

[参考资料]

课本&课件

《操作系统教程》 陈向群、杨芙清编著 北京大学出版社

《操作系统——内核与设计原理》 William Stallings 著 电子工业出版社

《操作系统教程》 孙忠秀、谭耀铭、费翔林、谢立、衣文国著 高等教育出版社

<http://baike.baidu.com>

[注] 黑色字体为参照参考资料给出的正式定义,灰色字体为自己的一些理解,仅供参考。
另外,带有英文注解的为课件上标红的部分。

如果大家发现哪个地方有什么问题,欢迎指正~最后希望大家不要着急,慢慢理解概念,一定可以把操作系统学好的。

Wrote by: 潘玲

Revised by: 詹雪莹

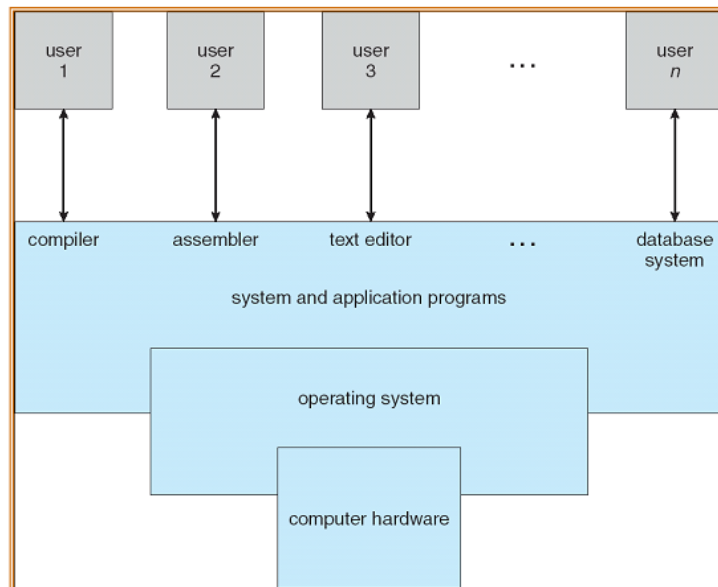
第一章 介绍

1、操作系统

操作系统是管理和控制计算机硬件与软件资源的计算机程序,是直接运行在“裸机”上的最基本的系统软件,任何其他软件都必须在操作系统的支持下才能运行。

操作系统是用户和计算机的接口,同时也是计算机硬件和其他软件的接口。它能**有效地**组织和管理计算机系统中的硬件及软件资源;**合理地**组织计算机工作流程,控制程序的执行;并向用户提供各种服务功能,使得用户能够灵活、**方便**、有效地使用计算机,并且使得整个计算机系统能高效地运行。

其中,“有效”指的是操作系统在管理资源方面要尽可能地提高处理器的利用率;“合理”主要是指操作系统对于不同用户程序要公平,以保证系统不发生死锁和饥饿现象;“方便”主要是指人机交互方面。



用一句话来说,其实就是操作系统是用户和计算机硬件的接口/中介(intermediary),使得计算机硬件能够被有效(efficient)地利用,用户程序能够被合理地执行,并且能够为用户提供方便(convenient)的操作界面。它能合理分配资源、避免冲突,是资源的分配器,同时也是控制程序以避免错误或不当操作。

2、内核(kernel)

内核是操作系统中最基本的部分,它是为众多应用程序提供对计算机硬件的安全访问的一部分软件,并且决定一个程序在什么时候对某部分硬件操作多长时间。它从始至终一直在

电脑上运行，其中“始”表示它是最先开始运行的程序，开机的时候就会运行；“终”表示它是最后退出的程序，关机的时候才退出。

3、并发性(concurrency) [课外了解]

并发性是指在计算机系统中同时存在多个程序。并发的实质是一个物理 CPU（也可以多个物理 CPU）在若干道程序之间多路复用，并发性是对有限物理资源强制行驶多用户共享以提高效率。并发性是关于软件过程分解成进程、线程并处理相关的效率、原子性、同步和调度问题。实现的关键之一是如何对系统内的多个活动（进程）进行切换。

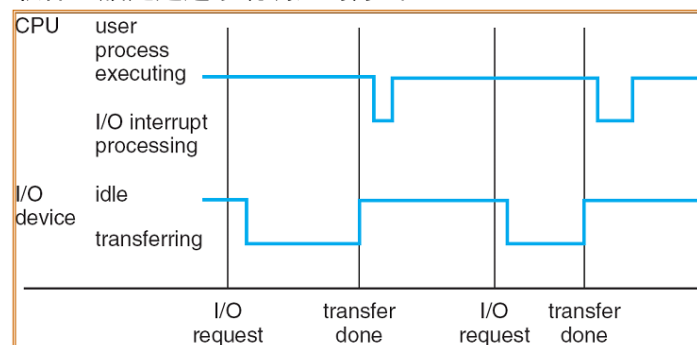
[注] 并发性与并行性是不一样的。并行性指的是两个或两个以上的事件或活动在同一时刻发生。而并发性是指两个或者多个时间在同一时间间隔内发生，带有一定的统计含义，是一个宏观概念。

p. s. 在单 CPU 环境下，本质是宏观上并行，微观上串行；在多处理器系统中，不仅在宏观上是并发的，在微观上（处理器这一级上）也是并发的，这在分布式系统中能更充分地体现。

4、中断(interrupt)

中断指当出现需要时，CPU 暂时停止当前程序的执行转而执行处理新情况的程序和执行过程。即在程序运行过程中，系统出现了一个必须由 CPU 立即处理的情况，此时，CPU 暂时中止程序的执行转而处理这个新的情况的过程称为中断。

中断的来源既可以是硬件也可以是软件，例如硬件一般是通过系统总线向 CPU 发送信号来触发一个中断，软件一般是通过系统调用的方式。



其工作流程为当 CPU 被中断时，处理器在响应中断前结束当前指令的执行。然后为从中断程序中转移控制做准备（保存从中断点回复当前程序所需要的信息），然后，把响应这个中断的中断处理程序入口地址装入程序计数器。接下来即可处理中断。处理完成后，恢复前面被中断的程序的执行。

5、直接存储器存取/直接内存存取(DMA, direct memory access)

1) 动机

尽管中断的引入大大提高了处理器处理 I/O 的效率，但是当处理器和 I/O 之间传送数据的时候，处理器还要做很多工作，效率仍然不高。即中断处理方式的缺点是每传送一个字符都要进行中断，其中中断控制器，还要保留和恢复现场以便能够继续原程序的执行，花费的工作量很大，这样如果需要大量的数据交换，系统的性能会很低。

2) 解决方法：DMA 技术

① 特点

不采用软件而是采用一个专门的控制器来控制内存与外设之间的数据交流，无须 CPU 介入，大大提高 CPU 的工作效率

② 工作过程

在进行 DMA 数据传送之前，DMA 控制器会向 CPU 申请总线控制权，CPU 如果允许，则将控制权交出。因此，在数据交换时，总线控制权由 DMA 控制器掌握，在传输结束后，DMA 控制器将总线控制权交还给 CPU

③ 意义

对于大量数据的 I/O 传送来说，DMA 技术是很有价值的

6、高速缓存(cache)

高速缓冲存储器是存在于主存与 CPU 寄存器之间的一级存储器，由静态存储芯片 SRAM 组成，容量比较小但速度比主存高得多，接近于 CPU 寄存器的速度。在计算机存储系统的层次结构中，是介于 CPU 寄存器和主存储器之间的高速小容量存储器。它和主存储器一起构成一级的存储器。其中，最重要的技术指标为其命中率。**一般意义上，每一个速度高一级的存储器都自动作为低一级存储器的高速缓存。**但是可能会带来高速缓存一致性(cache coheency)的问题。

7、多道程序设计(multiprogramming) [课外了解]

为了提高计算机系统中各种资源的利用效率，缩短作业的周转时间，常常利用多道程序设计使得多种硬件资源能够并行工作。让处理器和 I/O 设备同时保持忙状态，根本目的是为了**提高整个系统的效率。**

多道程序设计指的是允许多个程序同时进入一个计算机系统的主存储器并启动进行计算的方法。也就是说，计算机内存中可以同时存放多道（两个以上相互独立的）程序，它们都处于开始和结束之间。从宏观上看是并行的，多道程序都处于运行中，并且都没有运行结束；从微观上看是串行的，各道程序轮流使用 CPU，交替执行。引入多道程序设计技术的根本目的是为了**提高 CPU 的利用率，充分发挥计算机系统部件的并行性，现代计算机系统都采用了多道程序设计技术。**

• 小结：

问题 - 相对于处理器，I / O设备的速度太慢，使得处理器经常是空闲的

解决方案 - 多道程序设计multiprogramming

实现方法 - 当一个作业需要等待I / O时，处理器可以切换到另一个可能并不在等待I/O的作业

利用技术 - I/O中断驱动和DMA

好处 - 资源利用率提高

特点 - 允许处理器同时处理多个作业；宏观上并行，微观上串行

8、分时技术(timesharing) [课外了解]

使一台计算机采用片轮转的方式同时为几个、几十个甚至几百个用户服务的一种操作系统。分时操作系统将系统处理机时间与内存空间按一定的时间间隔，轮流地切换给各终端用户的程序使用。由于时间间隔很短，每个用户的**感觉就像他独占计算机一样。**

分时操作系统的特点是**可有效增加资源的使用率。**

| | 多道程序设计 | 分时 |
|---------|-----------|-------------|
| 主要目标 | 最有效地使用处理器 | 有效使用&减小响应时间 |
| 操作系统指令源 | 作业提供的命令 | 终端键入的命令 |

第二章 操作系统结构

1、系统调用

由于操作系统的特殊性，应用程序不能采用一般的过程调用方式来调用这些功能过程，而是需要利用一种系统调用命令去调用所需的操作系统过程。

所谓系统调用，就是用户在程序中调用操作系统所提供的一些子功能，这是一种特殊的过程调用，通常是由特殊的及其指令实现的。

它是操作系统提供给变成人员的唯一接口。编程人员利用系统调用，动态请求和释放系统资源，调用系统中已有的系统功能来完成一些工作。

因此，系统调用就像一个黑盒子那样，对用户屏蔽了操作系统的具体作用而只提供有关的功能。

从本质上来，系统调用是应用程序请求操作系统核心完成某一特定功能的一种过程调用。

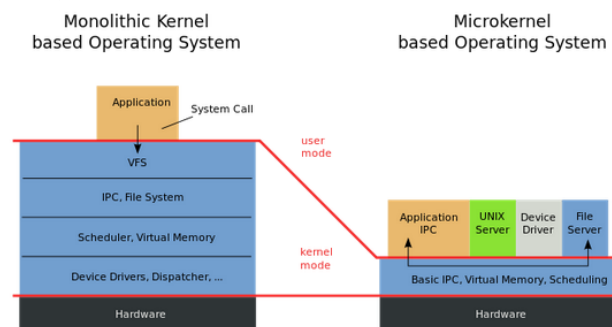
2、API

由操作系统实现提供的所有系统调用所构成的集合即为 API（应用编程接口，Application Programming Interface）。

[对比] 程序员调用的是 API，然后通过系统调用共同完成函数的功能。因此，API 是一个提供给应用程序的接口，一组函数，是与程序员进行直接交互的，它侧重于向上层提供给定的服务。而系统调用则不予程序员进行直接交互，它根据 API 函数，通过一个软中断机制向内核提交请求，以获取内核服务的接口，它侧重于向下层的内核发出一个明确的请求。

系统给调用是向下的，和硬件打交道（在内核模式下）；API 是向上的，为应用程序提供服务（在用户模式下）。程序访问通常是通过高层次的 API 接口而不是直接进行系统调用，把系统调用封装到一个库里面，然后调用这个库里面的函数，从而间接进行系统调用。（如果 API 需要和内核打交道就需要系统调用，否则不需要。）

3、微内核(microkernel)



微内核是一种能够提供必要服务的操作系统内核，其中这些必要的服务包括任务、线程、交互进程通信、内存管理等。由于它是提供操作系统核心功能的内核的精简版本，具有很好的拓展性，也增加了移植性。

其实微内核就是指保留了操作系统的核心模块，而把所有传统内核中的非核心部分移到 user mode 中。好处是显然的，使得系统易于拓展，增强了可靠性与安全性，不过在用户控件和内核控件的通信方面带来了一些影响。

优点主要包括以下几点：

- a) 增加一个新的服务不需要修改内核
- b) 在用户模式中比在内核模式中更安全、更易操作
- c) 增加可靠性

缺点：用户程序和系统服务通过使用进程间的通信机制在微内核中相互作用，例如发送消息。这些消息由操作系统运送。微内核最主要的缺点是与进程间通信的过度联系，以及为了保证用户程序和系统服务相互作用而频繁使用操作系统。

4、虚拟机(virtual machine)

虚拟机是指通过软件模拟的具有完整硬件系统功能的、运行在一个完全隔离环境中的完整计算机系统。进入虚拟系统后，所有操作都是在这个全新的独立的虚拟系统里面进行，不会对真正的系统产生任何影响，使得在同一台 PC 上可以运行多个操作系统。