

并行编程模式及分析*

何元清** 孙世新 傅彦

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】讨论了主从模型、单控制流多数据流、分治策略、数据流水线、投机策略、混合模型六种并行编程模型的结构和工作方式、性能及特点,分析统计了大量应用程序。指出了并行程序设计中,并行编程模式的选择直接影响了并行程序的正确性和效率,而主从模型和单控制流多数据流在整体性能上最优,程序执行效率也最好,是两种最常用的并行编程模型。

关键词 并行; 模型; 编程; 设计

中图分类号 TP388.6

Analysis of Parallel Programming Models

He Yuanqing Sun Shixin Fu Yan

(College of Computer Science and Engineering, UEST of China Chengdu 610054)

Abstract In this paper, the structure of several models is introduced, such as Master/Slave, Single program Multiple Data, Data Pipelining, Divide and Conquer, Speculative Parallelism, and the working is expatiated, both characteristic and performance analysis of parallel models is also given, A quantity of application programming is analyzes and stated. The result shows parallel models influence directly correctness and efficiency of parallel programming. Both master-slave and SPMD is best programming model because of that is best of all in common use models with the all and the one performance.

Key words parallel; model; programming; design

并行程序设计是在并行计算机上编写求解应用问题的并行程序技术,实现并行程序的四个要素是^[1]:并行体系结构、并行系统软件、并行程序设计语言和并行算法,并行语言是程序员进行并行程序设计的文本,也是编译系统对并行程序编译所依据的文本,它具有以下3个特性:1)并行模式;2)并行操作粒度;3)并行任务之间的通信模式。其中,并行模式的选择直接影响了并行程序的正确性和效率,从而影响整个系统的性能,因此选择一种有效的并行编程模式可以更好地提高系统的性能和效率。

1 并行编程模式

并行程序可以归纳为一些有明确定义的不同编程模式,每一种模式都具有相同控制结构的一类算法,大部分典型的分布式计算机应用都使用非常流行的客户/服务器模型,这种模式是通过远地过程调用(RPC)进行通信,并不具有内在并行性。并行计算领域的一些专家将模式从各种角度进行了分类^[2-4],目前,并行编程中用到的模式有以下6种:1)任务播种;2)单控制流多数据流;3)数据

2001年11月20日收稿

* 信息产业部预研基金资助项目

** 男 34岁 在职博士生 讲师

流水线；4) 分治策略；5) 投机并行；6) 混合模型。

1.1 任务播种

任务播种模型包括主任务和多个从任务两部分。主任务主要负责将问题分解为多个小任务，同时收集各个子问题的解，将其汇总得到最终结果。从进程执行如下一个很简单的循环：接受消息、进行计算、向主进程发送结果^[5]，其结构如图1所示。

1.2 单控制流多数据流

单控制流多数据流是一种普遍使用的模型，每个进程基本上执行同一段代码，但处理的数据不同。这种方法需要将应用程序分给不同的处理器，又称为几何并行、域分解或者数据并行，其结构如图2所示。

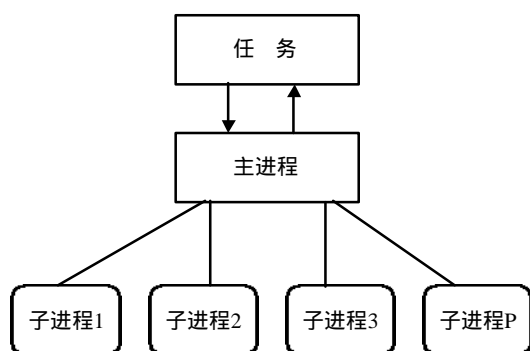


图1 主从模式基本结构

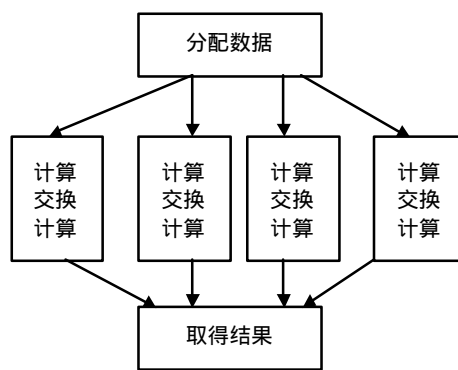


图2 单控制流多数据流结构

1.3 分治策略

分治策略是将一个问题分解成两个或两个以上的子问题，独立的解决每个子问题，然后将这些结果结合起来就得到最后的解，所以在分治策略中常见的三种操作是分解、计算、汇总。程序被组织成一种虚拟树的结构，一些进程派生出子任务，并且汇总这些任务的结果，计算出一个综合解。计算任务由虚拟树的叶子结点完成，执行过程如图3所示。

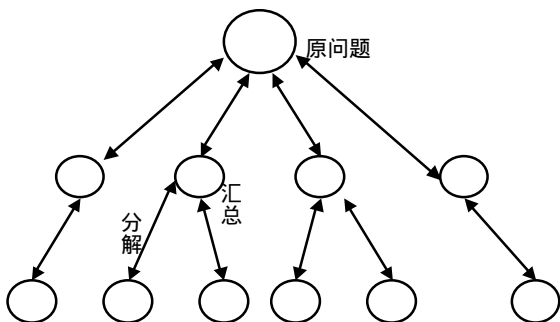


图3 分治策略的虚拟树

1.4 数据流水线

数据流水线是一种细粒度更高的并行机制，它建立在功能划分的基础上，其首先确定出算法的任务，然后每个处理器执行整个算法的一小部分。它又是一种最简单、常用的功能划分模型，结构如图4所示。

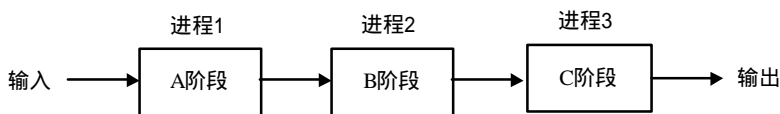


图4 数据流水线结构

1.5 投机策略

有些问题具有复杂的数据相关性，降低了开发并行性的可能性，在这种情况下，一种正确的解决办法是将问题划分为小块来解决，同时利用某些投机或者优化算法来提高并行执行效率。

1.6 混合模型

各种不同模型之间的界限有时候非常模糊，在某些应用中需要综合使用不同的模型，包括至少两种基本模型的混合方法，一般在大型应用程序中使用。将数据并行和任务并行混合起来同时使用，或者在同一程序的不同部分使用，是一种行之有效的方法。

2 性能比较

并行编程模式各有其特点, 本文对其
中几种常用的并行编程模式进行了性能分
析^[6], 比较结果如表1所示, 表中
表示级别的高低。

表1 编程模式性能比较

	效率	进程通信	可靠性	可扩展性	可重用
主从模式					
SPMD					
数据流水线					
分治策略					

3 数据分析

本文对现有的各种应用程序作了大量的分析统计^[7], 对其所使用的编程模式进行了归纳, 得出
如图5所示的分布图。图中, 主从模型和SPMD模型是一种普遍使用的模型, 在实际应用中使用最多,
其整体性能也最高。

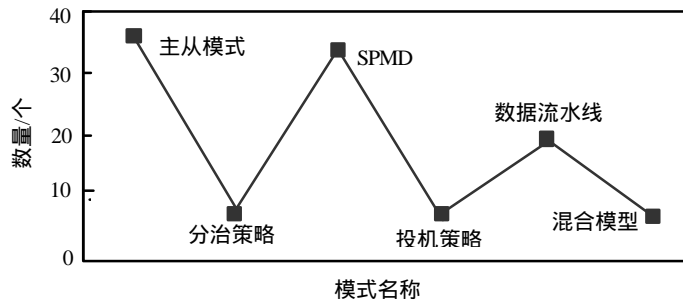


图5 应用模型分布图

4 结束语

在并行程序设计中, 并行模式是程序员必须考虑的重要因素之一。本文通过分析并行程序设计中常用到的六种典型模型, 详细介绍了这六种模型的结构, 比较了它们的特点及其性能。在并行编程模式中, 主从模型和SPMD在整体性能上最优, 是两种最常用的并行编程模型。

参 考 文 献

- 1 沈志宇, 胡子昂, 廖湘科, 等. 并行编译方法. 北京: 国防工业出版社, 2000
- 2 Pritchard D. Mathematical models of distributed computation. In Proceedings of OUG-7, parallel programming on transporter based machines. IOS Press, 1988. 25-36
- 3 Fox G. Parallel computing comes of Age: Supercomputer level parallel computations at Caltech. Concurrency: practice and experience, 1989, vol.1(1): 63-103
- 4 Wilson G. Parallel programming for scientists and engineers. MIT Press, Cambridge, MA. 1995
- 5 孙家昶, 张林波, 迟学斌, 等. 网络并行计算与分布式编程环境. 北京: 科学出版社, 1997
- 6 王剑波. 曙光2000培训教材. 成都: 西南交通大学, 2001
- 7 Fox G. What have we learn form using real parallel machines to solve real problems . In proceedings 3 rd conf. Hypercube concurrent computes and applications, 1988