国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY
第19卷第1期 1997年2月 Vol. 19 No. 1

工作站网络并行计算的性能分析*

孙安香 宋君强 李晓梅
(国防科技大学计算机系 长沙 410073)

摘 要 本文介绍了异构工作站网络环境下一种新的并行计算性能评价标准。应用该标准分析了在四台异构的SGI工作站上浅水波方程并行计算的性能。

关键词 工作站网络, 消息传递, 加速比, 并行效率, 可扩放性
分类号 TP301.6

Performance Analysing of Parallel Computing on Networks of Workstations

Sun Anxiang Song Junqiang Li Xiaomei
(Department of Computer Science, NUDT, Changsha, 410073)

Abstract In this paper, we describe a new criterion of the parallel computing performence of heterogenous networks of workstations. With the criterion we analyse the parallel computing performance of the shallow-water equations on four workstations.

Key words networks of workstations, passing message, speedup, efficiency, scalability

由于工作站网络的高性能及低消耗, 其应用越来越广泛。实际应用中, 工作站网络一般都是异构的, 特殊情形是同构的。工作站网络的并行计算基于网络计算, 通常通过高速互连网络相连结, 用可移植的消息传递机制MPI、PVM 或EXPRESS 等来分配任务, 进行并行计算。工作站网络环境下, 应用消息传递进行并行计算需解决的主要问题: 什么样的计算模型是最合理的, 传递的性能评价标准应作何修改, 新的通讯机制如何减小延迟, 在特定的消息传递平台上如何设计并行算法。

针对90年代和未来计算机系统结构的特点, 1993年美国的David 提出了Log$P$ 并行计算模型[1]。该模型基于消息传递原理, 准确地描述了分布式互连网络的性能特征, 为分布存储计算机的算法设计提出了一般原则。Log$P$ 模型也适合工作站网络的并行计算。

* 国家863项目资助
1996年6月13日收稿

1 工作站网络并行计算的性能评价标准

并行计算的性能评价应考虑并行算法和并行系统的结构。基于工作站网络结构的新特点，评价工作站网络的并行计算性能不能采用传统的标准，下面介绍美国Texas大学高性能计算与软件实验室的有关研究成果[2]。

一个异构的工作站网络定义为：$HN(M,C)$，这里：$M=\{M_1, M_2, \cdots, M_m\}$，$C$ 是工作站的互连网络，如：Ethernet 网、ATM 或 FDDI，每两台处理机间的网络具有相同的带宽。定义工作站 M_i 的计算能力权重

$$W_i = S_i(A) \setminus \max_{1\le j\le m}\{S_j(A)\}, \quad i = 1, 2, \cdots, m \tag{1}$$

其中 $S_i(A)$ 是工作站 M_i 独占解决问题 A 的速度。(1) 也可写成

$$W_i = \min_{1\le j\le m}\{T(A, M_j)\} \setminus T(A, M_i), \quad i = 1, 2, \cdots, m \tag{2}$$

这里 $T(A, M_i)$ 是独占 M_i 解决问题 A 的时间。下面我们介绍性能评价标准模型。

1.1 加速比和并行效率

异构型工作站网络的各 CPU 速度、I/O 和内存带宽不同，并行计算一个问题 A 时，各工作站的速度不一样。定义加速比

$$SP(A) = \min_{1\le j\le m}\{T(A, M_j)\} \setminus T(A, HN) \tag{3}$$

式中 $T(A, HN)$ 是并行计算 A 所需的时间，$T(A, M_i)$ 是 M_i 独立完成 A 所需的时间。

考虑100台工作站的网络系统，设最快的工作站上计算权重为1，其余99台的计算权重为0.5. 若最快的工作站速度再增加一倍，而其它99台的速度不变，则这99台机器的计算权重降为0.25，那么加速比将比前一种情况减少将近50%. 为了描述这种情况的变化，定义工作站网络的异构性

$$H = \sum_{j=1}^{m}(1 - W_j(A)) \setminus m \tag{4}$$

为了进一步描述异构网络的加载作用，定义并行度

$$P_{\deg}(A) = \sum_{j=1}^{m} T_j^{act} \setminus T(A, HN) \tag{5}$$

这里 T_j^{act} 为并行计算 A 时各工作站的计算时间($j = 1, 2, \cdots, m$)。

这样，加速比 SP、并行度 $P_{\deg}$ 和异构性 H 的关系为：

$$SP = P_{\deg}^{*}(1 - H) \tag{6}$$

由此可见，加速比随着并行速度的增加而增加。当异构性增加时，加速比减少。前面提到的100台机器的异构网络，第一种情况其异构性接近50%，第二种情况异构性为75%，第二种情况的加速比较第一种情形低。

设 $A = (A_1, A_2, \cdots, A_m)$，独占系统 $HN(A, M)$ 的并行计算效率定义为

$$E = \sum_{j=1}^{m}(W_j * A_j / S_j) \setminus \sum_{j=1}^{m} T(A, HN) W_j \tag{7}$$

由加速比的定义可以得到并行效率 E 与加速比 SP 的关系

$$E = SP \setminus \sum_{j=1}^{m} W_j \tag{8}$$

对于同构的工作站网络系统 $W_j = 1 (j = 1, 2, \cdots, m)$，则

$$E = SP \setminus m \tag{9}$$

即为传统的并行效率与加速比的关系，也进一步验证了上述异构工作站网络并行计算效率和加速比的合理性。

1.2 可扩放性

可扩放性用于评价问题规模扩大和系统处理机数目增加时并行处理系统的性能。问题 A 在异构网络 HN 的运行时间可分为两部分: CPU 时间 $Teff(A_i, M_i)$ 和系统开销延迟 $overhead(i)$，每一个计算权重单元其平均系统开销延迟为:

$$Le(A(I), HN) = \sum_{j=1}^{m} overhead(j) W_j \Big\backslash \sum_{j=1}^{m} W_j \qquad (10)$$

在异构网络环境 HN_1 上，问题 A 的计算规模为 $A(I_1)$，平均系统开销延迟为 $Le(A(I_1), HN)$。在 HN_2 上计算规模为 $A(I_2)$，延迟为 $Le(A(I_2), HN_2)$，系统由 HN_1 扩大到 HN_2，可扩放性可定义为

$$scale(A, (HN_1, HN_2)) = Le(A(I_1), HN_1) \backslash Le(A(I_2), HN_2) \qquad (11)$$

显然上式的值是小于或等于1的。

2 实验环境和结果

实验环境为异构网络工作站环境，由1台 Indigo-2 SGI 工作站和3台配置相同的 Indy-SGI 工作站组成，用10Mb/s 带宽的 Ethernet 网络联结起来。Indy-SGI 工作站的主频为100MHz，内存为32MB。Indigo-2 SGI 工作站的主频为150MHz，内存为64MB. 消息传递平台为EXPRESS，实验程序为用差分方法解浅水波方程[3]，有三种程序模式: 串行程序 test.f、采用一维自动域分解法(ODI: One Domain Image)的并行程序 tesf1.f 和采用二维自动域分解法(TDI: Two Domain Image)的并行程序 test2.f. 为了分析该实验环境的并行计算性能，我们测试了不同问题规模下三种程序模式的墙上时间和系统开销。

2.1 加速比和并行效率分析

按照第1节给出的评价标准，我们计算各种规模下 test1.f 和 test2.f 的加速比和并行效率。各计算规模大小为

问题1: 网格点为64×64，时间积分为120步;

问题2: 网格点为128×128，时间积分为240步;

问题3: 网格点为256×256，时间积分为480步;

问题4: 网格点为512×512，时间积分为960步。

由表1可以看出，对问题1，得到的加速比和并行效率很低。主要原因是问题规模小，通讯开销所占比例大。问题2计算规模是问题1的8倍，加速比和并行效率均有所提高，这是因为问题规模扩大，通讯所占比例减小而提高了并行效率，但还是不理想。网格在纵向和横向再扩大2倍，即问题3在问题2的基础上计算规模再扩大8倍，其并行性能较问题2好，这是因为规模扩大后，系统开销与计算开销分配的进一步合理。对问题4，其计算规模是问题3的8倍，是问题2的64倍，是问题1的512倍，由表1可以看出，加速比和并行效率的值与问题3较接近。这说明问题3的系统开销与计算开销的分配比例已较为合理，其并行计算性能指数接近稳定。

表1　浅水波方程并行计算测试结果

网　格	串行墙上时间	并行计算权重	程序模式	墙上时间	加速比	并行效率
64×64	2.7	2/3 2/3 2/3 1	test1.f	2.66	1.014	0.338
			test2.f	2.61	1.027	0.342
128×128	24.0	2/3 2/3 2/3 1	test1.f	13.79	1.751	0.584
			test2.f	12.63	1.907	0.636
256×256	226.6	2/3 2/3 2/3 1	test1.f	84.39	2.686	0.895
			test2.f	81.19	2.792	0.931
512×512	1842.1	2/3 2/3 2/3 1	test1.f	677.33	2.719	0.906
			test2.f	657.65	2.801	0.934

2.2　可扩放性分析

第1节已经提到，可扩放性主要考虑异构系统的系统开销，即问题规模变化对系统开销延迟产生的影响。受环境和解浅水波方程的特点的限制，这里只讨论问题规模扩大的可扩放性。根据第1节平均延迟的定义，我们测试了各问题的系统开销，从而得到程序 test1.f 关于问题1、2、3、4的平均延迟分别为 $L_1=2.66_s$，$L_2=8.16_s$，$L_3=14.00_s$，$L_4=116.18_s$. 程序 test2.f 关于问题1、2、3、4的平均延迟分别为 $L_1=2.37_s$，$L_2=7.98_s$，$L_3=13.95_s$，$L_4=137.18_s$. 进一步根据可扩放性的定义得到 test1.f 的可扩放性

$$scale_{1,2}=\frac{L_1}{L_2}=0.3265\quad scale_{2,3}=\frac{L_2}{L_3}=0.5828\quad scale_{3,4}=\frac{L_3}{L_4}=0.1205$$

从以上 scale 的值可知，计算规模由问题1扩大到问题2，可扩放性好，由问题2扩大到问题3，可扩放性值也好，但由问题3扩大到问题4，可扩放性值不大。这里的分析结论与2.1的结论一致。test2.f 的平均延迟与 test1.f 相差不大，可扩放性也一样。上述数值实验表明：针对工作站网络这种新的并行机系统，我们介绍的并行计算性能评价标准更具有合理性。

参 考 文 献

1　李晓梅. 并行计算模型及其算法设计. 数值计算与计算机应用，1995，16 (3)

2　Zhang X，Yan Y. Modeling and characterizing parallel computing performance on heterogenous networks of workstations. IEEE Symposium on Parallel and Distributed Processing，October，1995

3　Sadourney R. The dynamics of difference models of shallow-water equations. J Atmos. Science，1975，32：680～689

(责任编辑　张　静)