基于 Beowulf 集群的可扩展性模型的研究

孔令鑫,祝永志,侯秀杰
(曲阜师范大学 计算机科学学院,山东 日照 276826)

摘 要:可扩展性是衡量并行算法与并行系统匹配程度的一项重要指标。分析了传统的等并行开销计算比评价准则,指出其优缺点。为了适用于基于 Beowulf 集群的分布式并行计算环境,对传统的等并行开销计算比评价准则进行改进,得到 Beowulf 环境下的新的可扩展函数。该可扩展函数能够直观地反映基于 Beowulf 集群的分布式并行系统在机器规模和问题规模扩展时,其性能的扩展特性。用该评价准则分析并论证了编制的并行算法与 Beowulf 集群相结合的可扩展性。

关键词:可扩展性;等并行开销计算比;Beowulf 集群;分布式计算
中图分类号:TP311 **文献标识码**:A **文章编号**:1673-629X(2009)07-0127-03

Research of Scalability Model Based on Beowulf Cluster

KONG Ling-xin, ZHU Yong-zhi, HOU Xiu-jie
(College of Computer Science, Qufu Normal University, Rizhao 276826, China)

Abstract: Scalability has become an important indicator to measure the match degree of parallel algorithms and parallel system. The iso ratio of parallel overhead to computation is reviewed in this paper, the merit and deficiencies of this metric are pointed out. Then in order to apply the distributed parallel computation environment based on Beowulf cluster it is improved, obtain the new extensible function which reflects the scalability of distributed parallel systems more directly and precisely when the size of machines and the scale of problems are extending in the environment of Beowulf cluster. Finally, the new metric is used to analyze and prove the scalability of parallel algorithms and Beowulf cluster.

Key words: scalability; iso ratio of parallel overhead to computation; Beowulf cluster; distributed computation

0 引 言

在网络速度和微处理器性能迅速提高的情况下,集群系统的研究和应用得到迅速发展,文中,分布式并行系统是基于 Beowulf 集群的环境[1]。

可扩展性是衡量算法与并行系统匹配程度的一项重要指标,也一直是分布式并行计算所追求的一个重要目标。人们非常注重随着处理机台数的增加,其计算性能如何,即并行计算的可扩展性如何。目前可扩展性研究主要集中在并行算法与并行机相结合的可扩展性上,已有很多的研究成果[2~4]。但是并行机由于系统结构的限制或成本较高或研究经费等诸多的因素,其市场受到一定的限制。集群由于其投资风险小、编程方便、性能/价格比高等因素,已成为实现并行计算的一种主流技术。但目前针对集群系统的可扩展性研究还很少。在一般的集群系统中,系统规模扩展[5]可以从两个方面进行:一是增加物理节点个数,即物理扩展;二是通过升级来提高节点的处理能力,称为能力扩展。在文中,采用的是物理扩展。可扩展性是与加速比和效率的概念紧密相关的,增加处理器数和求解问题的规模都可能提高加速比。所以,对于可扩展性的评测可以通过加速比和系统效率来描述。

尽管可扩展性如此重要,但目前仍无一个公认的、标准的和被普遍接受的严格定义和评判它的标准。在保持某种性能指标的前提下,去预测一个扩展的集群系统能够求解多大的问题规模,这是个值得研究的问题[6]。笔者分析了传统的等并行开销计算比可扩展评价准则[7],然后将传统的等并行开销计算比模型推广到集群系统中,提出了一个面向 Beowulf 集群系统的可扩展模型,获得问题规模与系统规模的扩展关系式。

收稿日期:2008-10-20;**修回日期**:2009-01-04
基金项目:山东省高等学校实验研究项目基金(2005-400);曲阜师范大学科研资助项目(XJ0734)
作者简介:孔令鑫(1982-),女,硕士研究生,研究方向为分布式计算;祝永志,教授,硕士生导师,研究方向为网络与分布式系统。

1 Beowulf 集群简介

集群(Cluster)可如下定义:一组相互独立的计算机系统构建的一个松耦合多处理机系统,系统中各进

程借助网络实现通信、共享内存传递信息,从而实现分布式并行计算。在网络和微处理器性能迅速提高的情况下,集群系统的研究和应用得到迅速发展,对重新定义超级计算概念也产生了重要作用。

在 1994 年夏季,Thomas Sterling 和 Don Becker 在 CESDIS 用 16 个节点和以太网组成了一个计算机集群系统,并将这个系统命名为 Beowulf 集群。Beowulf 是一种集群系统,它可以使用很多普通的 PC 做成一个集群来解决人们所面临的问题,并且这种集群所具有的价格优势是传统的并行计算机所无法比拟的。目前,Beowulf 集群是科学计算中流行的一类高性能并行计算机集群结构[8]。

Beowulf 集群把具有相同配置的 PC 作为组装单元,运行 Linux 等自由软件,各个单元之间通过 TCP/IP 协议的局域网以及有关的程序库分配计算任务和通信。一般由服务节点来控制整个集群,服务节点是集群的控制台和对外的网关。在规模较大的集群中是可以有多个服务节点的。通常,除了服务节点以外,集群中的其他节点都是哑成员,这些成员由服务节点来管理,执行服务节点分配的任务。Beowulf 的性能一般取决于以下几个因素:节点本身、节点之间互联设备、底层通信软件、全局资源管理系统以及计算环境(PVM 或 MPI)等。现在 Beowulf 集群使用的操作系统扩展到 Microsoft Windows 在内的很多操作系统[9]。

2 等并行开销计算比评价准则

2.1 传统的等并行开销计算比可扩展评价准则

在等并行开销计算比可扩展模型中,采用扩展后和扩展前的平均速度之比来衡量实际系统的可扩展性。

设 T_{comp} 表示最后结束的处理机的纯计算时间,T_0 表示最后结束的处理机所有的开销时间,包括等待、同步和通信时间,整个并行算法的并行执行时间:

$$T_p = T_0 + T_{comp}$$

并行开销计算比即 $c = \frac{T_0}{T_{comp}}$

设 P 台处理机的串行算法计算量为 W,设 W' 为 $P'(P' > P)$ 台处理机时维持并行开销计算比 c 的串行算法计算量,则维持并行开销计算比不变时从处理机台数为 P 到 P' 的并行系统的可扩展性为

$$\mathrm{Scal}(P,P') = \frac{V_p'}{V_p} \tag{1}$$

即

$$\mathrm{Scal}(P,P') = \frac{W'}{W} \cdot \frac{T_p}{T_p'} \cdot \frac{P}{P'} \tag{2}$$

$$\mathrm{Scal}(P,P') = \frac{W'}{W} \cdot \frac{T_{comp}}{T_{comp}'} \cdot \frac{P}{P'} \tag{3}$$

$$\mathrm{Scal}(P,P') = \frac{W'}{W} \cdot \frac{T_0}{T_0'} \cdot \frac{P}{P'} \tag{4}$$

设 α 为通信建立的延迟时间,β 为传递每个字节所需时间,不考虑网络竞争,处理机间发送或接受由 s 个字节组成的消息所需时间为 $\alpha + s\beta$,设 T_i 为每台处理机的总开销,则

$c = \frac{T_m}{T_{comp}}$,其中 $T_m = \max_{1 \leqslant i \leqslant P}(T_i)$,是并行算法的通信复杂性。

人们在设计一个算法时,往往要给出算法的通信复杂性来说明算法的通信性能,使用此式就可以推理出问题规模 W 怎样随处理机台数变化,系统才是可扩展的。

可以得出:该评价准则通过扩展前后的平均速度之比来度量实际系统的可扩展性的方法可以用来评估并行算法的可扩展性;但是,该评价准则可能使得执行时间过长。

2.2 改进的可扩展性评价准则

在 Beowulf 集群系统中,假设执行 n 种不同类型的 k 个计算任务所需的时间为 T_k,其中第 i 种类型由 $k_i(1 \leqslant i \leqslant n)$ 个计算任务构成,$t_i(1 \leqslant i \leqslant n)$ 为第 i 种类型的执行时间,则 k 个计算任务的总的执行时间 $T_k = \sum_1^n k_i t_i$,其中,$k = \sum_1^n k_i$

那么,系统的执行速度 $V_k = \frac{k}{T} = \frac{k}{\sum_1^n k_i t_i}$

定义 1. 系统单个节点上的开销计算比:

$$c_i = \frac{T_i^0}{T_i^{comp}}(i = 1,\cdots,p)$$

定义 2. 若系统的问题规模为 W、执行速度 V 增长使得开销计算比为 $c_i = c(i = 1,\cdots,p)$,此时 W、V 和 P 满足关系式 $W = F(P,V)$,称 F 为此系统的可扩展函数。

根据 $c_i = c(i = 1,\cdots,P)$ 这个条件,可以获得每个节点上 W_i、V_i 和 P 的关系式:

$$W_i = f_i(i = 1,\cdots,P)$$

加权后可推出关系式

$$W = F(P,V) \tag{5}$$

3 可扩展性评测

以下试验是在 Beowulf 集群上实现的。每台微机上运行 Redhat Linux 8.0 操作系统,并采用 MPICH.NT 1.2.5.2 作为分布式并行计算的支撑环境。

下面来看以高斯列主元素消去法求解线性方程组。算法的思路是,首先寻找系数矩阵第一列中绝对值最大的数,称之为主元。接着,选取该行作为参照行,把它运行的第一个元素消为零。然后,对剩余的行依次重复该操作,直到系数矩阵通过换行变成一个上三角矩阵。这个过程称为消元,消元后就可以快速地解出未知数的值。

并行程序设计方法:将计算按照节点个数进行划分任务,每一个任务负责消去矩阵的一块,并不仅仅是一行。对于算法运行的环境,一方面,当问题规模较小时,程序的通信量并不大。当问题规模增大后,用户需要的计算量不断增大,用户时间上升的比较快。另一方面,当计算量超出通信量时,并行程序才能发挥作用,计算通信量越大,并行程序的性能越高,并行效果越好。所以问题规模较小时采用并行方法求解有时候会适得其反,尤其是在分布式环境下。

在 Beowulf 集群的环境下,对 1000 阶的线性方程组求解问题进行实验测试,所得结果见图 1。

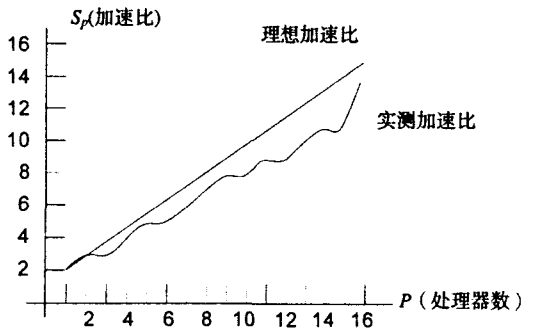

图 1　基于 Beowulf 集群系统的加速比测试结果

表 1　高斯列主元消去法的可扩展性

P	W	S_p	CPU	Scal(16,P)
8	256 * 256 * 256	4.13	42%	
16	512 * 512 * 512	7.08	44%	1.000
32	1024 * 1024 * 1024	14.46	50%	0.901
64	1380 * 1380 * 1380	28.88	52%	0.868

表 1 是高斯列主元素消去法在异构 Beowulf 集群的环境下的测试结果,其中 P 为处理机数、W 为问题规模,表格包括对加速比 S_p、CPU 利用率、可扩展性的分析。通过测得的数据可以看出 CPU 的利用率符合预测的先递减后递增的规律,经过改进的传统等并行开销计算比可扩展模型是适用于 Beowulf 集群环境的。

4　结束语

近年来,随着微处理器性能迅速地提高和以太网等局域网技术的日益成熟,用普通 PC 机建立的 Beowulf 集群得到了广泛的应用,基于此环境下的并行算法与 Beowulf 集群的可扩展性研究成为现今研究的重要课题之一。如何在处理器增多的情况下拓展问题规模,并使得问题的执行时间合理且 CPU 利用率合理,这值得学者们去研究。文中通过对传统的等并行开销计算比可扩展模型进行了分析,并对其改进,使得该准则在 Beowulf 集群环境下,能够评价 Beowulf 集群系统的可扩展性。

参考文献:

[1] 祝永志,王国仁. Beowulf 并行计算系统的研究与实现[J]. 计算机工程,2006,32(11):242 - 244.

[2] Grama A, Gupta A, Kumar U. Isoefficiency function: A scalability metric for parallel algorithms and architectures[J]. IEEE parallel & Distributed Technology, 1993, 1(3): 12 - 21.

[3] Sun X, Rover D. Scalability of parallel algorithm - machine combinations[J]. IEEE Trans . Parallel and Distributed System, 1994, 5(6): 599 - 613.

[4] Hwang K, Xu Z. Scalable parallel computing: Technology , architecture, programming[M]. Boston: Mc Graw - Hill Companies, 1998.

[5] 陈　军,李晓梅. 近优可扩展性:一种实用的可扩展性度量[J]. 计算机学报,2001,24(2):179 - 182.

[6] 王与力,杨晓东. 一种更有效的并行系统可扩展性模型[J]. 计算机学报,2001,24(1):86 - 90.

[7] 迟利华,刘　杰,李晓梅. 并行算法与并行机相结合的可扩展性[J]. 计算机研究与发展,1999,36(1):47 - 51.

[8] 祝永志,李丙峰,魏榕晖. Beowulf - T 机群系统性展性的研究[J]. 计算机科学,2008,35(2):298 - 302.

[9] 朱亚超,李胜利. Beowulf 集群系统性能评测技术研究[D]. 武汉:华中科技大学,2006.

(上接第 126 页)

[4] An Lan, Zhang Liansheng, Chen Meilin. A Parameter - Free Filled Function for Unconstrained Global Optimization[J]. Journal of Shanghai University, 2004, 8(2): 117 - 123.

[5] Yang Y J, Shang Y L. A New Filled Function Method for Unconstrained Global Optimization[J]. Applied Mathematics and Computation, 2006, 173: 501 - 512.

[6] 骆世云,叶仲泉. 求解无约束全局优化的改进的单填充函数法[J]. 计算机技术与发展,2008,18(8):108 - 111.

[7] Liang Y M, Zhang L S, Li M M, et al. A Filled Function Method for Global Optimization[J]. Journal of Computational and Applied Mathematics, 2007, 205: 16 - 31.

[8] 王晓丽,周国标. 实现快速全局优化的跨越函数方法[J]. 应用数学,2006,19(1):56 - 60.