

使用资源利用率作为 负载均衡系统的负载指标*

鞠九滨 杨 鲲 徐高潮

(吉林大学计算机科学系 长春 130023)

摘要 一个可以正确反映当前系统负载情况的负载指标对一个成功的动态负载均衡系统来说是至关重要的。但是,很少负载均衡系统对其选定给出一个实验说明,而且多数系统使用资源队列长度(即 CPU 队列长度)作为负载指标。本文说明了这种负载指标存在的缺点,并对一种新的负载指标——资源利用率进行了讨论。实验表明,在动态负载均衡系统中,使用资源利用率作为负载指标比使用资源队列长度作为负载指标对资源利用率及作业响应时间的改进更有效。

关键词 动态负载均衡系统,负载指标,资源利用率,队列长度,作业响应时间。

在分布式系统中,负载均衡 LB(load balancing)是将用户提交给系统的工作负载分散到各主机上以提高系统性能。其前提是存在一个能正确反映系统当前负载情况的负载指标,负载指标被负载均衡机制用于决定作业放置的位置。虽然这个领域的研究者们提出并使用了很 多负载指标,但他们多数未对其所用负载指标的合理性给出实验上的证实。

负载均衡的一个重要目标就是提高系统性能,即缩短作业平均响应时间。一个作业的响应时间依赖于其所运行的主机上的负载。负载越重,其运行时间越长。因而,理想的负载指标应与作业响应时间有一种单调关系。作业的执行需要各类计算机资源,主要包括 CPU 周期、磁盘带宽及主存空间。负载均衡的另一重要目标是均匀地、充分地利用整个系统的资源。这个目标与上边提到的目标是一致的。资源的使用越平衡,作业响应时间就越短。

理想的、体现系统负载情况的负载指标应当满足以下条件:(a)测量开销低,这意味着可以频繁测量以确保信息最新;(b)能体现所有竞争资源上的负载;(c)各个负载指标在测量及控制上彼此独立。

Ferrari^[1,2]建议使用各种资源队列长度的线性组合作为负载指标,但其条件是假定系统处于稳定状态,并且要求资源的排队规则是 FCFS,PS 或 LCFSPR。实际上的系统并不满足这些条件。Zhou^[3]建议使用 CPU 队列长度作为负载指标,他发现 CPU 队列长度和磁盘队列长度分别与 CPU 类作业和 I/O 类作业有密切关系。但资源队列长度与资源利用率并没

* 本文研究得到国家 863 高科技项目和国家教委博士点基金资助。作者鞠九滨,1935 年生,教授,主要研究领域为分布式系统与计算机网络。杨鲲,1970 年生,助教,主要研究领域为分布式系统与计算机网络。徐高潮,1966 年生,讲师,主要研究领域为分布式系统与计算机网络。

本文通讯联系人:鞠九滨,长春 130023,吉林大学计算机科学系

本文 1995-02-27 收到修改稿

有紧密的关系. Bonomi^[4]等人使用处理机上活动的进程数的瞬时信息和周期地搜集到的平均 CPU 运行队列等信息的组合作为负载指标, 并且进行了实际测量, 性能上有所改进, 但仍不能正确反映资源利用率. Kunz^[5]通过实验比较了运行队列任务数、系统调用率、CPU 进程切换率、CPU 利用率、空闲主存大小和 1 分钟负载平均值 6 个单项负载指标以及它们的“或”及“与”组合发现, 效果最好的是单项指标中的运行队列任务数, 任何其它单项指标或其组合都不比它好, 但实验条件有很大局限性. Banawan 和 Zahorjan^[6]使用预言(Oracle)模拟方法比较了作业到达数、作业离开数、瞬时队列长度、平均队列长度、平均响应时间、平均伸展系数(作业从到达离开的时间与在此阶段收到的服务时间之比)、利用率、未完成的工作量等负载指标, 得到的结论是使用瞬时队列长度、利用率、平均队列长度、平均响应时间和平均伸展系数中任何一个都可以极大地改进系统性能. Stumm^[7]指出 CPU 队列很少与 CPU 利用率有关系, 特别是在交互式环境中. Mehra 和 Wah^[8]使用了比较元(Comparator)神经网络, 该网络能学习预测一个作业的相对执行时间, 但该作业必须只使用在其到达之前便观察到的资源利用率模式. 他们的负载指标包含着如 CPU、磁盘、网络及主存等主要被争用资源的信息. 但此方法只适用于重复执行的小作业, 随着执行时间的增多, 其准确性便下降了.

大多数分布式系统使用进程数作为负载指标^[9,10], 但该指标没有区分进程的性质和大小. 一个大进程对资源的占用量及其运行时间可能比几个小进程对资源的占用量总和还要多. I/O 类作业和 CPU 类作业对资源的使用情况也是大不相同的: 一个长时间运行的 I/O 类作业占用很少 CPU 周期, 而一个短的 CPU 类作业在运行时可以占用全部的 CPU. 若表示系统负载情况的 CPU 队列长度将系统进程也考虑进去, 则更不准确, 因为系统进程很少占用资源. 例如, 一个工作繁忙的 SUN4/65 工作站开机 4 个小时, 其 swap 进程和 inetd 进程才分别使用了 7s 和 2s 的时间. 因而只使用进程队列长度作为负载指标是不合适的. 但直接使用资源利用率, 即 CPU 利用率和 I/O 利用率却很直观、准确. 只有当资源利用率达到 100% 时, CPU 队列长度对作业响应时间的影响才是主要的. 例如, 同一个计算量较大的作业在 CPU 利用率均为 100%, 但 CPU 队列长度分别为 2.4、3.4 的 2 个 SUN4/65 工作站上的执行时间分别为 68 秒和 99 秒.

本文通过实验给出了系统资源利用率与作业响应时间之间的关系. 实验系统为 SUN4/65 工作站.

1 资源利用率与作业响应时间

在作业执行所需的资源中, 通信带宽主要在诸如发送信号、等待信息到达等同步及通信延迟方面影响作业的响应时间, 它的变化性较大, 这里不予考虑. 另外, 我们假设网络利用率很低, 不会发生争用. 现代工作站的软、硬件技术对运行的作业提供足够内存, 使得内存很少争用. 这样, 我们主要讨论 CPU 周期和磁盘带宽的变化对作业响应时间的影响.

1.1 CPU 利用率

CPU 是系统的主要资源, 也是作业争用的主要对象. CPU 利用率(%CPU)对不同类型作业响应时间的影响是不同的.

首先是 I/O 类作业, 表 1 是对一个典型 I/O 类作业在不同 %CPU 的主机上的响应时间测定的结果. 被测作业在空载下运行时的 %CPU 为 7% 左右, I/O 利用率约为 86%. 表 1 中

所列数据均是 8 次测得的平均值. 结果表明,不同忙碌程度的 CPU 对 I/O 类作业的响应时间影响甚微.

表 1 I/O 类作业在不同 %CPU 下的响应时间

| 环境 %CPU | 空载 | 17% | 25% | 36% | 47% | 57% | 67% | 77% | 87% | 99% |
|---------|------|------|------|------|------|------|------|------|------|------|
| 响应时间(s) | 52.6 | 52.4 | 52.5 | 52.2 | 52.4 | 52.6 | 52.8 | 52.5 | 52.6 | 53.1 |

图 1 表示出在不同 %CPU 负载环境下分别运行不同 %CPU 作业的响应时间的变化情况. 从图 1 可以看出,同一被测作业在不同 CPU 利用率的负载环境下,其作业响应时间增加的百分比是不同的,环境负载越重则其响应时间越长,而且当该作业占用 CPU 越大时情况越明显. 从图 1 还可以看出,在同一负载环境下,占用 CPU 周期(%CPU)程度不同的作业,其响应时间增加的百分比也是不同的,占用 CPU 越多的作业,其作业响应时间增加的百分比也越大. 为研究在系统重载情况下是否值得进行作业迁移,我们从图 1 找出在系统环境负载为 95.4%的情况下,占用 CPU 程度不同的作业的响应时间增加的百分比,并由此绘制成图 2. 从图 2 可以看出,在系统重载情况下,轻作业(%CPU<40%)不值得迁移,因其迁移不会使响应时间提高很多(不到 10%);而运行在重载机上的、%CPU 大于 40%的作业则值得迁移.

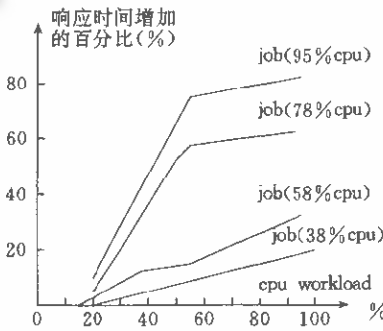


图1

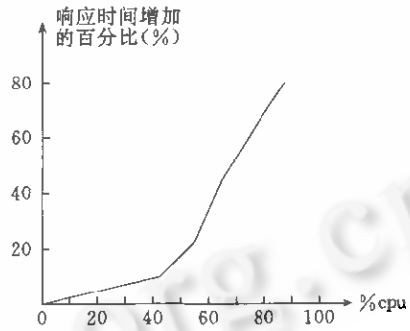


图2

1.2 I/O 利用率

磁盘(访问)带宽也是系统的主要资源之一,是 I/O 类作业争用的主要对象. I/O 利用率(%IO)表示一个工作站上磁盘带宽被使用的情况,该值和 CPU 利用率的值可以用 SunOS 现有的命令(Iostat)获得.

CPU 类作业在具有不同的 %IO 负载但 CPU 占用量很小的工作站(SUN4/65)上运行此类作业的响应时间见表 2. 被测程序是求小于 1000,000 的最大质数.

表 2 CPU 类作业在不同 %IO 下的响应时间

| 环境 %IO | 空载 | 12% | 27% | 36% | 42% | 55% | 66% | 78% | 86% | 96% |
|---------|------|------|------|------|------|------|------|------|------|------|
| 响应时间(s) | 64.0 | 64.0 | 64.1 | 64.8 | 64.6 | 65.1 | 65.1 | 66.2 | 66.5 | 66.9 |

可见,环境负载的 %IO 对 CPU 类作业的响应时间影响很小.

但是,不同 %IO 的负载环境对不同程度 I/O 类作业的影响规律却是很大的,见图 3. 从图 3 可以看出,环境 %IO 越大则被测作业的响应时间越长,而且当该作业 I/O 操作越多时

情况越明显. 这表明环境 %IO 对 I/O 类作业影响很大. 从图 3 还可以看出, 在同一负载环境下, 具有不同的 I/O 利用率的 I/O 作业, 其响应时间增加的百分比也是不同的. 大 I/O 类作业, 其百分比也大. 图 4 给出了在 workstation I/O 满载 (%IO 为 96%) 的情况下, 具有不同 I/O 利用率的作业的响应时间增加的百分比. 从图 4 可以看出, 当作业的 %IO 大于 10% 时, 该作业值得迁移.

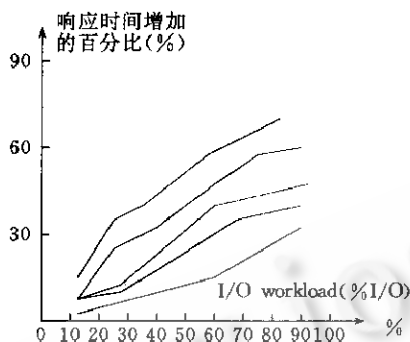


图3

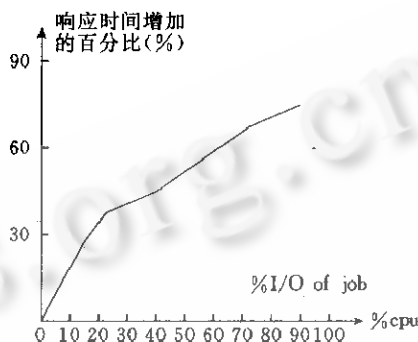


图4

1.3 CPU 队列长度

正如前边所提到的, 当 CPU 利用率达 100% 时, 必须以 CPU 队列长度作为负载指标对 CPU 类作业进行迁移位置的决策. 表 3 是在主机 %CPU 达 100% 并具有不同 CPU 队列长度的情况下运行 CPU 类作业的响应时间. 可以看出, CPU 队列长度对 CPU 类作业的响应时间影响很大, 两者之间关系的具体细节见文献[2].

表 3 CPU 类作业在不同 CPU 队列长度下的响应时间 (%CPU=100%)

| 平均 CPU 进程数 | 1.1 | 2.1 | 2.9 | 3.8 |
|----------------|------|------|-------|-------|
| CPU 类作业响应时间(s) | 62.8 | 92.1 | 123.8 | 235.0 |

从前边的实验结果可得出如下结论:

- (1) CPU 类作业和 I/O 类作业分别使用 %CPU 和 %IO 为负载指标;
 - (2) 当被提交作业的类型为混合类型时, 同时使用 %CPU 和 %IO 作为负载指标;
 - (3) 若各主机 %CPU 均达到满载 (100%), 则选用 CPU 队列长度作为负载指标.
- 作业类型的测定因篇幅所限不再说明, 详见文献[11].

2 资源利用率与负载平衡

为了研究使用不同负载指标的负载平衡算法的性能, 实现 3 个 LB 系统: QLB, cpuLB 和 JLB, 它们在负载指标上分别使用 CPU 队列长度、CPU 利用率、CPU 利用率+I/O 利用率+CPU 队列长度. 环境是由 Ethernet 连接起来的 6 台无盘工作站 (Sun4/20) 和一个服务器 (Sun4/65).

我们分别在无平衡系统 (NoLB) 和这 3 种 LB 系统上运行同一组命令流 (Script), 测量出每种情形下的平均作业响应时间、平均 CPU 利用率及平均 CPU 队列长度, 而且计算出每组值的均方差及相对于 noLB 情形的改进百分比. 这一组 script 在 6 台机器上的负载分

布属于典型情况,即轻载($\%CPU < 30\%$)、中等($40 \sim 60\%CPU$)、重载($\%CPU > 70\%$)各 2 台. 均方差是表示一组数据离散程度的,这是衡量一个 LB 系统性能的一个重要指标. 一个好的 LB 系统应使各主机的负载均衡,这就表现为各主机负载的均方差较小.

组成 script 的命令中包括多种多样的 unix 系统命令及典型的用户程序,例如包含 ls, ps, df, cp, rm, cc, split, compress 等,还包含求大质数、求菲波那契数程序等. 测试的结果如表 4.

表 4 各种 LB 算法比较

| 算法 | 平均作业响应时间(s) | | | | 平均 CPU 利用率(%) | | | 平均队列长度 | | |
|-------|-------------|-------|--------|-------|---------------|-------|-------|--------|------|-------|
| | 测量值(s) | 改进(%) | 均方差(s) | 改进(s) | 测量值 | 均方差 | 改进(s) | 测量值 | 均方差 | 改进(s) |
| noLB | 24.98 | 0 | 69.54 | 0 | 59.20 | 20.87 | 0 | 1.92 | 1.35 | 0 |
| QLB | 16.72 | 33.1 | 42.66 | 38.6 | 67.27 | 7.86 | 62.3 | 1.71 | 0.28 | 79.2 |
| cpuLB | 14.32 | 42.7 | 41.57 | 40.2 | 69.87 | 5.97 | 71.4 | 1.31 | 0.46 | 65.9 |
| JLB | 12.65 | 49.4 | 28.65 | 58.8 | 67.96 | 5.65 | 72.9 | 1.54 | 0.31 | 77.1 |

由上表可以看出使用 CPU 利用率作为负载指标的 LB 系统在性能上优于使用队列长度作为负载指标的 LB 系统,而同时考虑了资源利用率和 CPU 队列长度时则具有最好的性能. 为了比较,表 4 也列出了平均队列长度. 在平均队列长度的均方差方面,使用资源利用率不如 CPU 队列长度是意料中的事.

3 结 论

本文讨论了动态负载平衡系统中负载指标的选定. 负载指标对动态负载平衡系统的性能影响很大. 我们建议使用资源利用率作为负载指标来代替目前普遍使用的资源队列长度这种负载指标.

作业进入系统后只有获得足够的资源才可运行. 不同性质的作业对各类资源的需求是不一样的. 以计算为主的作业主要占用 CPU 资源,能否最大限度地缩短此类作业的响应时间关键在于能否尽量多地获得 CPU 资源. 以 I/O 为主的作业主要占用 I/O 通道,能否最大限度地缩短此类作业的响应时间关键在于它能否尽量多地获得 I/O 带宽. 资源队列长度与资源利用率并没有直接的关系,资源的队列长度小并不能说明资源利用率就低,资源利用率高也不意味着资源的队列长度就大. 国际上现有的负载平衡设施多使用 CPU 队列长度作为负载指标,因为它比较容易获得,而且它与多数作业的作业响应时间有密切的关系. 若使用资源利用率作为负载指标则必须知道作业需要哪些资源,在作业执行之前获得这些资源是很困难的. 我们使用了在作业执行开始时用跟踪的方法预测作业行为进而得出作业类型的方法,从而解决了这个问题.

我们的实验结果表明在动态负载平衡系统中使用资源利用率作为负载指标比使用资源队列长度作为负载指标更好.

参考文献

- 1 Ferrari D, Zhou S. A load index for dynamic load balancing. 1986 Fall Joint Computer Conference, Nov. 1986.
- 2 Ferrari D. A study of load indices for load balancing schemes. Report No. UCB/CSD85/262, Computer Science

Division, UCB, Oct. 1985.

- 3 Zhou S. An experimental assessment of resource queue lengths as load indices. Winter USENIX Conference, Jan. 1987.
- 4 Bonomi F, Fleming P J, Steinberg P. An adaptive join—the—biased—queue rule for load sharing on distributed computer system. 28th Conference On Decision and Control, Dec. 1989.
- 5 Kunz T. The influence of different workload description on a heuristic load balancing scheme. IEEE Transaction On Software Engineering, July 1991, 17(7).
- 6 Banawan S A, Zahorjan J. On comparing load indices using oracle simulation. 1990 Winter Simulation Conference, 1990.
- 7 Stumm M. The design and implementation of a decentralized scheduling facility for a workstation cluster. 2nd Conference on Computer Workstations, March 1988.
- 8 Mehra P, Wah B W. Automatic learning of workload measures for load balancing on a distributed system. Int'l Conf. on Parallel Processing, 1993.
- 9 Zhou S, Ferrari D. A measurement study of load balancing performance. 7th ICDCS, 1987.
- 10 Ezzat A K. Load balancing in NEST; a network of workstations. Fall Joint Computer Conference, Nov. 1986.

USING RESOURCE UTILIZATION AS LOAD INDEX IN DYNAMIC LOAD BALANCING

Ju Jiubin Yang Kun Xu Gaochao

(Department of Computer Science Jilin University Changchun 130023)

Abstract Load indices which accurately reflect the current load status at a computer system are crucial to a good dynamic load balancing system. However, few load indices have been experimentally validated that they are suitable for load balancing. Resource queue length (CPU queue length only exactly) are currently used in implemented dynamic load balancing systems extensively. In this paper, it's drawback is indicated and a new load index, resource utilization, is suggested. Either resource utilizations or response times of jobs in a dynamic load balancing system using resource utilizations as load indices are better than the ones in a system using CPU queue length as load index.

Key words Dynamic load balancing, load index, resource utilization, queue length, response time.