基于MPI的网络并行计算环境及应用

罗省贤　何大可　编著

西南交通大学出版社

基于 MPI 的网络并行计算环境及应用

罗省贤　何大可　编著

西南交通大学出版社
·成 都·

内 容 简 介

分布式网络机群并行系统易于实现、可扩展性强、I/O 并行度高，并且在充分利用资源、建立异构并行计算环境方面具有无可比拟的灵活性。消息传递通信是分布式网络机群并行系统节点机间的通信方式，现在广泛使用的可移植消息传递界面 MPI 已经成为国际标准，并且使用方法比已经较成熟的 PVM 更简便。本书在介绍并行计算基础知识、概念以及分布式网络机群并行计算环境的基础上，重点介绍了 MPI 的通信模式、点到点通信、集群通信以及进程组虚拟拓扑结构等内容，并给出了一些可供读者实际应用的并行算法和 MPI 应用程序实例，使读者能够较全面地了解和掌握应用 MPI 编写消息传递机制并行程序的方法和技术。

本书可用作科研与工程技术人员的技术参考书以及高等院校教师、研究生的教学参考书，也可供学习并行技术的读者阅读和参考。

图书在版编目（CIP）数据

基于 MPI 的网络并行计算环境及应用 / 罗省贤、何大可编著. —成都：西南交通大学出版社，2001.12
ISBN 7-81057-632-1

Ⅰ. 基… Ⅱ. ①罗…②何… Ⅲ. 计算机通信－通信理论 Ⅳ. TN91

中国版本图书馆 CIP 数据核字（2001）第 092174 号

基于 MPI 的网络并行计算环境及应用

罗省贤 何大可 编著

*

出 版 人 宋绍南
责任编辑 张华敏
封面设计 肖 勤

西南交通大学出版社出版发行
（成都二环路北一段 111 号 邮政编码：610031 发行科电话：7600564）
http://press.swjtu.edu.cn
E-mail: cbs@center2.swjtu.edu.cn
四川森林印务有限责任公司印刷

*

开本：787mm × 1092mm 1/16 印张：13.125
字数：313 千字 印数：1—1000 册
2001 年 12 月第 1 版 2001 年 12 月第 1 次印刷
ISBN 7-81057-632-1/TN · 283
定价：29.50 元

目　录

试读结束，需要全本PDF请购买 www.ertongbook.com

第一章　并行计算机与并行计算概论

随着计算机的应用日趋复杂，从数据处理、信息处理、知识处理到智能处理，应用范围越来越广，处理问题的规模越来越大，因此对计算机的运算速度和处理能力要求也越来越高。尤其是一大批具有挑战意义的科学工程计算问题，如气象预报、流场计算、石油勘探及核物理等方面的计算问题需要计算速度达到百亿次每秒、甚至万亿次每秒浮点运算。但是无论如何发展技术，单台计算机的速度受到材料的物理限制，根据电子在硅材料中运动的极限速度估算，单机的极限速度是十亿次每秒。解决这一问题的办法是用成千上万个微处理器组成并行机，才能突破单机的极限速度。并行计算机是一种具有并行结构的高性能计算机，在计算机的发展历史中，并行技术一直是提高计算速度的一条重要途经。

1.1　并行计算机发展简史

具体地说，并行机是由两个以上的处理器连接起来并发操作的计算机。并行机最早的雏形诞生于 1963 年 2 月 18 日，美国西屋（Westing House）宇航实验室的工程师将 9 个 CPU 部件连接成一个 3×3 阵列，并用它进行计算求得了一个偏微分方程的解，这在当时是一个创举。不过世界上第一台真正能称为并行机的是美国 Illinois 大学研制的 ILLIAC-Ⅳ。1966 年美国 Illinois 大学开始研制 ILLIAC-Ⅳ，当时美国投资 3 000 万美元。ILLIAC-Ⅳ由一个控制部件和 64 个处理单元组成，64 个处理单元连接成 8×8 阵列，每个处理单元能做 64 位字长的浮点运算，局部存储器为 16 KB，设计速度为 400 万次每秒。ILLIAC-Ⅳ的理论峰值运算速度可达 2.5 亿次每秒。但是由于受当时元器件的限制，设计几经修改，历时 9 年，直到 1975 年才勉强能提供使用，实测的最大运算速度仅接近 5 000 万次每秒，而且此后每年用于维护的投资达 200 万美元。ILLIAC-Ⅳ作为并行计算机产品不是一个很成功的例子，它花费了四倍于合同规定的经费，其性能甚至未达到设计指标的十分之一，但是它对并行机发展的影响是深远的。ILLIAC-Ⅳ已作为计算机发展史上的划时代产品陈列在波士顿美国国家博物馆。

在此之后的十几年内，以向量化结构为主的计算机占了统治地位，比较成功的代表是美国 Cray 公司推出的向量—并行机 Cray-1。Cray-1 机的研制从 1972 年开始，1976 年第一台样机交付使用。Cray-1 机有 12 个流水功能部件，时钟周期为 12.5 ns，其主要创新点是设有 8 个向量寄存器组，每组能保存 64 个 64 位浮点数。向量计算机的特点是应用流水线的概念，将一个计算部件分解成可以并行操作的相对独立的几部分，形成流水线操作结构，流水线的每部分只完成一个子功能，通过时间重叠实现并行操作，达到运算加速的目的。这样，如果一次加法需要 3 个时钟周期，对于 3 级流水线上的向量加法，则几乎平均一个时钟周期完成

一次加法，效率是原来的 3 倍。由于 Cray-1 机大量采用了向量流水结构、实测运算速度达到 1.3 亿次每秒浮点运算、有很高的性能价格比，因此迅速而成功地实现了商品化生产。

尽管向量机很成功，但是单机的极限速度限制了计算机性能的不断提高，不能满足大型计算的需求。要突破单机的极限速度限制，最终要用大规模并行机（MPP）和网络机群并行系统来克服向量机的局限。Cray 公司 1983 年以后推出的改进型 CrayX-MP 机，实际上是由 2 ~ 4 台 Cray-1 通过共享存储器互连而成，属于一种并行的流水结构，其速度达到 4 亿次每秒浮点运算。自从 VLSI（Very Large Scale Integration 超大规模集成）微处理器出现以后，采用大量微处理器来组成并行机成为并行机发展的新方向，而且取得很大的成功。因为 VLSI 微处理器可以大规模生产，这样就使得用成千上万个微处理器来组装并行机成为可能，并且 VLSI 微处理器比较便宜，达到同样性能指标时可比传统的巨型机便宜 10 ~ 100 倍，这些优势使大规模并行机的性能迅速提高。

并行机的发展前景令人鼓舞，使许多国家的科研机构及计算机公司竞相投入研究和竞争。比较有名的是美国 Intel 公司 1985 年推出的个人超级计算机 iPSC（Personal Super-Computer）。该机最初以 Intel 80286 作为节点机，采用超立方体互连结构，20 世纪 80 年代后期又更新为用 i860 微处理器作为节点机，使运算速度最高可达 50 亿次每秒。20 世纪 90 年代，Intel 公司又推出 500 个 i860 组成的二维阵列结构的并行机 Touchstone Delta，峰值速度达到 400 亿次每秒。1986 年美国思维公司推出的 Connection Machine CM-1，由 65 536 个 1 位微处理器组成，运算速度为 10 亿次每秒，推出的第一年就生产了 16 台。英国、西德也推出了有特色的并行机，有的并行机峰值速度达到 4 000 亿次每秒。总之，运算速度达几百亿到上千亿次每秒的并行机已经有商品化产品。2000 年 6 月，美国 IBM 公司又推出了迄今世界上最快的超级计算机 ASCI White（Advanced Strategic Computing Initiative White），运算速度达 12 万亿次每秒。21 世纪的运算速度将有可能达到百万亿次每秒。在大规模并行机快速发展的同时，网络并行计算也有很大发展，1994 年 4 月 26 日美国宣布破译了世界上最长最难的 RSA129 密码，这项破译工作在 Internat 网上动用了 1 600 台计算机，共有 600 多人工作 8 个月终于获得成功。

现在我国的并行机技术也发展到一个新阶段。我国自行研制成功的并行机“曙光系列”、“神威系列”及“银河系列”是我国高性能计算机技术的代表。曙光 1000 是我国第一套大规模并行机系统，峰值运算速度为 25.6 亿次浮点每秒。目前曙光、神威及银河新一代产品运算速度已达到数千亿次浮点每秒，我国高性能计算机的研制能力正在不断快速地发展。

1.2 并行处理的定义和分类

1.2.1 并行处理的定义

并行是指两个以上的事件在同一时刻或同一时间段内发生。

并行处理是一种信息处理的有效方式，这种信息处理方式着重于开发计算过程中的并发（Concurrency）事件。并发性的概念有三层含义：同时性（Simultaneity）、并行性（Parallelism）和流水线（Pipelining）。因此并发事件也分三类：

· 同时事件　即同一瞬间发生的事件。

· 并行事件　在同一时间段内发生的事件。

· 流水线事件　在重叠的时间段内发生的事件。

同时性利用资源的重复实现并行，同时性是最严格的并行性。多处理机系统就是利用资源的重复实现并行的例子。

广义的并行性是利用资源共享实现并行。例如分时操作系统可使多个用户的多个进程在一个时间段内并行运行，通过时间分片、资源共享支持并发事件。

流水线通过时间重叠实现并行。向量机就是流水线并行技术的典型代表。

一般来说，高性能的计算机都可称为并行机，因为实际上现代计算机都采用了各种各样的并行技术，如每个机器周期内可执行多条指令的超标量结构、每个机器周期内可执行多级流水操作的超流水线等，尤其是 CPU 执行指令的流水线结构已经成为基本技术，使一般计算机特别是高性能计算机都具有并行能力。由于现在 VLSI 迅猛发展，价格便宜的微处理器大量涌现，从而使资源重复成为实现并行机的重要途径。因此，现在的并行处理就定义为多个资源同时工作，有多个节点处理机的计算机系统才称为并行计算机。

1.2.2 并行处理的分类

在并行处理中，参与并行处理的基本单位的大小称为粒度（Granularity）。粒度一般可分为三级：

· 粗粒度（coarse granularity）　以大块的程序为并行处理单位。

· 细粒度（fine granularity）　以语句、表达式甚至一个简单的算术或逻辑操作为并行处理单位。

· 中粒度（medium granularity）　介于粗细粒度之间。

也可以按编程级划分粒度：

· 粗粒度　作业级（或程序级），多个作业（或程序）并行处理。

· 中粒度　任务级（或过程级），多个任务（或过程）并行处理。

· 细粒度　指令级，多个指令并行执行。

粒度是并行处理技术中的一个基本要素，粒度大小对并行机系统的效率有很重要的影响。如果并行系统的网络通信速度不高，但却采用细粒度并行处理，由于细粒度要求频繁的通信，将使并行机系统效率很低。只有当粒度选择适当，计算问题的并行算法与硬件结构相适应时，才能充分发挥并行机的潜力。节点机数少的并行机一般都采用粗粒度并行，节点机数多的并行机大都采用细粒度并行。商品并行机大多数都属于中粒度并行。

并行处理的最高级是作业级粗粒度并行处理。多个作业并行处理比较简单，因为各个作业相互独立计算和处理，不需要通信，互不干扰。各个作业的处理程序都按传统的串行程序编写方法相互独立编写，不存在并行编程的通信和执行时序问题。通常中、大型机的分时系统就支持多作业并行处理。对于多机系统，可以给各个节点机分配不同的作业并行处理。

任务级并行处理为粗、中粒度（以并行处理的程序单元大小为依据）并行处理。一个用户的任务分为多个子任务，或一个程序分为多个过程（程序块），以便并行地、分工合作地完成同一个计算任务。这样就需要利用并行算法把一个任务分解成多个子任务，按并行程序编程方法编程，各个子任务相互联系进行通信，才能高效地分别在不同的处理机上运行。并

行程序的编写不同于传统的串行程序，需要有相应的并行语言、并行软件、并行处理环境等，这一问题也是影响并行计算机是否能推广应用的重要因素。

指令级并行为细粒度并行。要实现指令级并行，必须事先分析数据相关性。如果先后执行的指令要用到的操作变量之间有依赖关系，称为数据相关。没有数据相关性的指令才能并行执行。

下面的简单例子说明了数据相关性对指令级并行的影响：

（1）有数据相关性

```
        x(0) = 1
        do 100 i = 1，n
        x(i)=2*x(i - 1)
100     continue
```

此段程序中，数组的第 i 个元素 x(i)的计算，要依赖前一步的计算结果即第 i - 1 个元素 x(i - 1)的值，因此运算具有数据相关性，指令无法并行执行。

（2）无数据相关性

```
        do 10 i = 0，n
100     x(i) = 0;
```

此段程序中的计算无数据相关性，n 次循环用并行处理来完成只需要花费执行一条赋值指令的时间。

1.3 并行计算机系统结构的分类模型

并行机系统结构一般按两种方式分类，即指令流/数据流分类法（弗林分类法）和节点机耦合程度分类法。

1.3.1 指令流/数据流分类—弗林分类法

弗林（Flynn）1964 年提出了按指令流和数据流的多少对计算机系统结构进行分类的方法。指令流指计算机在计算过程中执行的一系列指令，数据流指执行指令过程中用到的一组操作数据。弗林分类法将计算机结构分为四类。

1）单指令流单数据流（SISD：Single Instruction Stream & Single Data Stream）

SISD 计算机系统由一个处理器、一个控制器和一个存储器系统组成。SISD 计算机的处理器仅接收和处理单个指令流，而且这个指令流仅在一个数据流上执行。所有的单处理器串行计算机系统属于 SISD 类。

实际上，SISD 计算机是传统的冯·诺依曼计算机，属于共享存储器计算机系统。

2）多指令流单数据流（MISD：Multiple Instruction Stream & Single Data Stream）

MISD 计算机系统由多个处理器、多个控制器、多个存储器部件和连接网络组成，每个处理器都有自己的控制器。所有处理器共享由多个存储部件组成的一个存储系统，各处理器通过连接网络存取存储系统。每个处理器执行一个独立的指令流，多个指令流同时在一个数据流上执行自己的操作。MISD 计算机属于共享存储并行机，其并行性体现在多个处理器并

试读结束，需要全本PDF请购买 www.ertongbook.com

行地完成不同的计算任务。MISD 计算机系统的结构如图 1.1 所示。

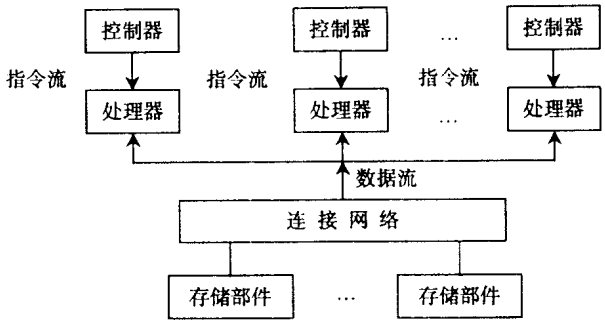

图 1.1　MISD 计算机系统结构

下面的例子说明 MISD 并行计算方式。

例如，判定一维数组 z 中 n 个数是否为素数。

方法：只要逐一判出 z 中某个数 z[i] 是否满足仅能被 1 和 z[i] 本身整除，若满足条件，x[i] 为素数。

为便于理解和简化问题，假定每个 z[i] 只有 p 个因子（除 1 和 z[i] 本身外），记为 $f_{i,1}$、$f_{i,2}$，…，$f_{i,p}$。对一个具有 p 个处理器的 MISD 并行机，可用如下的并行算法判别数组 z 中的 n 个数是否为素数：

For　i = 1　To　n　Do

　　主进程从存储器读 z[i]，广播 z[i] 到 p 个处理器；

For　j = 1　To　p　Do　　（各处理器并行执行）

　　从存储器读 $f_{i,j}$ 到处理器 P_j；　　（处理器编号为 P_1，…，P_p）

For　j=1　To　p　Do　　（各处理器并行执行）

处理器 P_j 判断 $f_{i,j}$ 是否可整除 z[i]；

If（P 个处理器都回答不能整除）Then　z[i]是素数；

Else　z[i]是合数。

上述算法表明，p 个处理机一次并行判别，就可知 z[i] 是否为素数，如果用单机处理，p 次判别才能确定 z[i] 是否为素数。

如果 z[i] 的因子为 n×p 个，按卷帘式负载分配法将数据流的数据分配给 n×p 个处理器，上述算法中的后面两个 For 循环控制可作如下修改：

设 myid 为当前处理器上所运行进程的进程号。

For　j = myid + 1　To　n×p　j = j + p　Do　　（各处理器并行执行）

　　从存储器读 $f_{i,j}$ 到处理器 P_{myid}；

For　j = myid + 1　To　n×p　j = j + p　Do　　（各处理器并行执行）

　　处理器 P_{myid} 判断 $f_{i,j}$ 是否可整除 z[i]；

即，第 1 个处理器判断因子 $f_{i,1}$，$f_{i,p+1}$…，$f_{i,(n-1)p+1}$ 是否能整除 z[i]

　　第 2 个处理器判断因子 $f_{i,2}$，$f_{i,p+2}$…，$f_{i,(n-1)p+2}$ 是否能整除 z[i]

　　⋮

　　第 p 个处理器判断因子 $f_{i,p}$，$f_{i,p+p}$…，$f_{i,(n-1)p+p}$ 是否能整除 z[i]

3）单指令流多数据流（SIMD：Single Instruction Stream & Multiple Data Stream）

SIMD 计算机由一个控制器、多个处理器、多个存储器部件和连接网络组成。控制器负责向多个处理器广播指令，所有处于活动态的处理器分别在来自存储器的不同数据流上并行执行相同的指令流。处理器通过连接网络访问存储器系统。SIMD 计算机系统的结构如图 1.2 所示。

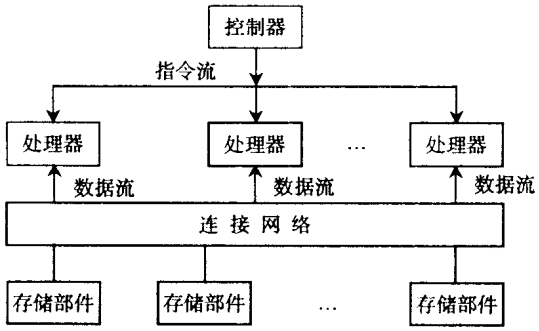

图 1.2　SIMD 计算机系统结构

4）多指令流多数据流（MIMD：Multiple Instruction Stream & Multiple Data Stream）

MIMD 计算机系统由多个控制器、多个处理器、多个存储部件和连接网络组成。每个处理器在各自的数据流上执行自己的指令流，因此 MIMD 计算机系统具有多个指令流和多个数据流。连接网络用来实现处理器之间或处理器与存储器部件之间的通信。

SIMD 计算机和 MIMD 计算机的主要区别在于：SIMD 计算机的各处理器同步运行，同步地使用连接网络，而 MIMD 计算机的各处理器则异步运行、异步地使用连接网络。MISD 计算机系统的结构如图 1.3 所示。

多个节点机组成的并行机都属于 MIMD 类。

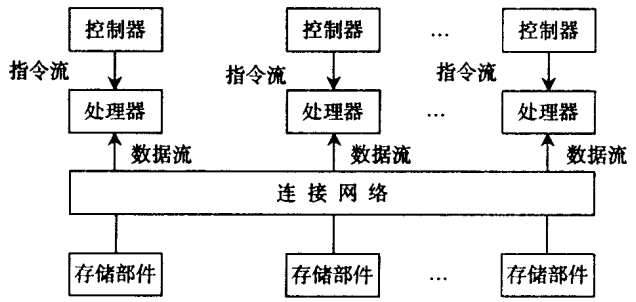

图 1.3　MIMD 计算机系统结构

1.3.2　节点机耦合程度分类法

并行机的基本特征是具有多处理机，每个处理机可称为节点机。各节点机通过互连网络耦合成一个整体，根据耦合的紧密程度可分为紧耦合和松耦合两大类系统，耦合程度是否紧密的判别标准由系统的存储器类型确定。

1）紧耦合系统

具有共享存储器的系统称为紧耦合系统。因为每个节点机通过互连网络与共享存储器相连，每个节点机离开共享存储器都不能独立工作，系统运行过程中各节点机和共享存储器之间通信频繁，故称为紧耦合系统，也称为多处理机系统。

紧耦合系统的优点是整个系统共享一个存储空间，可以沿用传统的高级语言编程，显然编程方便，容易推广使用。紧耦合系统的缺点是，当多个处理机同时访问共享内存时会产生内存争用现象，从而严重影响系统效率，并且同时也会产生通信瓶颈使系统性能下降。为此也有一些方法，如多总线结构（每条总线挂多个处理器，减少通信瓶颈影响）及非均匀存储访问（减少内存争用）等可用于克服上述问题。此外紧耦合系统还有一个缺点是可扩展性差，系统出厂时在硬件结构上已经定型。

MISD 只有紧耦合系统类型。SIMD 和 MIMD 都有紧耦合系统类型，图 1.2 和图 1.3 所示的系统结构图都是具有共享存储器的 SIMD、MIMD 紧耦合系统。

2）松耦合系统

具有分布式存储器的系统称为松耦合系统。因为系统中各节点机有自己的局部存储器，指令和大部分数据都可在本地处理机内访问到，各节点机具有相对的独立性，只有少数共享数据需要在节点机之间通过通信方式交换，这样可使通讯量大大减少，因此称为松耦合系统。

松耦合系统中的节点机都是可以独立工作的计算机，节点机之间主要采用链路（消息传递）方式进行通信。松耦合系统属于分布式存储多计算机系统，也可称为无共享资源结构的系统，其主要的优点是：

① 通过最小化共享资源减小资源竞争带来的系统性能下降；

② 通信网络上的数据通信量较小，缓和了通信瓶颈问题；

③ 具有很高的可扩展性。

由于松耦合系统中的节点机都是独立的计算机，所以节点机数可以大量扩展，并且不会增加处理机间的干扰。

只有 SIMD 和 MIMD 并行机可以具有松耦合系统类型。松耦合系统可以分为三类。

（1）松耦合 SIMD 并行机系统

松耦合 SIMD 系统由一个控制器、多个处理单元（每个处理单元有自己的局部存储器）组成，多个处理单元由连接网络连接在一起，系统的结构如图 1.4 所示。

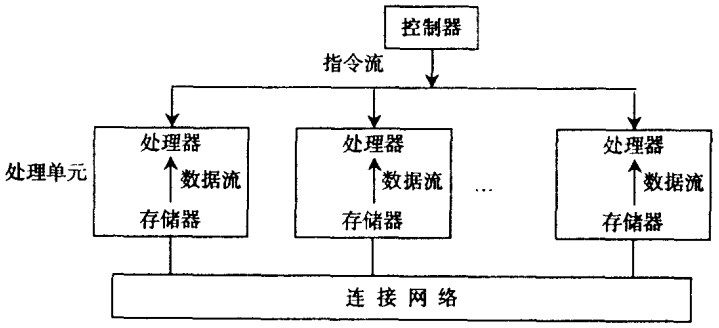

图 1.4　松耦合 SIMD 系统结构

下面以一个简单的向量加法计算为例说明松耦合 SIMD 系统的效率。

设 x、y 为已知向量，计算向量和 z = x + y。完成求和计算的单处理机串行程序为：

```
For  i = 0  To  N - 1  Do
    z[i] = x[i] + y[i]
```

上述串行程序需要执行 N 次求和计算。

如果在 SIMD 并行机中完成相同的求和运算，数据分布存储在 N 个处理器局部存储器中，各处理器并行地执行求和运算，每个处理器只需要执行一条指令：

```
z[i] = x[i] + y[i]          （由第 i 号处理器执行）
```

就完成整个计算。

（2）松耦合 MIMD 并行机系统

松耦合 MIMD 系统由多个处理单元组成，多个处理单元由连接网络连接在一起，每个处理单元有自己的控制器、处理器和局部存储器，系统的结构如图 1.5 所示。

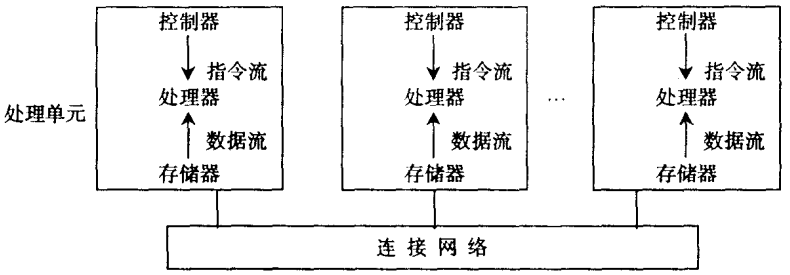

图 1.5　松耦合 MIMD 并行机系统结构

（3）计算机群并行系统

近些年来，由于工作站尤其是 PC 机的价格下跌很快，应用已十分普及，因此兴起了一种新的计算机群（Computer Cluster）并行系统结构。将数量可多可少的 PC 机或工作站用高速网络互连起来，可以很方便地建立起 PC 机群并行系统或工作站机群并行系统，还可以建立成 PC 机和工作站组合机群异构并行系统。机群并行系统的性能可以与高性能巨型机相匹敌，而且价格便宜得多。据有关报导，几百台工作站白天分散工作，晚上互连成机群，并行计算大型题目，其功能可相当于几十台 Cray Y-MP 机！这种机群并行系统可以看作一台松散偶合的 MIMD 并行机，是一种很有意义、很重要的并行系统结构。尤其是对于并行数据库的支持来说，这种并行系统结构是一种最好的并行结构。

在高性能微处理器芯片大量涌现的今天，松偶合系统的一个突出优点就是其规模便于扩展，大规模并行机系统一般都采用松偶合结构。松偶合系统采用消息传递机制实现各节点机间的通信。编写基于消息传递的并行程序对程序员的要求较高，程序员要负责数据分割、任务分配以及考虑数据交换、通信操作的时序问题，并且没有全局地址的概念，程度不易调试。因此，松偶合分布式存储并行系统的编程比共享存储并行系统的编程更困难些。现在大规模并行处理程序主要采用 SPMD（Single Program Multiple Data）模式，即各节点机运行同一个程序处理不同的数据。如果各节点机运行不同的程序处理不同的数据，就构成 MPMD（Multiple Program Multiple Data）模式。由于 SPMD 得到了成功的应用，从而加速了松偶合系统的发展。

第二章　分布式网络并行计算环境

计算机网络是计算机技术与通信技术结合的产物，是以计算机间信息传输为主要目的而连接起来的计算机系统的集合，以便实现计算机系统之间资源共享、通信、信息服务与网络并行计算等功能。本章简要介绍分布式网络并行系统常用的网络拓扑结构以及网络并行计算软、硬件环境。

2.1　常用的计算机网络和网络拓扑结构

2.1.1　计算机网络的发展

计算机网络的发展开始于 20 世纪 60 年代中期。最早出现的一种广域网（Wide Area Network，WAN）ARPAnet 网是由美国国防部下属的高级研究计划署建立的，该网正是 Intrnet 网的前身。20 世纪 70 年代中期，由于小型机的出现及短程通信技术的发展，1975 年美国 Xerox 等公司首先推出局域网（Local a Area Network，LAN）的以太网（Ethernet）技术。以太网采用广播总线技术，传输速率 10 Mb/s。20 世纪 80 年代又出现了令牌网（Token Ring）和光纤网（Fiber Distributed Data Interface）。令牌网是 IBM 公司推出的环状网络产品，环网上的节点机在接到令牌后才能发送消息，发出的消息由网上多站访问部件依次通过环网的每个节点机，即使网上负载很重，数据传输仍有确切的响应时间，效率高，实时性好。光纤网是由光纤为介质构成的共享媒体的环形网络，网络传输也采用类似令牌网的协议，传输速率为 100 Mb/s，一般用作以太网或令牌网等低速网的主干网。由于微处理机的飞速发展和社会信息化的需求，20 世纪 90 年代以来，计算机网络化的趋势尤为明显，出现了高速以太网和 ATM 网等先进技术。尤其是 ATM 网采用了异步传输模式，可以在公共网络上提供传输速率为 100 Mb/s 到数 Gb/s 的传输服务，ATM 是实现高速、宽带网络信息传输与通信的关键技术。

总之，计算机网络正朝着广域国际化、宽带低延迟、多媒体综合技术以及智能化、无线化等方向发展，计算机与通信正在进一步相互渗透和融合，计算机产业的重心将逐步转移到以网络为中心的模式。

2.1.2　网络的拓扑结构

现代并行机系统中采用各种各样的通信网络实现节点机互连，连接网络是并行计算机的关键组成部分，关系到多处理机的通信方式和多存储器模块的存取模式，对并行算法、并行程序有十分重要的影响。高性能的并行机系统要求连接网络结构简单，以最小的造价提供快速、灵活的通信方式。为了达到这个目标，人们提出了大量的连接网络，可分为静态网络和

动态网络两大类。静态网络的互连结构是固定不变的，动态网络的互连结构是可变的或者说是可重构的。每种连接网络都有其优点和缺点，没有哪一种连接网络在任何情况下都是最优的。

1）动态连接网络

动态连接网络可分为总线网络和开关网络两大类，主要用于共享存储器并行计算机系统，实现多存储器模块的共享存取，本节仅简要介绍两种动态连接网络。

（1）共享总线连接网络

共享总线连接网络是最简单的动态连接网络，可以实现多处理器存取共享存储器，也可实现多处理器间的通信。

共享总线连接网络的优点是结构最简单，缺点是资源竞争最严重，通信效率低，但处理器加局部高速缓存可缓解通信瓶颈。以共享总线连接网络互连的并行机系统结构如图 2.1 所示。

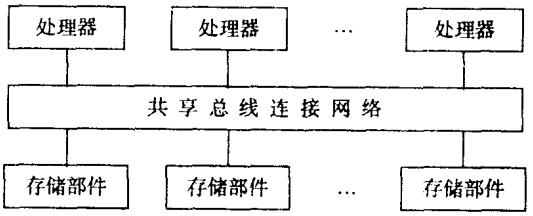

图 2.1　以共享总线连接网络互连的并行机系统结构

（2）开关网络

开关网络又分交叉开关网络和多级开关网络两大类。交叉开关网络主要用于实现多处理器对多个存储模块的共享，使所有处理器能够并行地存取所有共享存储器。

交叉开关网络的优点是具有最小的资源竞争，缺点是具有最高的硬件成本和复杂性。交叉开关网络结构如图 2.2 所示。

交叉开关网络由相互交叉的一组横向通路和一组纵向通路构成，每个交叉点上都有一个开关，通过控制开关可以使任一横向通路与任一纵向通路连通。交叉开关网络特点如下：

① 每条通路上只能有 1 个节点机发送消息。例如当 P_1 向 Q_1 发送消息时，$P_2 \sim P_4$ 就不能向 Q_1 发送消息，但 P_1 可以同时向 $Q_1 \sim Q_4$ 发送消息。

② 当纵、横向通道数均为 N 时，可以有 N 条并行通道同时通信，因此有很高的通信带宽。

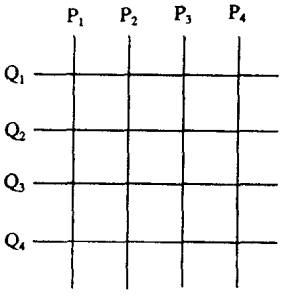

图 2.2　交叉开关网络结构

③ 开关的数量随 N^2 增长。当 N 较大时，所需的开关数量极大，导致成本昂贵。

总之，交叉开关网络和总线连接网络相比，通信性能很好，但结构复杂价格高，一般适合 N 较小的并行机。

由于单总线网络太简单，交叉开关网络太贵，于是人们又提出了折衷的单级或多级开关网络结构，折衷的关键之处是减少了网络的开关数。常用的单级网络有网格网、立方体网、混洗交换网及加减 2^i 网等。多套单级网络串联起来可以组成多级开关网络。

试读结束，需要全本PDF请购买 www.ertongbook.com