一、计算机操作系统概述

1.1计算机系统

1.1.1计算机系统概述

电子数字计算机：一种能过够按照已经设定的程序进行数据处理的电子设备，是软件和硬件相结合的面向系统的、侧重于应用的自动化求解工具

计算机的发展：电子管计算机🡪晶体管计算机🡪集成电路计算机🡪大规模与超大规模集成电路计算机

计算机系统的组成：硬件+软件

软硬件层次：计算机硬件（内存、I/O接口、CPU🡪机器指令系统）🡪系统软件🡪支撑软件🡪应用软件🡪用户

虚拟机观点：裸机+操作系统就是虚拟机，操作系统作为扩展机器

资源管理观点：OS作为系统资源管理者 处理机、存储器、外设及信息

1.1.2硬件系统

硬件组成：中央处理器、主存储器、外围设备等

冯诺依曼模型：控制器、运算器、存储器、输入、输出

总线：（传输类型分类）控制总线、数据总线、地址总线

总线：（分级）内部总线、系统总线、通信总线

北桥（主存控制器）：连接CPU、主存储器、南桥控制器

南桥（I/O控制器）：通过各种协议连接输入输出设备和北桥控制器

中央处理器（CPU，运算核心和控制核心）：运算逻辑部件（ALU）+高速存储部件（通用、状态、控制寄存器以及cache）+控制部件（译码、控制）

存储器：（速度）cache（三级结构）🡪主存（内存）🡪硬盘🡪光盘

输入输出控制方式：轮询、中断、直接存储器访问

1.1.3软件系统

系统软件：操作系统、语言处理程序、实用程序、数据库管理系统、专用文件系统（其中OS和语言处理程序是最重要的）

支撑软件：（支撑软件也可以看作是系统软件的一部分）接口软件、工具软件、环境数据库，用来支持用户操作计算机的环境、提供软件开发工具

应用软件：用户按需求自身需求的专用程序

计算机硬件 🡪 机器指令

操作系统和实用程序 🡪 扩展机器指令

数据库管理系统 🡪 数据库子语言

语言处理程序 🡪 高级语言

支撑语言 🡪 程序员

1.2计算机操作系统

1.2.1计算机操作技术的发展

手工操作阶段🡪半自动化控制阶段🡪操作系统

1.2.2操作系统及其分类

操作系统：是管理系统资源、控制程序执行、改善人机界面，提供各种服务，合理组织计算机工作流程和为用户有效使用计算机提供良好运行环境的最基本的一种系统软件。

操作系统提供多道程序设计，隔离其他上层软件，并为其提供接口和服务，使得上层软件可以获得远较硬件所能提供的更强的功能

操作系统分类：（设计目标）批处理操作系统、分时操作系统、实时操作系统

（另外几种书中没有介绍）

核心理念：管理、控制、服务

核心目标：方便用户使用、扩大机器功能、管理系统资源、提高系统效率、构筑开放环境

批处理操作系统：服务于成为批（batch）的一系列作业，最早出现的操作系统

作业：把程序、数据联通作业说明书组织起来的任务单位（程序进入到处理器中）

分时操作系统（最常用的操作系统）：多个练级用户同时使用一个计算机系统，在各自的终端上进行交互式会话，在会话过程中提供程序、数据和命令，以问答的方式控制程序运行，系统把处理器的时间划分为时间片轮流分配给各个联机终端，若时间片用完则产生时钟中断，将控制权转至操作系统兵重新进行调度。

分时操作系统特点：分时、独立、及时、交互

批处理和分时操作系统的特点：追求目标不同、适用作业不同、资源利用率不同

实时操作系统：用于满足实时控制和实时信息处理的需要，可以支撑实时任务。包括不限于过程控制系统、信息查询系统、事务处理系统

实时任务：分为硬实时（在规定时间内完成）、软实时（在一个柔性时间段中尽早完成）

1.3深入观察操作系统

1.3.1资源管理的角度

计算机系统的资源：硬件资源（处理器、主存、外围设备等）+信息资源（数据、程序）

操作系统：为用户提供一种简单有效使用资源的方法，充分发挥各种资源利用率，为此对每种资源进行管理（在相互竞争的应用程序之间有序的控制软硬件资源的分配、使用、回收，使资源能够在多个程序之间共享）

研究内容：记录资源使用状态;使用或未使用，谁使用等

资源分配原则：确定如何分配（策略），何时分配

执行分配：根据分配的策略进行资源分配

资源回收：在某些用户，或程序不再需要资源时，系统应及时回收（修改记录），以便进行再分配。

资源管理技术：复用、虚拟（虚化）、抽象

复用：可以创建虚拟资源以解决物理资源数量不足的问题，包括时分复用共享和空分复用共享

空分复用共享：从“空间”上分割资源为更小的单位供进程使用

时分复用共享：从“时间”上分割资源为更小的单位，做个进程轮流使用资源一段时间（时间片轮转🡪若干个程序同时进行）

虚化：对资源进行转化、模拟、整合，把一个或多个物理资源变成一个或多个逻辑上的产物，物理资源是实际存在的，逻辑资源是虚拟的

抽象：通过创建软件来屏蔽硬件资源物理特性和接口细节，简化对硬件资源的操作、控制、使用的技术（不考虑物理细节而资源进行操作的技术），包括单级资源抽象和多级资源抽象

复用和虚拟的主要目标是解决物理资源数量不足的问题，抽象则用于处理系统复杂性，重点解决资源易用性

进程抽象🡪处理器+虚存抽象 进程是对进入主存执行的程序在处理器上操作的状态集的一种抽象

虚存抽象🡪主存+文件抽 虚存是对主存的一种抽象

文件抽象🡪设备 文件是设备的一种抽象

多层次抽象：磁盘🡪分区🡪扇区🡪簇🡪文件系统分区

1.3.2程序控制的角度

多道程序设计：允许多个程序同时进入计算机的主存，通过竞争处理器资源获得交替进行

多道程序设计特点：宏观🡪程序是并发的，均处于开始和结束之间 微观🡪程序是串行的，轮流占用CPU交替地执行

计算机系统运行特征：

1. I/O操作慢而CPU运算速度快，程序在执行I/O操作时并不占用CPU资源

（2）计算机系统配有多个硬件设备，因此多道程序设计使得多个程序在执行时可并行使用计算机的不同部件，从而充分大会计算机因不见的并行性，提高CPU的利用率以及外围设备的使用率，发挥计算机系统的处理能力

多道程序设计优点：

（1）提高CPU、主存、设备的利用率

（2）提高系统的吞吐率，使得单位时间内完成的作业数量增加

（3）充分发挥系统的并行性，使得设备与设备之间、CPU与设备之间均可并行工作

注意：并不是道数越多系统效率越高，这依赖于CPU与设备的并行度

为支持多道程序设计，计算机系统必须从软硬件综合设计的角度解决如下问题：

（1）怎样使用资源

（2）怎样复用CPU

（3）怎样使得CPU与I/O设备充分并行

（4）怎样让正在运行的程序让出CPU

为了支持多道程序设计，首先要为进入主存计算的程序建立管理实体—进程管理数据结构，之后在解决如下问题：

（1）处理器的管理与分配

（2）存储管理与存储保护

（3）其他资源的管理与调度

1.3.3操作控制计算机的角度

操作系统规定了合理操作计算机的工作流程，该流程由操作系统的操作接口—系统程序提供，为用户提供了操作控制计算机的所有服务。操作系统提供脱机作业控制方式（批处理系统的作业控制方式）和联机作业控制方式（当前最常见的作业控制接口）两种作业级接口

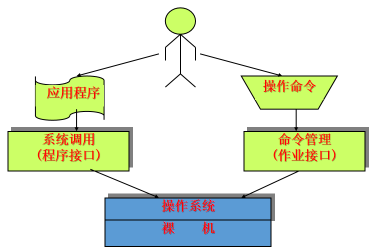
命令解释程序时接受和执行一条用户命令并对作业加工处理的程序包，无论以何种形态出现在用户面前，处理过程均类似（P.22太多了）

1.3.4人机交互的角度

操作系统的人机交互部分用于改善人机界面，为用户使用计算机提供良好的使用环境

人机交互设备：传统的终端设备、信心的模式识别设别

操作系统的用户界面：命令行界面；图形用户界面；触摸屏用户界面；语音用户界面

1.3.5系统接口的角度

操作系统的接口类型：命令行用户界面；图形用户界面；程序接口

命令行控制接口：

联机用户接口：用户不用编写作业控制说明书；用户使用系统提供的操作命令和系统会话；交互式地控制程序的执行和管理计算机系统

脱机用户接口：也称批处理用户接口，专为批处理作业的用户提供；操作命令：控制语句，用于编写作业控制说明书用户将作业的执行顺序和出错处理方法一并以命令文件方式提交给系统；系统按照命令文件中所规定的顺序控制作业执行执行过程中，用户无法干涉

Linux与Windows的命令控制接口：

命令控制界面：

传统的字符界面（CUI）：用户在系统给出的提示符下敲入特定的命令；系统接受用户输入的命令，并执行系统向用户报告执行结果，并等待用户输入新的命令；

图形用户界面（GUI）多窗口：命令被开发成用鼠标点击的菜单或者图标；Linux的图形化界面：xWindow

Linux shell

交互型命令解释程序、shell程序解释系统

Shell程序：带形参的批命令文件

Shell程序的组成：命令或其他Shell程序；位置参数；变量及特殊字符；表达式比较；控制流语句；函数

Windows的命令控制界面

Windows命令控制界面分两部分命令

解释程序cmd.exe：接受键盘输入的命令

窗口：通过鼠标或键盘进行操作

命令：Dos基本命令+windows自有命令；系统信息命令，如time，date，mem；系统操作命令，如shutdown,taskkill；文件系统命令，如copy, del, mkdir；网络通信命令，如ping, netstat, route

系统调用（访管指令）：现代操作系统内核提供一系列具有预定功能的服务条例。系统调用把应用程序的访问请求传送至内核，调用响应的服务历程完成所需处理，再将处理结果返回给应用程序

系统调用涉及的概念

计算机运行状态：

系统态：也称为管态或核心态，操作系统在系统态运行

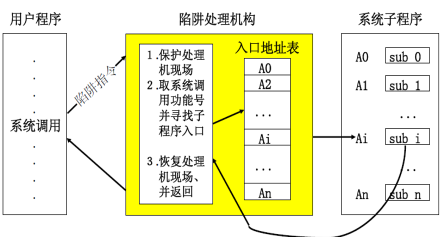
用户态：也称为目态，应用程序只能运行在用户态

处理机在系统态和用户态之间切换

特权指令：只允许操作系统使用的、在系统态运行的指令，对内存空间的访问范围不受限制；启动各种外部设备、设置系统时钟、关中断、清内存、修改存储器管理寄存器、执行停机、转换执行状态等。

非特权指令（访管指令）：由应用程序在用户态使用的命令。对内存的访问范围局限于用户空间

系统调用是用程序获得的操作系统服务的唯一途径，作用有两点：内核可以基于权限和规则对资源进行裁决，保证系统的安全性；系统调用对资源进行抽象，提供一致性接口，避免用户在使用资源时发生错误，且使变成效率大大提高

系统调用的实现过程

陷入（trap）机构：系统中为控制系统调用服务的机构

陷入指令：由于系统调用引起处理机中断的指令，又称为“访管指令：

（1）当用户使用系统调用时，产生一条相应的陷入指令，处理机在执行该指令时发生中断，并发信号给“陷入处理机构”，该机构启动相应处理程序去完成系统调用所要求的功能。

（2）每个系统调用都拥有一个功能号，在陷入指令中给出，有的陷入指令中还带有传递给系统调用处理程序的参数

（3）为实现系统调用，需要设计一张系统调用处理程序入口地址表，与系统调用功能号相对应

（4）陷入处理机构把陷入指令中的功能号与入口地址表中有关项对应起来，从而使用功能号驱动有关子程序的执行

系统调用实现的要点：

（1）编写系统调用服务例程

（2）设计系统调用入口地址表，每一个入口地址都指向一个系统调用的服务条伦理，有的包含系统调用自带参数的个数

（3）陷阱处理机制，需要开辟现场保护区，保存发生系统调用是应用程序的处理器现场

系统调用分类

进程控制：创建、中止、暂停、改变优先级等控制：

exit进程自我终止 fork创建进程 sleep进程睡眠

wait阳塞当前进程 getpid读父进程标识

进程通信：消息队列、共享存储区、socket等通信的建立、使用和删除：

存储管理：内存的申请和释放

线程管理：线程的创建、调度、执行、撤销等

系统管理：设置和读取时间、读取用户和主机标识等

gtime读取时间 stime设置时间 getuid读取用户标识

设备管理：请求和释放某个设备，启动设备操作，设备的读写和控制：

loctl设备配置 open设备打开 close设备关闭

read读设备 write写设备

文件管理：文件读写、创建删除和文件控制

creat文件创建 seek读写指针定位 open文件打开

close文件关闭 write写文件 read读文件

chmod修改文件属性 stat读文件状态 mount安装文件系统

传递参数三种方法：

（1）陷入指令自带参数：可以规定陷入指令之后的若干单元存放参数，叫做直接参数，或者在指令之后紧邻的单元存放参数的地址，叫做简介参数

（2）通过CPU的通用寄存器传递参数，不适用与大量参数，一般子啊主存的某个区域表中存放参数

（3）在主存中开辟专用堆栈区存放参数

系统调用与普通函数的区别（pdf130+P.27）

（1）运行在不同的系统状态

一般过程调用：调用过程和被调用过程均运行在用户态

系统调用：调用者运行在用户态，被调用者运行在系统态

（2）通过软中断进入

（3）返回问题

一般过程调用：返回调用过程中调用点之后

系统调用：系统调用执行结束后，执行CPU调度模块，重新选择进程执行

（4）嵌套问题

一般过程调用：对嵌套深度没有限制

系统调用：一般规定嵌套深度不超过6层

（5）调用形式不同

一般过程调用：按转向地址调用

系统调用：按功能号调用

（6）实现方式不同

（7）被调用代码位置不同

（8）提供方式不同

1.3.6系统结构的角度

软件设计呈现的特征（软件危机？）：

（1）复杂程度高

（2）开发周期长、成本高

（3）正确性难以保证

操作系统结构设计三层含义：

（1）研究操作系统的整体结构，同时考虑构造的过程和方法

（2）研究操作系统的局部构造，包括数据结构和控制结构

（3）研究操作系统的运行模型，即操作系统自身程序如何组织运行

结构设计方法

（1）单体式结构：把模块作为操作系统的基本单位，按照功能需要而不是程序和数据的特性，把整个系统分解为若干个模块，每个模块具有一定的独立功能，若干个关联模块协作完成某个功能

（2）将操作系统划分为若干个模块（进程）按照功能的调用次序将他们排列成若干层次各层之间只能存在单向以来或者单向调用关系，即底层为高层服务，反之则不能，这样的系统结构清晰且不构成循环调用

（3）物理计算机组员通过复用和共享技术可以改造成多个虚拟机，通过用一类物理设备来模拟另一类的物理设备，或通过分时地使用一类物理设备，把一个物理实体改造成若干个逻辑上的对应物

（4）操作系统金鸡奖所有应用必须的最小核心功能集放入内核，其他功能都在内核之外，由在用户态运行的服务进程实现，通过微内核提供的消息传递机制完成进程之间的消息通信

优点：

1）对进程的请求提供一致性接口，不区分内核和用户，所有服务借助消息传递机制完成

2）可扩充性好、易修改，增加性服务或替换老功能只需要增加或替换服务器

3）可移植性好，于CPU有关的代码均在内核中

4）对分布式系统提供有力支撑，当消息传递时，不必知道在哪台机器上

内核：一组中断驱动的程序模块，作为可信软件来提供支撑进程并发执行的基本功能和基本操作，在内核空间，运行于内核态

内核设计四大方法：单内核、微内核、混合内核、外内核

1.4 操作系统的作用与功能

对内是“管理员”和“魔术师“，对外是“服务员”

用户接口和服务提供者 作为扩展机或虚拟机 资源管理者和控制者

1.4.1处理机管理功能

作业和进程调度：后备队列上(外存空间)的谓度，作业调度(并不是所有类型机器都

具)。CPU调度，进程调度

进程通信;由于多个程序(进程)彼此间会发生相互制约关系，需要设置进程同步机制

进程之间往往需要交换信息，为此系统要提供通信机制

1.4.2存储器管理功能

内存分配：记录整个内存，按照某种策略实施分配，或回收释放的内存空间

地址映射：硬件支持下解决地址映射，即逻辑到物理地址转换

内存保护：保证各程序空间不受“进犯”

内存扩充：通过虚拟存储器技术虚拟成比实际内存大的多的空间来满足实际运行的需要

1.4.3设备管理功能

缓冲区管理：管理各类I/0设备的数据缓冲区，解决CPU和外设速度不匹配的矛盾

设备分配：根据UO请求和相应分配策略分配外部设备以及通道、控制器等，设备蜜动：实现用户提出的I/O操作请求，完成数据的输入输出。这个过程是系统建立和维持的

设备无关性：应用程序独立于实际的物理设备，由操作系统将逻辑设备映射到物理设备

1.4.4文件管理功能

文件存储空间的管理：包括：记录空闲空间、为新文件分配必要的外存空间，回收释放的文件空间，提高外存的利用率等

目录管理：目录文件的组织、及实现用户对文件的“按名存取”、目录的快速查询和文件共享等

文件的读写管理和存取控制：根据用户请求，读取或写入外存。并防止未授权用户的存取或破坏，对各文件(包括目染文件)进行存取控制

1.4.5用户接口

命令界面：系统提供一套命令，每个命令都由系统的命令解释程序所接收、分析，然后调用相应模块完成命令所需求功能

图形界面：考虑用户使用计算机的方便性，现代操作系统都提供了图形用户界面。它也是一种交互形式，只不过将命令形式改成了图形提示和鼠标点击

程序界面：也称系统调用界面，是程序级上用户与操作系统打交道的方式。

1.5 操作系统的主要特性

并发性 共享性 异步性

1.5.1操作系统中的并发性

并发性：指两个或两个以上的事件或活动在同一时间间隔内发生

实质：一个物理CPU(也可以多个物理CPU)在若千道程序之间多路复用

作用：发挥并发性能够消除系统中部件和部件之间的相互等待，有效地改善来统资源的利用率，改进案统的吞吐率，提高案统效率采用并发技术的条统称多任务系统

并发性使系统变得复杂化，带来问题：

（1）如何从一个活动切换到另一个活动

（2）急样将各个活动隔离开来，使之互不干扰，免遭对方破坏

（3）怎样让多个活动协作完成任务

（4）怎样协调事个活动对资源的竞争

（5）如何保证每个活动的资源不该其它进程侵犯

（6）多个活动共享文件数据时，如何保证数据的一致性

并行性：两个或两个以上的事件或活动在同一时刻发生。并行的事件或活动一定是并发的，但反之并发的事件或活动未必是并行的。

并行性是并发性的特例，而并发性是并行性的扩展。

1.5.2操作系统中的共享性

共享指操作案统中的资源可被多个并发执行的进程所使用

透明资源共享：资源隔离与接权访

显式资源共享：临界资源与独占访问

与共享性有关的问题：资源分配、信息保护、存取控制等

1.5.3操作系统中的异步性

随机现象：进程何时执行？何时暂停？怎样的速度向前推进？都是异步(随机)的

原因：

（1）作业到达系统的类型和时间是随机的

（2）操作员发出命令或按按钮的时刻是随机的

（3）程序运行发生播误或异常的时刻是随机的

（4）各种各样硬件和软件中断事件发生的时刻是随机的

并步性给系统带来潜在危险，有可能导致与时间有关的错误

操作系统的一个重要任务是必须确保捕捉任何一种随机事件，正确处理可能发生的随机事，正确处理任何一种产生的事件序列，否则将会导致严重后果