景中的动目标进行识别与跟踪仿真。新方法通过先检测符合要求的参考点后再进行匹配运算,明显减少了运算量,提高了目标识别的实时性。同时,对于干扰物体的形状中部分区域与目标一样的情况,采用自动更新模板的方式,有效地降低了目标的错判率,提高了目标识别率。

参考文献:

[1] 周西汉,刘勃,周荷琴. 一种基于对称差分和背景消减的运动检测方法计算机仿真[J]. 计算机仿真,2005, 22(24): 117-119.

[2] McKenna S, Jabri Z Duric Z. Tracking groups of people[J]. Computer Vision and Image Understanding, 2000, 80(1): 42-56.

[3] Gavrila D M. The visual analysis of human movement: a survey[J]. Computer Vision and Image Understanding, 1999, 73(1): 82-98.

[4] 贾云得. 机器视觉[M]. 北京:科学出版社,2000: 26-235.

[5] 周敬兵. 复杂背景下目标的检测与跟踪技术研究[D]. 南京:南京理工大学,2007.

[6] 田娟,郑郁正. 模板匹配技术在图像识别中的应用[J]. 传感器与微系统,2008,27(1):112-113.

[7] 许波,李正明. 一种新的基于自适应模板的相关跟踪算法[J]. 光学与光电技术,2004,2(4):62-64.

[8] Barron J, Fleet D, Beauchemin S. Performance of optical flow techniques [J]. International Journal of Computer Vision, 1994, 12(1): 43-77.

(责任编辑:韦廷宗)