操作系统复习提纲

第一章

1. 计算机系统从下到上划分为四个层次：硬件、操作系统、实用程序、应用程序
2. 外层的软件必须以事先约定的方式调用内层软件或硬件提供的服务。称这种约定为接口
3. OS是由一系列程序模块组成的一个大程序。对计算机的软、硬件资源进行统一的管理和调度，合理地组织计算机的工作流程，以提高资源的利用率。
4. 操作系统的设计目标：方便性（方便用户使用）&有效性（提高系统资源利用率）。（此外便于设计、实现和维护）
5. 顺序处理、简单的批处理系统、多道成批处理系统、分时系统、实时系统、嵌入式系统
6. 顺序处理无操作系统。
7. 简单的批处理系统有一个监控程序软件常驻内存。一次装入一个作业
8. 多道成批处理系统（以主存为中心）（通道、中断、缓冲技术——全部和输入输出有关）
9. 多道程序设计技术：是指在内存同时存放若干道程序，使它们在系统中交叉运行，共享系统中的各种资源。当一道程序暂停执行时，CPU立即转去执行另一道程序。[特点]：多道、宏观上并行（不同的作业分别在CPU和外设上执行）、微观上串行（在单CPU上交叉运行）。
10. 引入多道程序设计技术的根本目的是**提高CPU的利用率**，**充分发挥系统设备的并行性**。这包括程序之间、CPU与设备之间、设备与设备之间的并行操作
11. **资源利用率**：指在给定时间内，系统中某一资源（如CPU、存储器、外部设备等）**实际使用时间**所占比率。

**吞吐量**：指单位时间内系统所处理的信息量。它通常是用每小时或每天所处理的作业个数来度量。

**周转时间**：指从作业进入系统到作业退出系统所用的时间。而平均周转时间是指系统运行的几个作业周转时间的平均值。

**周转时间**：作业进入系统到作业退出系统所经历的时间。

**PPT OS1 P25例子。**

1. 资源利用率：指在给定时间内，系统中某一资源（如CPU、存储器、外部设备等）实际使用时间所占比率。
2. 吞吐量(Throughput)：指单位时间内系统所处理的信息量。它通常是用每小时或每天所处理的作业个数来度量。
3. 周转时间：指从作业进入系统到作业退出系统所用的时间。而平均周转时间是指系统运行的几个作业周转时间的平均值。
4. 批处理系统特点：

优点: **系统吞吐量大，资源利用率高**。适合计算量大、自动化程度高的成熟作业。

缺点: **不能直接控制作业运行，作业的周转时间太长**（I/O时CPU处于等待状态）

1. 多道程序设计技术的实现（实现要解决的问题）：

存贮分配和存贮保护

处理机的管理和调度

系统其它资源的管理和调度

1. 分时系统：一台主机连接有若干个终端。用户交互式地向系统提出命令请求，系统接受命令，采用时间片轮转方式处理请求，并在终端上显示结果。如在大型数据库上的查询。
2. 分时系统的特点：

**同时性：**若干用户同时使用一台计算机。

**独立性：**每个用户占有一台终端，彼此独立操作，互不干扰。

**交互性：**用户可通过终端与系统进行人机对话。

**及时性：**用户的请求能在较短时间内得到响应。通常以用户能够接受的等待时间来确定（2～3s）。

1. 批处理与分时系统的差别：

**批处理系统：**

目标是**提高系统资源的利用率**。

适用于比较成熟的大型作业。

可在后台执行。不需要用户频繁干预。

**分时系统：**

目标是**对用户请求的快速响应**。

适用于短小作业。

终端键入命令。

1. 实时系统：

当对处理机操作或数据流动有严格时间要求时，就需要使用实时系统。

实时过程控制：工业生产中的自动控制，飞机导航、导弹发射等。

实时信息处理：民航机票的预订、查询，银行系统的借贷，情报信息检索等系统。

1. 硬实时：系统的所有可能的延迟是一定的。对于关键的任务必须在指定时间范围内完成。软实时：即使任务没有在规定时间内完成，也还是允许的。
2. 实时系统特点：

（1）**实时性**。计算机对随机发生的外部事件能够及时地响应和处理。

*及时：是指响应所具有的速度，足以控制发出实时信号的那个设备。*

（2）**可靠性**。要具有容错能力*，可采用双工机制：一台主机；一台后备机。*

（3）**确定性**。是指系统按照固定的、预先确定的时间执行指定的操作。*其可确定性取决于系统响应中断的速度和处理能力。*

1. 实时系统与分时系统的区别：

**实时系统**

（1）实时性和可靠性较高。

（2）交互能力较差。

**分时系统**

（1）以用户的容忍程度为依据，对实时性没有要求。

（2）允许出错，可重复运算。

1. 嵌入式系统：（一般都是实时的）

软件要求固化存储。

通常是一个多任务可抢占式的实时操作系统核心，只有满足实际需要的有限功能，如任务调度同步与通信、主存管理、时钟管理等。

1. 操作系统的三种基本类型：批处理系统；分时系统；实时系统。

通用操作系统：一个系统兼有批处理、分时和实时处理三者或其中两者的功能。

1. 操作系统的基本功能：

(1)**处理机管理：**进程管理。处理机如何调度的问题：FCFS、优先级、时间片轮转？

(2)**存储器管理：**主存管理。存储分配、存储保护、主存扩充。

(3)**设备管理：**涉及对系统中各种输入、输出设备的管理和控制。分配设备，控制设备传输数据。

(4)**文件管理：**将程序、数据、操作系统软件等组织成文件，存放在磁盘或磁带上，方便用户访问。

1. 为了管理系统资源，操作系统必须掌握系统资源的当前状态信息。

进程----进程表

存储器----存储表

I/O设备----I/O设备表

文件----文件表

1. 操作系统的特性：

(1)**并发性：**并发是指系统中存在着若干个逻辑上相互独立的程序，它们都已被启动执行，都还没有执行完，并竞争各种资源。

(2)**共享性：**是指系统中的资源可供内存中多个并发执行的进程共同使用。如打印机（互斥共享）、磁带机、磁盘等（同时共享）。（支持系统并发性的物质基础是资源共享）

(3)**虚拟性：**把共享资源的一个物理实体变为若干个逻辑上的对应物。如，CPU的分时共享；虚拟存储器技术。

(4)**异步性（随机性）：**有限的共享资源使并发进程之间产生相互制约关系。各个进程何时执行、何时暂停、以怎样的速度向前推进、什么时候完成等都是不可预知的。

1. 操作系统性能评价：

（1）系统效率：系统效率是操作系统的一个重要性能指标。体现系统效率的指标包括资源利用率、吞吐量和周转时间以及响应时间等。

（2）系统可靠性：是指系统能发现、诊断和恢复硬件和软件故障的能力。

（3）可移植性(Portability):是指把一个操作系统从一个硬件环境移植到另一个硬件环境时系统仍能正常工作的能力。移植时，代码修改量要小，效率要高。

1. 多处理机操作系统：

多处理机采用紧耦合方式进行连接，共享主存

非对称多处理(ASMP)：主处理机运行操作系统，其他处理机运行用户作业，主处理机为其他处理机分配和调度任务，主从模式。

什么是对称多处理？它有什么好处？

**对称多处理(SMP)：**操作系统和用户程序可安排在任何一个处理机上运行，各处理机共享主存和各种I/O设备。**增加了系统的吞吐率；增加了系统的可靠性。**

1. \*网络操作系统：网络中的各台计算机都配有各自独立的操作系统，网络操作系统把它们联系起来，并为它们提供通信和网络资源共享。

客户/服务器（Client/Server)模式：服务器是一个瓶颈。

对等模式（Peer-to-peer)：系统内的节点机（nodes）是对等的，既可作为客户机，又可作为服务器。在网络中既无服务处理中心，也无控制中心。

1. \*分布式系统：是由多个分散的计算机通过网络连接而成的一个统一的计算机系统。可以获得极高的运算能力和广泛的数据共享。
2. \*智能卡系统：智能卡实际上是一台单片机系统。有非常严格的运行能耗和存储空间的限制（从几KB到几百KB)。
3. 用户与操作系统接口：

**操作接口：**命令语言或窗口界面是用户使用计算机系统的主要接口。

**编程接口：**系统调用是用户与操作系统之间的编程接口。

1. 核心态/用户态（为了实现系统保护）

在计算机系统中存在两类性质不同的程序：**操作系统内核程序**；**用户自编程序和系统外层的应用程序**。

前者是后者的管理者和控制者，所以操作系统内核应该享有某些特权。为此将CPU的运行状态分为**核心态（管态）**和**用户态（目态）**。

1. 各种状态下分别执行什么程序？什么时候发生状态转换？状态转换由谁实现的？：

**在核心态下**，允许执行处理机的全部指令集，访问所有的寄存器和存储区；

**在用户态下**，只允许执行处理机的非特权指令，访问指定的寄存器和存储区。

**用户态到核心态**的转换由硬件完成；**管态到目态**的转换由操作系统程序执行后完成。

1. 什么是系统调用？什么是特权指令？特权指令执行时，CPU处于哪种工作状态？

**系统调用**：操作系统内提供的一些子程序。提供了用户程序与操作系统内核的接口。用户通过系统调用接口，向操作系统提出资源请求或获得系统服务。

**特权指令**：关系系统全局的指令。执行时处于核心态。

1. 通过中断和异常，CPU能从用户程序的运行转入操作系统内核程序的运行。

中断：异步事件。也称外中断，指来自CPU正在执行指令以外的事件。如I/O结束中断，时钟中断、硬件故障。不同中断有不同优先级。

异常：同步事件。异常，也称内中断、例外或陷入(trap)，指来自CPU正在执行指令的内部事件。异常不能被屏蔽，一旦出现应立即处理。

1. 处理机优先级：指处理机当前正运行程序的中断响应级别。当处理机处于某一优先级时，只响应比该优先级高的中断，而屏蔽优先级低的中断。

第三章

1. P操作顺序很重要，一定要能申请再给控制权，否则会死锁
2. Pv操作太可怕
3. 管程看一下（记住一致性，感觉代码不会考）

第四章

P122 为什么会有6次？大概就是六个都没有都要调进来？

为什么说段式是二维，页式（页表）是一维

缺页处理过程简述

1．查页表的状态位。

2．状态位为0，缺页中断。

3．操作系统处理缺页中断，寻找一个空闲的内存页。

4．若有空闲页，则把从磁盘读入信息装入该页面。

5．若无空闲页，则按某种算法选择一个已在内存的页面，暂时调出内存。若修改过还要写磁盘。调入需要的页。之后要修改相应的页表和内存分配表。

6．恢复现场，重新执行被中断的指令。

P155结合书看下

第5章

文件分配表怎么就快了？