

第3章

计算机操作系统



只有硬件系统的计算机,绝大多数用户是无法使用的。操作系统是用户和计算机之间的接口,通过它用户可以方便、高效地使用计算机中的各种资源。本章主要介绍操作系统的基本概念、主要功能及其基本原理和方法,包括处理器管理、存储管理、设备管理和文件管理等。从软件的角度来看,操作系统是管理计算机软件 and 硬件的系统程序,因此掌握操作系统设计的基本原理与方法,对于培养程序设计的思维能力,具有十分重要的作用。

3.1 操作系统概述

计算机系统是由硬件系统和软件系统组成的,用户所面对的计算机系统是在硬件基础上经过软件层层包裹后的计算机。操作系统是配置在计算机硬件上的最底层软件,是对硬件系统功能的第一次扩充,它直接控制、管理各种硬件资源。当计算机安装了操作系统以后,就为其他软件 and 用户提供了良好的工作环境,用户就不再直接操作计算机硬件,而是利用操作系统所提供的命令来操作和使用计算机。因而,操作系统是用户与计算机系统的接口,也是其他软件与计算机硬件的接口。

操作系统(Operating System, OS)是控制和管理计算机软硬件资源,以尽量合理有效的方法组织多个用户共享多种资源并为用户提供操作界面的程序集合。形象地说,操作系统就像一个大家庭里面的大管家,一方面要管理好所有资源,使资源得到合理高效的利用;另一方面要为“主人”提供最为优质的服务。

3.1.1 操作系统的功能

从资源管理的角度看,可把操作系统视为计算机系统资源的管理者。计算机系统资源包括硬件资源和软件资源。硬件资源有处理器、存储器、外部设备;软件资源有程序、数据和文档。现代计算机系统一般采用多任务并发执行方式,并且由多个用户同时联机使用,这使得计算机资源不是由一个程序独占,而是由多个并发运行的程序共享使用,必然会产生对资源的相互竞争。为了使计算机系统中所有软、硬件资源协调一致,有条不紊地工作,就必须有一个地位高于各应用程序的软件来进行自动的、统一的管理和调度,以提高整个计算机系统的性能,这个软件就是操作系统。具体来说,操作系统具有处理器

管理、存储管理、设备管理、文件管理等功能。

资源管理的主要任务有：①跟踪记录资源的使用情况。②合理分配和及时回收资源。③协调多个任务对资源的请求。④科学地组织系统的工作流程，如采用多道、分时和缓冲等技术，提高系统资源的利用率。

从一般用户的角度看，操作系统为用户提供以下方面的主要服务：

(1) 操作系统为用户提供操作界面。用户是通过操作系统来使用计算机的，操作系统程序一般安装在联机硬盘上，用户开机的过程就是将操作系统核心程序加载到内存并执行的过程。操作系统为用户操作计算机提供了操作命令，它由字符方式下的键盘命令和图形方式下的鼠标命令组成。例如图 3-1、图 3-2 分别是 Windows 操作系统的命令行界面和图形用户界面。

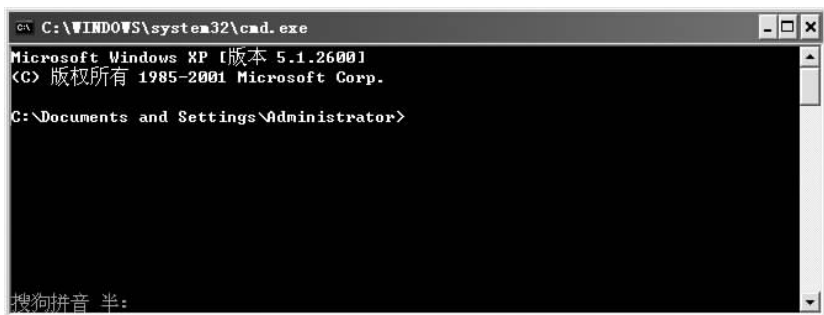


图 3-1 Windows 系统的命令行界面



图 3-2 Windows 系统图形用户界面

在命令行界面中,用户通过键盘将字符型控制命令传送给操作系统,例如,查看目录内容命令 DIR、文件复制命令 COPY、改变当前目录命令 CD、磁盘格式化命令 FORMAT、用来检查网络是否通畅或者网络连接速度的命令 PING 等;在图形用户界面,用户不用记住并输入命令,系统的可执行命令被组织在菜单中,用户只需要用鼠标或键盘选择并单击命令就能控制计算机完成任务。

(2) 操作系统向用户提供方便、统一的接口来使用各种 I/O 设备。操作系统屏蔽了各种 I/O 设备复杂的物理特性和差异,为用户提供了统一接口使用各种设备,如打印机、显示器、USB 设备等。

(3) 操作系统向用户提供方便地使用文件的接口,可使用户方便地完成对文件的复制、删除等操作。

(4) 操作系统为用户自行安装和运行其他应用程序提供了方便。

3.1.2 操作系统的分类

操作系统是基于硬件体系结构的,因此随着计算机硬件的发展,操作系统也在迅速地更新换代,以适应各种硬件环境和各种应用。目前操作系统种类繁多,对操作系统类型的划分有各种分类标准。下面主要是以系统功能为标准介绍不同类型的操作系统。

1. 批处理系统

在计算机应用的早期,把用户交给计算机做的工作称为作业,用户事先把作业的程序、数据和作业说明书一起交给操作员,操作员让成批的作业进入计算机系统,由计算机自动完成这批作业并输出结果,操作系统控制作业的执行和作业间的交替。

批处理操作系统可分为单道和多道两种。批处理单道系统一次只允许一个作业进入内存,I/O 设备与 CPU 串行工作,浪费资源。批处理多道系统允许多个作业同时在内存中运行,提高了资源的利用率、系统的吞吐量,可以实现作业流程的自动化。

2. 分时系统

批处理系统采用脱机控制方式,用户无法干预自己程序的运行。分时操作系统采用时间片轮转技术同时向多个用户提供联机作业控制方式,多个用户能够同时在各自的终端上观察并控制自己程序的运行,及时获得运行结果,还可以进行程序的调试和改错。

分时系统的主要特征是:交互性、独立性、及时性。

最早的分时操作系统是美国麻省理工学院在 1963 年研制的 CTSS(Compatible Time-Sharing System),现在的 Mac OS、UNIX、Windows 等都是分时操作系统。

3. 实时系统

实时系统的主要特点是计算机能及时响应随机发生的外部事件请求,以足够快的速度完成对事件的处理,并且控制实时设备协调一致地工作,或做到协调一致地执行实时任务。所谓外部事件,是指来自与计算机系统相关联设备的服务请求。

实时系统的专用性很强,主要适用于信息处理和过程控制等有实时要求的领域,可以

据此把实时系统分为两种：一是实时信息处理系统，如飞机、火车订票系统，证券交易系统；二是实时过程控制系统，如导弹发射系统、飞机自动导航系统。

4. 通用系统

上述的批处理、分时和实时 3 种是操作系统的基本特征，如果一个操作系统兼有批处理、分时和实时中的两种或多种操作系统的功能，则该操作系统称为通用操作系统。

例如，分时和批处理相结合的系统，通常采用分时运行方式，而当用户少时则以批处理为主的方式运行。

5. 网络操作系统

网络操作系统是为网络上的计算机提供网络通信和网络资源共享功能的操作系统。网络上的每台计算机都是独立的，因此，网络操作系统除了具有普通操作系统的功能外，还要能够具备网络通信能力，管理网络上的共享资源，协调各个主机上任务的运行，并向用户提供多种网络服务和网络接口，保障系统安全等功能。如文件传输、电子邮件、远程打印、远程管理等。

Netware、OS/2 warp、UNIX、Windows NT 和 Windows Server 都是典型的网络操作系统。

6. 分布式操作系统

分布式操作系统对分布资源的控制调度能力比网络操作系统更强，且能调度全网各计算机以合作完成大型的运算处理任务，并为用户提供一个统一的界面实现所需的操作并使用系统资源。对于用户的资源需求，操作系统要在系统的各台计算机上搜索，找到所需资源后进行分配。当某个部分出现故障时，可以调度后备资源，重构系统。

在网络操作系统中，如果一个计算机上的用户想使用网上另一台计算机的资源，他必须指明被访问计算机的地址。而在分布式系统中，用户使用哪台计算机的资源用户是不必知道的，是由以全局的方式管理全网计算机资源的分布式操作系统自动分配的。分布式系统和网络系统的区别在于操作系统的不同，而不是硬件连接。

代表性的分布式操作系统有：Amoeba、Chorus、Mach 和 DCE。

7. 嵌入式操作系统

嵌入式操作系统是为嵌入式硬件系统设计的系统软件。嵌入式硬件不同于一般的计算机硬件系统，它是由嵌入式处理器及相关支撑设备组成的，典型的处理器有 ARM 系列等。嵌入式系统是与应用紧密结合的，它具有很强的专用性，可以结合实际系统的需求进行合理的裁剪利用。嵌入式操作系统一般固化在嵌入式系统的 ROM 中，而不是存储在磁盘中，通常包括与硬件相关的底层驱动程序、系统内核、设备驱动接口、通信协议、图形界面和标准浏览器等。嵌入式操作系统除了具备一般操作系统的基本功能外，还要求具有很强的针对性、强稳定性、强大的网络功能，并且具备实时高效、软件固化、占用内存较小等特点。

目前具有代表性的嵌入式操作系统有： μ CLinux、Chorus、Diba、Navio、OS-9、Psos、QNX、VxWorks 和 Win CE。智能手机的安卓操作系统、塞班操作系统等也属于嵌入式操作系统。

嵌入式操作系统正向着微内核、模块化、虚拟机、多任务多线程方向发展。

3.1.3 操作系统的特征

操作系统的主要特征体现在以下 4 个方面。

1. 并发性

并发性是指多个程序同时在系统中运行。并发性体现了操作系统同时处理多个活动事件的能力。通过并发减少了计算机中各部件间由于相互等待而造成的资源浪费，提高了资源利用率。

2. 共享性

共享性是指计算机系统资源能够被并发执行的程序共同使用，共享是在操作系统控制下实现的。操作系统要对资源进行管理与调度，使得并发执行的多个程序能够合理地共享这些资源。共享的实质是多个并发的程序在操作系统的统一指挥下交替使用资源。

3. 虚拟性

虚拟性是指操作系统通过某种技术将一个实际存在的实体变成多个逻辑上的对应体，并发的多个程序访问这些逻辑对应体，提高了实体的利用率。例如，虚拟内存技术，即匀出一部分硬盘空间来充当内存使用，当计算机运行程序所需的内存不足时，操作系统会将计算机的内存和硬盘上的部分空间组合，统一进行逻辑编址，这样，程序访问的便是空间扩大的虚拟存储器。操作系统的虚拟性体现在 CPU、内存、设备和文件管理等各个方面，正是操作系统的虚拟性才把裸机变成了功能更强更易于使用的虚拟机。

4. 异步性

异步性也称为不确定性，是指在多个程序并发运行环境中，每个程序何时开始执行、何时暂停、推进速度和完成时间等都是不确定的。因此，操作系统的设计与实现要充分考虑到各种可能性，以便稳定、高效、可靠、安全地达到程序并发和资源共目的。

3.2 处理器管理

处理器是计算机系统最重要的资源，计算机的工作过程就是快速执行程序的过程，而程序只有获得了处理器才能运行。因此，处理器管理是操作系统的重要组成部分，它的主要任务是对处理器的分配和运行进行合理有效的管理与调度，以提高处理器的利用率。处理器管理可分为作业管理和进程管理两种。

3.2.1 作业管理

1. 作业的概念

作业是用户在一次事务处理过程中要求计算机所做工作的总和。它由程序、数据和作业说明书组成。用户在使用计算机解题时,需要编写相应的程序,称为源程序,还要准备好程序运行时需要的初始数据,也称为输入数据,同时向计算机系统提出控制要求,计算机系统按照要求完成解题。在高级语言中由字符代码组成的源程序计算机是无法识别运行的,计算机系统要调用编译程序进行编译,形成二进制代码的若干目标程序段;再调用连接程序将目标程序连接成可以执行的程序;最后调用运行程序运行可执行程序,输出计算结果。我们可以把这整个过程作为一个作业,也可以将部分过程作为一个作业。对于普通用户,虽然不使用作业这个概念,但实际上也是在运行作业。

2. 作业管理

在多道程序设计系统中,系统要在多个作业中按一定的策略选取若干个作业,为它们分配所需资源,让它们能够同时执行,这就是作业调度。作业是在操作系统控制下执行的。

由作业调度所选择的作业只是有了获得处理器的资格,但不一定就能占有它,只有被进程调度所选择的一个进程,才能真正地在实际的处理器上运行。

3.2.2 进程管理

1. 进程的概念

处理器的任务就是运行程序,程序以文件的形式保存在外存储器上。一个程序开始执行,就从外存储器被调入内存开始运行。一个程序被加载到内存,系统就创建了一个进程。

进程的定义:程序及其数据在计算机上的一次执行过程,是操作系统进行资源分配和调度的一个独立单位。

如果说程序是提供给计算机完成任务的一系列操作步骤的话,那么进程就是按照这些步骤具体的执行过程。程序和进程是两个完全不同的概念,但又有密切联系,它们之间的主要差异是:

(1) 程序是一个静态的概念,它是存放在外存上的程序文件,而进程是程序的一次执行过程,是一个动态的概念。

(2) 由于程序只是外存上的文件,所以可以脱离机器长期保存;而进程是执行着的程序,执行完毕,进程也就不存在了,所以进程的生命是暂时的。

(3) 程序和进程不存在一一对应关系,一个程序可多次执行,因此可产生多个不同的进程。

(4) 各个进程在并发执行过程中会产生相互制约的关系,所以前进速度是不可预测的,而程序是静态的,不存在这种异步特征。

2. 进程的特征

从前面的进程定义、程序与进程的区别,可以分析出进程具有如下特征。

(1) 动态性。一方面进程是程序及数据的一次执行过程,程序从开始运行到结束,中间可能暂停多次,这个过程就具有动态性;另一方面,进程由创建而产生,由调度而执行,因得不到资源而暂停,因撤销而消亡,所以进程具有一定的生命周期。可见,动态性是进程的基本特征。

(2) 并发性。这是指多个进程共存在内存中,在一段时间内同时运行。在多道程序系统中,并发性是进程的重要特征,也是操作系统的重要特征。进程的并发性提高了系统的效率。

(3) 独立性。进程是一个能独立运行、独立获得资源和独立调度的基本单位。只有进程才能向系统申请资源,才能作为一个独立单位参与运算,未建立进程的不能。

(4) 异步性。异步性是指每个并发执行的进程按各自独立的、不可预知的速度向前推进。进程完成的时间是不可预知的,因而具有异步性的特征。

(5) 结构性。从结构上看,进程至少由程序、数据集合和进程控制块 3 部分组成。程序是指示 CPU 完成本进程所需要的操作,数据集合则是执行进程的进程所需要的数据,进程控制块是一个数据结构,是进程存在的唯一标识,用来跟踪记录进程动态变化着的各种调度信息。

3. 进程的基本状态及其转换

(1) 进程的基本状态

根据进程的特性,一个进程从创建到撤销的过程中,具有执行-暂停(等待)-执行的活动规律,也就是说要经过不同的状态。在操作系统中,进程通常分为以下 3 种状态:

① 就绪状态。是指进程的外部条件满足,而 CPU 正被其他进程占用,所以进程暂时不能运行,处于等待 CPU 的状态。系统中处于就绪状态的进程有多个,操作系统将这些进程组织成一个队列,称为就绪队列。一旦 CPU 空闲,就从队首取进程执行。

② 执行状态。进程已获得 CPU,正在执行。在单处理机系统中,任意时刻只有一个进程处于执行状态。而在多处理机系统中,有多个进程同时处于执行状态。

③ 阻塞状态。是指正在执行的进程由于等待 I/O 设备或某个事件的完成,暂时不能执行,即使 CPU 空闲也无法使用。系统根据进程阻塞的原因不同,把进程组织成多个队列,称为阻塞队列。

(2) 进程状态的转换

一个新创建的进程通常处于就绪状态,在进程执行过程中,进程不可能固定处于某个状态,它将随着自身的推进和外界条件的变化,由当前状态转换为另一种状态。具体的状态转换有:

① 就绪状态转换为执行状态。处于就绪状态的进程,在调度程序为之分配了 CPU 后,该进程就变为执行状态。

② 执行状态转换为就绪状态。正在执行的进程,由于分配给它的时间片已经用完而

被暂停执行时,该进程便由执行状态转回到就绪状态。

③ 运行状态转换为阻塞状态。一个处在执行状态的进程,如果因等待系统资源或发生某事件而使进程的执行受阻,如等待打印机或等待另一个进程的计算结果,使之无法继续执行,这时该进程就要放弃 CPU,由执行状态转换到阻塞状态。

④ 阻塞状态转换为就绪状态。在阻塞队列中的进程,如果获得了等待的资源或某事件已完成,就由阻塞状态转换为就绪状态。图 3-3 给出了进程的 3 种状态及其转换关系。

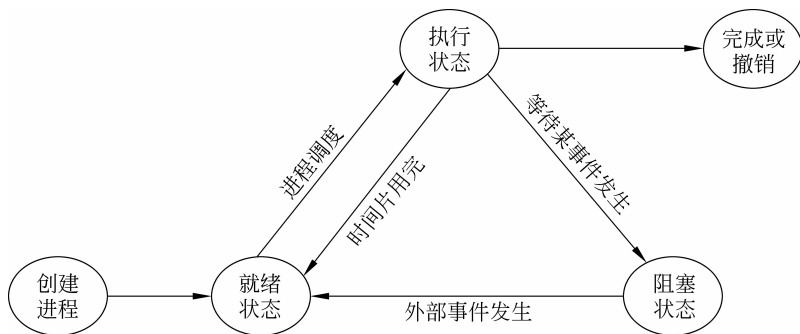


图 3-3 进程状态及转换关系

4. 处理器调度

处理器调度是指操作系统为进程分配处理器时所采用的原则和算法。

在单处理器系统中,程序有两种运行方式:单道程序顺序执行和多道程序并发执行。现代计算机系统采用多道程序并发执行。

单道程序顺序执行是指任一时刻只允许一个程序在系统中执行,一个程序执行结束后才能执行下一个程序。也就是说任一时刻内存中只有一个进程。这种运行方式的优点是实现简单,不需要在多个进程之间转换;缺点是资源利用率低,在处理器忙时设备空闲,在设备忙时处理器空闲,处理器和设备串行工作。并且,这样的系统无法提供多用户同时联机使用方式。

多道程序并发执行是指计算机内存中同时存放了多个相互独立的进程,处理器被各个进程分时共享,各个进程交叉运行,形成了“微观上串行、宏观上并行”的运行方式。该种运行方式可使处理器和设备并行工作,当某个进程在进行输入输出操作时,可以同时有另一个进程使用处理器计算。这种方式提高了系统资源利用率,但也引起了各个进程对处理器的竞争,如果有两个或更多的进程处于就绪状态,而只有一个 CPU 可用,那么 CPU 就必须在多个就绪进程中进行选择。操作系统中完成这个工作的程序称为调度程序,该程序使用的算法称为调度算法。下面介绍几种常用的调度算法。

(1) 先来先服务

进程按照它们请求 CPU 的时间先后顺序使用 CPU。在系统中所有就绪进程按照请求顺序链接成一个就绪进程队列,CPU 选择队首进程执行,执行中若有其他进程进入就被安排在队列尾部。一个进程一旦获得 CPU,便一直执行下去直到终止,除非进入阻塞状态才会释放 CPU。当阻塞状态的进程变为就绪状态时,又会被排到队列的末尾。这个

算法实现简单,但服务质量较差。

(2) 时间片轮转调度算法

轮转调度算法是把 CPU 的处理时间划分成若干时间片,并且把所有就绪进程按先入先出的原则组织成一个就绪队列。每次调度时把 CPU 分配给队首进程,当该进程运行完一个时间片后,系统的计时器发出中断请求,强迫该进程让出处理器,并把它放入就绪队列队尾,等待下一轮调度,直到运行完成。随后,进行进程切换,再把 CPU 分给当前就绪队列队首进程。同样让它运行一个时间片,如此循环。这种方法主要在分时系统中采用。图 3-4 给出了就绪队列中 A、B、C、D、E 五个进程的轮转调度示意图。

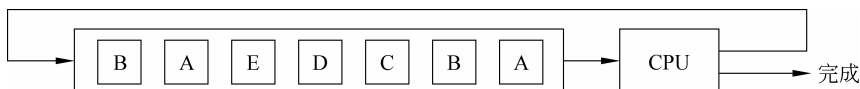


图 3-4 时间片轮转调度

(3) 优先级调度

轮转调度算法将所有进程视为同等重要,事实上系统中的进程是有轻重之分的,如系统进程要重于用户进程。优先级调度就是系统或用户按照某种原则为每个进程指定一个优先级,在调度时,调度程序根据优先级,允许优先级更高的就绪进程先运行。

5. 利用 Windows 任务管理器查看进程

进程管理是操作系统的重要组成部分,在 Windows 操作系统中查看进程的运行状态可以通过任务管理器。打开任务管理器有 3 种方法:方法一是按 Ctrl+Shift+Esc 组合键;方法二是按 Ctrl+Alt+Delete 组合键选择任务管理器按钮;方法三是右击任务栏空白处,在快捷菜单中选择任务管理器。打开任务管理器后的界面如图 3-5 所示。



图 3-5 任务管理器

使用任务管理器可以显示应用程序、进程和服务的运行状态,显示 CPU、内存和网络连接等的性能图表。

利用任务管理器的“应用程序”和“进程”选项卡,可以方便地管理程序和进程。使用任务管理器还可以查看每个进程的详细信息,可以查看计算机运行状态的基本信息,结束用普通方法无法结束的应用程序。

3.3 存储管理

操作系统中管理分层存储体系的部分称为存储管理器。寄存器和高速缓存由硬件管理,存储管理的主要任务是有效地管理内存。

内存和处理器组成计算机系统的主机系统。处理器运算的数据来自内存,运算结果还要写入内存。而用户程序则以文件的形式存储在外存上,要执行用户程序,首先要将其装入内存,那么用户程序怎么才能申请到内存,如何装入到内存的指定区域,内存容量不够怎么扩充,程序执行完怎么释放空间等问题,则是由操作系统中的内存管理模块来解决的。

内存管理的主要任务是对内存空间的分配和回收,内存与外存交换信息的管理,实现地址变换,实现内存的逻辑扩充和存储保护等工作。内存管理的目标是提高内存的利用率,内存管理的具体功能如下:

(1) 内存的分配与回收。系统首先要将整个的内存区域进行分割,如划分为操作系统区和应用程序区,再根据不同的分配方法,把大的区域划分为大小不同的子区域以适应不同的用户对内存的请求。这需要在操作系统中设置相应的数据结构,用于记录这些区域的起始地址、大小、状态和当前内存的使用情况。当有用户请求内存时,以适合的空间分配给它,并修改使用状态,以反映当前系统状态。内存对于计算机系统来说,是一种价格昂贵而容量不足的资源,所以当程序运行结束后,要及时回收该程序所占有的空间,来满足其他等待装入内存运行的程序。

(2) 地址转换。我们知道,一条指令是由操作码和操作数组成的,程序是有序的指令集合。用户编程时不可能知道该程序被装在内存的什么地方运行,因此它就不可能用实际的物理地址来编程,操作数地址也不可能是实际的物理地址。用户编程时是按程序和数据放在从零开始编址的空间中,把这个空间称为逻辑地址空间,这时操作数所在的地址称为逻辑地址,跟实际的内存无关。如一个操作数的逻辑地址为 500,则该操作数的位置在离程序开始 500 距离处。在多道程序并发执行系统中,只有当程序装入内存时,才开始将逻辑地址改为内存的实际地址(物理地址),这个转换是由操作系统的重定位装入程序完成的,称为地址转换。具体操作就是将程序装入内存的起始地址加上逻辑地址得到正确的内存地址。

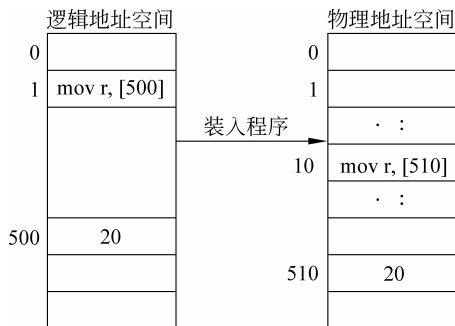


图 3-6 地址转换

图 3-6 给出了一个地址转换示意图,图中假设有一条指令“mov r,[500]”,意思是

500 号地址(逻辑地址)的内容送到寄存器 r 中。存储该条指令的逻辑地址是 1 而物理地址是 10, 因此在地址转换时将操作数对应的存储单元由逻辑地址 500 修改为物理地址 510, 实际是将物理地址 510 处的存储单元内容送到寄存器 r 中。

(3) 内存共享和保护。内存的共享是指两个以上的进程使用同一个子程序段或数据段, 该部分在内存中只保留一个副本。内存的共享节省了内存的空间, 减少了内外存的数据

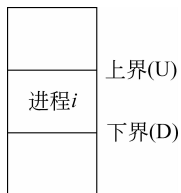


图 3-7 内存上下界

交换, 提高了系统的效率。内存保护的任务是确保每道用户程序都在自己的内存空间中运行, 互不干扰、相互保密, 即不允许访问其他程序的存储空间, 尤其是不允许用户程序访问操作系统的程序和数据(只能通过系统调用), 这样才能保障操作系统安全稳定, 用户程序正确执行。如图 3-7 所示, 操作系统常用上下界限保护的存储保护机制, 当进程访问内存越过上界(U)或小于下界(D)时, 能及时发现并进行处理。

(4) 内存的逻辑扩充。在多道程序环境中, 内存资源比较紧张, 内存容量的有限性直接制约了进程的并发执行效率和进程规模, 往往会出现多道程序对内存的需求总和超过实际内存容量, 有时甚至出现一个进程的空间要求都难以支持的情况。例如, 一个三维图像处理进程, 一幅图像本身的数据空间几乎达到了实际内存的大小。如果要求程序和数据必须全部转入内存才能进行的话, 这样的应用系统就无法在计算机系统上使用。为了解决内存容量不足的问题, 必须实现内存的扩充。

这里说的内存扩充不是指添加物理内存, 而是从逻辑角度扩充内存, 即采用虚拟存储技术将主存和辅存的地址空间统一编址。用户程序在运行之前, 操作系统采用了一些技术手段, 无须把整个程序全部装入内存, 只将当前要使用的部分装入内存, 便可开始运行, 其余暂时不用的部分先存放在外存中。而在运行期间由系统自动地、动态地置换出暂时不用的程序和数据, 腾出足够的空间来装入接下来要运行的、还在外存中的程序或数据。这样, 便可在较小的内存空间中运行较大的程序, 而且可以使更多的进程在内存中并发执行。在用户看来, 好像内存容量变大了, 而实际并非如此, 这种方法并没有扩充物理内存, 但其作用相当于扩充了物理内存。用户看到的大容量内存是虚拟的, 是用 CPU 时间和外存空间来换取的, 是内外存的有机结合, 这个存储空间就是虚拟存储器。

所谓虚拟存储器是用户能将其作为可编址内存对待的存储空间, 它是由操作系统提供的一个假想的特大的存储空间。如图 3-8 给出的是 Windows 系统中在“系统属性”中的“虚拟内存”项, 在其中可设置虚拟内存的大小。

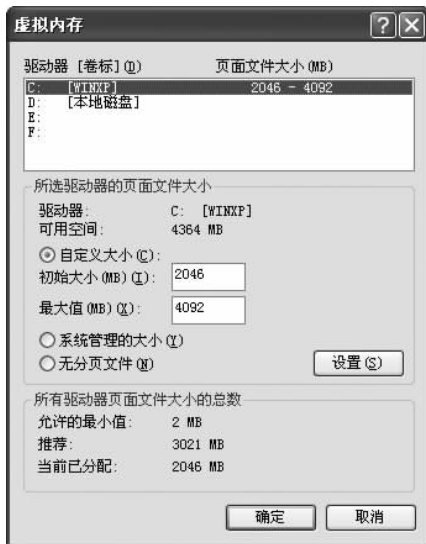


图 3-8 虚拟内存

3.4 设备管理

设备管理是用户与外设的接口,这里的设备是指除 CPU 和内存以外的各种设备,如键盘、显示器、磁盘、打印机、终端等,通常称为 I/O 设备。这些设备种类繁多,物理性能各不相同,一般用户很难直接使用。操作系统的设备管理为用户方便使用各种设备提供了统一的接口,用户只需要输入操作系统提供的命令就可以使用设备。例如,用户写完文档要使用打印机打印,只需用鼠标单击打印命令即可。设备管理的主要任务是管理各类外部设备,包括分配、启动和故障处理等,合理地控制 CPU 与外设间的数据交换。最大程度地实现 CPU 和 I/O 设备的并行工作。设备管理的目标一是方便用户使用 I/O 设备,二是在多道程序环境下提高设备的利用率。设备管理的具体功能如下:

(1) 设备的分配与回收。根据用户的 I/O 请求分配相应的设备,当多个并发进程对同一设备提出 I/O 请求时,要采用一定策略来决定将此设备分配给哪个进程。通常有两种分配算法:一是先来先服务算法,即所有请求按请求发生时间先后排队,当设备可用时,将设备分配给队首元素对应的进程,即最先提出 I/O 请求的进程;二是优先级算法,按一定原则设置进程优先级,按进程优先级由高到低排队,当设备可用时,分配给队首元素对应的进程,设备用完后,应立即加以回收。

(2) 实现设备独立性。设备独立性是指提供统一的 I/O 设备接口,使用户编写的应用程序与实际使用的物理设备无关,设备独立性也称设备无关性。例如用户编写的文档要保存在优盘上,无论什么品牌、什么型号的优盘接到 USB 口,系统都可以识别,完成对优盘的各种操作。

(3) 缓冲区管理。缓冲区是操作系统在主存中开辟的一个存储区域,专门用于临时存放 I/O 数据。输入设备输入的数据先传入到输入缓冲区,进程需要输入数据时,操作系统可把缓冲区中的数据传送到进程的数据存储区。进程需要输出时,先传送数据到输出缓冲区,等待输出设备提取。由此可见,缓冲管理可以提高 CPU 和 I/O 设备的并行性,以及 I/O 设备之间的并行性,从而提高整个系统的效率。

(4) 设备驱动。设备种类繁多,特性各异,必须有程序来控制设备的运行。操作系统对每类设备分别设置不同的设备驱动程序,控制设备完成 I/O 操作,实现设备的即插即用(Play-and-Play, PnP)。

3.5 文件管理

3.5.1 文件的概念

文件是有名称的一组相关信息的集合,每个文件都要用一个名字作为标识,称为文件名。在计算机系统中,所有的程序和数据都是以文件的形式存放在计算机的外存储器上的。例如,C 源程序、Word 文档、各种可执行程序等都是文件。操作系统对文件按名存取,通常文件名由两部分组成,文件主名和扩展名中间用符号“·”隔开。文件主名是有意

义的字符或数字组合,命名文件时要做到“见名知意”,以使用户识别。在搜索文件时可以使用通配符“*”和“?”。“*”代表任意一个字符串,“?”代表任意一个字符。文件的扩展名通常表示文件的类型。可以从不同角度对文件进行分类。

(1) 按文件的用途进行分类

① 系统文件:包括操作系统内核、系统应用程序等,系统文件对于系统的正常运行是必不可少的,这些文件不对用户直接开放。

② 库文件:包括标准的和非标准的子程序库,这类文件允许用户通过一定的方式调用,但用户不得修改。

③ 用户文件:用户自己的文件,如源程序、目标代码、数据、文档、视频、音频和图片等。

(2) 按文件数据的形式进行分类

① 源文件:源代码和数据构成的文件。

② 目标文件:源程序经过编译程序编译后的二进制目标代码文件。

③ 可执行文件:链接程序将目标代码链接后生成的可以执行的文件。

3.5.2 文件系统的概念和功能

在操作系统中,负责管理和存取文件信息的部分称为文件系统。文件系统的任务是为用户提供一种简便、统一的存取和管理文件的方法,实现对文件的按名存取;解决文件在外存上的存储、共享、保密和保护问题,采用合理的分配策略来提高外存资源的利用率和提高文件的存取速度。文件管理的具体功能如下:

(1) 实现文件的按名存取

用户可以按照文件名访问文件,而不必考虑各种外存的差异,也不必了解文件在外存上的具体物理位置以及是如何存放的。

(2) 目录管理

为了能方便地在外存中找到所需文件,要为文件建立目录,每个文件对应一个目录,包含文件名、文件属性、文件的存放路径等。目录管理的主要任务是建立外存中文件的目录结构。操作系统通常采用树型目录结构,便于文件的分类与检索,提高文件的访问效率。操作系统提供对目录和文件操作的手段,包括文件的创建、删除、复制、打开、关闭、读、写等。

(3) 文件存取权限控制

当用户对某一文件操作时,首先检查该用户是否具有对该文件的此种操作权限,有则允许,否则拒绝其对该文件的此种操作。文件存取权限控制机制解决了文件的安全和保护问题,文件信息的安全在多用户系统中尤为重要。

(4) 文件存储空间的管理

当用户新建文件时,文件系统必须为用户分配存储空间,当文件被删除后,系统及时回收文件所占用的存储空间。由于删除文件会造成外存储空间“碎片”很多,操作系统提供碎片整理程序,把“小空间”变为“大空间”,提高存储空间的利用率。

3.5.3 文件的操作

操作系统提供了对文件进行创建、打开、读写、关闭、删除等各种操作,具体如下:

(1) 创建文件

在创建文件时,系统会为新文件分配所需的磁盘空间,建立一个目录项来记录文件名及其在外存的位置等属性。

(2) 打开文件

用户要访问已经存在的文件,必须先打开文件。文件系统根据文件名,在文件目录区找到该文件的目录,将这个目录复制到内存中。

(3) 读写文件

对已经打开的文件,用户就可以对其进行读写。通过读指针,将位于外存的数据读入到内存缓冲区;通过写指针,将内存缓冲区的数据写入到位于外存的文件中。

(4) 关闭文件

当不再使用文件时,要及时关闭文件,以释放内存空间,节约系统资源。

(5) 删除文件

当不再需要某个文件时,便可以从文件系统中删除。系统先找到要删除的目录项,使之成为空,然后回收该文件的存储空间。一般系统中不允许删除正处于打开状态的文件,应该先关闭再删除。

3.5.4 图形界面下用户对文件与文件夹的操作

1. 驱动器、盘符和根目录

通常文件存放在磁盘上,磁盘容量很大,为便于管理,人们通过操作系统的分区程序把一个磁盘划分为几个分区,每一个分区称为一个卷,也称为逻辑驱动器,都有盘符。假如一个磁盘分了3个区,则盘符分别为C:、D:、E:。在磁盘分区时需要确定文件系统,不同的操作系统所支持的文件系统不同,目前,Windows支持的文件系统有FAT32和NTFS。磁盘分区后还不能使用,还需要对每个分区分别进行格式化,格式化完成以下任务:

(1) 把磁道划分为一个个扇区,每个扇区容量512B。

(2) 安装文件系统,建立根目录。

如C盘格式化后,C:\称为C盘根目录,C盘上的所有目录和文件被组织成以C:\为根的树形结构。

2. 文件夹

前面讲的目录也称为文件夹,用来分类存放大量的磁盘文件。用户可以在磁盘上创建文件夹,在文件夹下再创建子文件夹,将文件分门别类地存放在不同的文件夹中,这样就会将磁盘上的所有文件夹和文件组织成树状结构,如图3-9所示。图中矩形表示目录,椭圆表示文件。

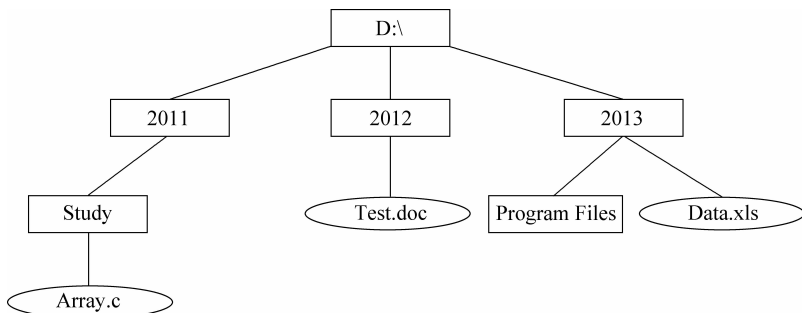


图 3-9 树状目录结构

在图 3-9 中,在 D 盘根目录下创建了 3 个一级子目录 2011、2012、2013,再在 2011 目录下创建二级子目录 Study,在 2013 目录下创建二级子目录 Program Files。所有文件都在叶子节点。这种结构像一棵倒置的树,树根为根文件夹,树中每一个分支为子文件夹,树叶为文件。为了查找方便,用户可以将同类型的文件放在同一个文件夹中,也可以将同一个项目的相关文件放在同一文件夹中;同名文件可以放在不同文件夹中,实现备份功能;也可以将访问权限相同的文件放在同一文件夹中,以保障文件安全;还可以将某一时间段的所有文件放在同一文件夹中,集中管理。

树状目录结构建立后,所有的文件夹和文件就会有自己的固定位置,要访问树中的文件,我们只需从树根(根目录)开始,沿着树枝(文件夹)一直到树叶(文件),经过的途径称为文件路径。如图 3-9 中,C 源文件“Array. c”的路径为“D:\2011\Study\Array. c”(文件夹之间用符号“\”分隔)。这种从根目录开始直到该文件为止的所有目录名称为文件的绝对路径。文件的绝对路径会记录在文件属性中,它表示了文件在磁盘中的绝对位置。如果我们当前不在根的位置,而在一分支上,要访问另一分支上的文件,首先回到根位置,再沿绝对路径即可。如果当前目录为 Program Files,则“Array. c”文件的路径为“..\..\2011\Study\Array. c”(用“..”表示上一级目录),这种文件路径称为相对路径,即从当前目录到文件的途径。

3. 文件类型

在大多数操作系统中,文件扩展名表示文件类型。在不同操作系统中,表示文件类型的扩展名也不尽相同,常见的文件扩展名及其意义如表 3-1 所示。

表 3-1 文件扩展名及其意义

文件类型	扩展名	说明
可执行程序	EXE、COM	可执行程序文件
源程序文件	C、CPP、BAS	程序设计语言的源程序文件
Office 文档	DOC、XLS、PPT	Word、Excel、Powerpoint 创建的文档
流媒体文件	WMV、RM、QT	能通过 Internet 播放的流式媒体文件
压缩文件	ZIP、RAR	压缩文件

续表

文件类型	扩展名	说明
网页文件	HTM、ASP	前者是静态的,后者是动态的
图像文件	BMP、JPG、GIF	不同格式的图像文件
音频文件	WAV、MP3、MID	不同格式的声音文件

4. 文件和文件夹的操作

(1) 创建文件夹或文件

为了便于管理,可以创建不同的文件夹来存放不同用途的文件。也可以在磁盘某位置新建某类型的文件。创建文件夹,在 Windows 系统中用“新建”命令。

(2) 选定文件夹或文件

对文件或文件夹进行操作之前,首先选定要操作的文件或文件夹。具体的操作有:

- ① 选定单个文件夹或单个文件:单击文件夹或文件。
- ② 选定相邻的文件夹或多个文件:先单击第一个文件或文件夹的图标,按住 Shift 键,再单击最后一个文件或文件夹图标。
- ③ 选定不相邻的文件夹或多个文件:先单击第一个文件或文件夹的图标,按住 Ctrl 键,再依次单击要选定的文件或文件夹。
- ④ 选定全部文件和文件夹:选择“编辑”菜单中的“全部选定”命令或按 Ctrl+A 键。
- ⑤ 取消选定:单击窗口的空白处。

(3) 对文件及文件夹的操作

包括更改文件或文件夹的名称、搜索文件或文件夹、移动文件或文件夹、复制文件或文件夹、删除文件或文件夹等操作,表 3-2 给出的是“编辑”菜单中对文件及文件夹的操作,表 3-3 给出的是“文件”菜单中对文件及文件夹的操作,表 3-4 给出的是“恢复文件”和“查找文件”的操作。

表 3-2 “编辑”菜单中文件和文件夹的操作

作用	“编辑”菜单中的命令	鼠标拖曳	快捷键或键盘命令
复制	“复制”、“粘贴”	直接拖曳(不同驱动器) Ctrl+拖曳(同一驱动器)	Ctrl+C、Ctrl+V
移动	“剪切”、“粘贴”	Shift+拖曳(不同驱动器) 直接拖曳(同一驱动器)	Ctrl+X、Ctrl+V
删除	“删除”	直接拖曳到回收站	Del

表 3-3 “文件”菜单中文件和文件夹的操作

作用	“文件”菜单中的命令	说明
发送	“发送”	可将文件发送到磁盘、我的文档、邮件接收者等地方,也可以用该命令在桌面创建快捷方式

续表

作用	“文件”菜单中的命令	说 明
新建	“新建”	新建文件夹、快捷方式或各种类型的文档
改名	“重命名”	重新命名文件或文件夹的名称
查看属性	“属性”	查看文件或文件夹的属性

表 3-4 “恢复文件”和“查找文件”的操作

作用	操 作 命 令	作 用
恢复文件	通过“回收站”或“编辑”→“撤销删除”	从回收站恢复到原有位置
查找文件	“开始”→“搜索”	搜索所需的文件

5. 创建文件与文件夹的快捷方式

(1) 什么是快捷方式

磁盘上的文件成千上万,但用户经常使用的文件并不是很多,如果每次访问文件都按路径去寻找,会很麻烦,操作系统为用户提供了一种方便、快速的访问方法,可以创建文件或文件夹的快捷方式。

快捷方式是原文件或文件夹的替身,它不是这个文件本身,而是指向这个文件的指针,如同一个人的照片。因此,打开快捷方式就打开了相应的文件,但删除快捷方式却不会删除它的原身。

快捷方式通常放在对用户来说最直观最便捷的地方,如桌面。桌面是计算机启动后,用户第一眼看见的整个屏幕背景。也可以放在用户感觉方便的磁盘上的任何地方。

(2) 创建快捷方式的方法

选定对象—>右键单击—>快捷菜单中选择“发送到”命令即可创建桌面快捷方式。创建好的快捷方式图标左下角有一个弧形箭头,如图 3-10 所示是 Winword 快捷方式。



图 3-10 Winword 快捷方式

桌面上的快捷方式可以复制到磁盘任何地方。在选定对象时,如果不知道对象路径,可使用“开始”菜单中的“搜索”命令查找到所要创建快捷方式的文件或文件夹。

本章小结

通过本章的学习,了解了操作系统的作用和工作原理,作为一个大管家,它承担着处理器管理、存储管理、设备管理、文件管理等一系列繁重的工作,没有它,计算机将陷入一片混乱之中。

在本章中,首先介绍了操作系统的基本定义和基本功能,对于用户来说操作系统是用户与计算机硬件之间的接口,使用户能便捷、有效地使用计算机。而对于系统程序员来说,操作系统的任务是高效地管理整个计算机系统的软硬件资源,协调各进程对资源的使

用冲突,提出使用资源的统一方法,最大限度地实现各类资源的共享并提高资源的利用率。

其次,本章介绍了操作系统的基本特征和分类。操作系统是一个大型复杂的并发系统,并发性和共享性是其两个最重要的特征,并发和共享能更好地改善资源利用率并提高系统效率,但有时却会引发一系列问题,使操作系统的实现复杂化。操作系统按系统功能可分为批处理系统、分时系统、实时系统、网络操作系统、分布式操作系统和嵌入式操作系统。

最后,本章进一步讨论了操作系统是如何进行管理的。对于处理器的管理,需要掌握如何将 CPU 分配给需要的作业和进程;对于存储管理,需要解决用户程序怎么才能申请到内存,如何装入内存的指定区域,内存容量不够怎么扩充,程序执行完怎么释放空间等问题;而设备管理的主要任务是管理各类外部设备,包括分配、启动和故障处理等,合理地控制 CPU 与外设间的数据交换;文件管理则是为用户提供了一种简便、统一的存取和管理文件的方法,实现对文件的按名存取,解决文件在外存上的存储、共享、保密和保护问题,采用合理的分配策略来提高外存资源的利用率并提高文件的存取速度。

在学习本章时,要特别注意操作系统在计算机系统中的重要地位,掌握它与硬件和其他各类软件之间的关系,将操作系统这一比较抽象的概念与其他比较容易理解的知识放在一起进行学习,可能会事半功倍。

习题

1. 操作系统的主要功能是什么? 操作系统有哪些特征?
2. 什么是进程? 进程与程序有什么区别?
3. 进程的特征是什么? 进程有哪几种状态?
4. 简述处理器调度的主要调度算法。
5. 存储管理的主要功能都有哪些?
6. 什么是文件系统? 简述文件系统的主要功能。
7. 文件和文件夹都有哪些操作? 结合现有的操作系统,针对实际的文件和文件夹完成这些操作。

第4章

网络基础



计算机网络是计算机技术与通信技术相互渗透、不断发展的产物,计算机网络已经渗透到我们的工作与生活的各个方面,彻底改变了人们的工作方式、生活方式,乃至思维方式。本章介绍计算机网络的基本概念、计算机网络的发展、局域网、网络协议与分层结构、计算机网络体系结构、常用的网络测试工具等内容。

4.1 计算机网络概述

计算机网络是通信技术与计算机技术相结合的产物,一方面,通信技术中分组交换概念的提出为计算机网络的研究奠定了理论基础。另一方面,计算机技术应用到通信技术中,提高了通信网络的各种性能。计算机网络的发展经历了从简单到复杂、由单机与终端通信到成千上万台计算机之间的互连通信进而到物联网的迅猛发展过程。

4.1.1 计算机网络的定义

计算机网络是指地理上分散的多台具有独立功能的计算机通过通信设备及传输媒体互连起来,在功能完善的网络软件(网络协议、网络操作系统等)的支持下,实现计算机之间资源共享、信息交换或协同工作的系统。

计算机网络具有如下功能:

- (1) 信息交换:是计算机网络的最基本功能,可实现计算机间的信息传输。例如电子邮件、发布新闻、电子商务、远程教育、远程医疗等活动。
- (2) 资源共享:是计算机网络的最主要功能,资源指的是网络中所有的软件、硬件和数据。共享指的是网络中的用户能够部分或全部地使用这些资源。
- (3) 分布式处理:网络系统中若干台计算机可以互相协作共同完成一个任务。例如火车票、飞机票在多个地点进行预售。

4.1.2 计算机网络的产生与发展

计算机网络的发展始于20世纪50年代,经历了从简单到复杂、由单机与终端之间的远程通信到全球范围内成千上万台计算机之间互相连接、进而发展到物联网等的发展历程,实现了人与人、人与物、物与物之间的联接。计算机网络的发展大体经历了4个阶段。

1. 以数据通信为主的第一代计算机网络

第一代计算机网络的特点是整个系统中只有一台主机,多个远程终端通过通信线路和主机相连,如图 4-1 所示,其中 H 表示主机,T 表示终端。远程终端只有输入输出设备,没有独立的处理能力。面向终端的计算机网络具有明显的缺点:

- (1) 只有主机有处理能力且要承担通信工作,因而主机负荷较重。
- (2) 采用集中控制方式,可靠性低。
- (3) 每个终端要单独占用一条通信线路,通信线路利用率低。

从严格意义上说,该计算机网络还不是真正的计算机网络,但这样的通信系统已具备了网络的雏形。

第一代计算机网络的典型代表是 1954 年美国军方的半自动地面防空系统(Semi-Automatic Ground Environment, SAGE),将远程雷达和其他设施获得的信息通过通信线路与基地的一台计算机相连,进行集中的信息处理与控制。

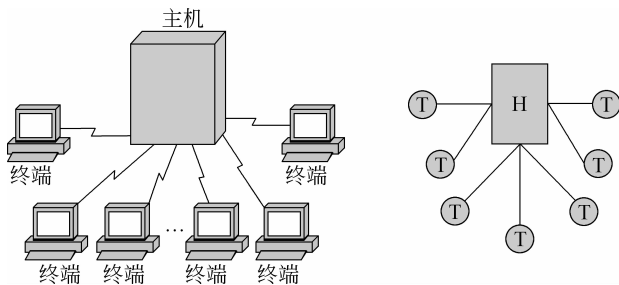


图 4-1 面向终端的计算机网络

2. 以资源共享为主的第二代计算机网络

随着计算机技术和通信技术的发展,计算机网络出现了多台主计算机之间通过通信线路互连起来。第二代计算机网络的特点是系统中有多台可以带有各自的终端的主机,这些主机之间通过通信线路相互连接,都具有自主处理能力,它们之间不存在主从关系。该阶段的网络系统,实现了计算机之间的相互通信和资源共享。这一阶段的标志性成果是 ARPANET 的成功运行。

1969 年,美国国防部高级研究计划局(Advanced Research Projects Agency, ARPA)建立了一个名为阿帕网(The Advanced Research Projects Agency Network, ARPANET)的网络,将分布在洛杉矶的加利福尼亚州大学洛杉矶分校、加州大学圣巴巴拉分校、斯坦福大学、犹他州大学 4 所大学的 4 台大型计算机通过租用的通信线路连接起来。该网络采用分组交换技术传送信息,保证了当网络的某一部分因遭受攻击而失去工作能力时,网络的其他部分仍能维持正常的通信工作。分组交换也称包交换,它是将用户传送的数据划分成多个更小的等长部分,每个部分叫作一个数据段。在每个数据段的前面加上一些必要的控制信息组成的首部,就构成了一个分组(Packet)。首部用以指明该分组的地址,然后网络中的交换机或路由器将源地址的分组接收后暂时存储在存储器中,再根据