一、概论

·批处理&分时

批处理是单用户多程序；

分时是多用户多程序。

·单道&多道程序

单道程序意味着内存只允许一个程序存在和运行；

多道程序意味着内存允许多个程序同时存在。

·实时&网络&分布式操作系统

实时的特征是高响应速度和高可靠性；

网络分为服务器（汇总、发送）和客户端（接收）；

分布式共享资源、互相通信、彼此对等。

·冯洛伊曼体系结构

存储程序式：集中顺序过程控制。

·四个操作系统功能

处理机分配（进程、资源管理）、存储器管理、设备管理、文件管理。

·四个现代操作系统特征

并发、资源共享、虚拟化存储、处理不确定事件。

·操作系统工作模式

内核态（CPU可以访问包括外围设备的所有资源）、用户态（只能受限地访问内存）。

相关概念：

·特权指令/保护指令：具有特殊权限的指令，只用于操作系统或其他系统软件，一般不提供给用户使用。

·异常：中断（Interruption外中断）源自CPU外的事件、陷入（Trap、Exception内中断）源自CPU指令的事件。切换到内核态，检查特殊寄存器进行对应操作。

·分析操作系统的观点

用户观点（提供哪些功能）、资源管理（由哪些部分构成）、进程观点（如何进行调度、沟通各资源）、模块分层（操作系统的构建）。

·微内核&整体内核

微内核只保留操作系统核心功能。规模小、可移植性好

整体内核把操作系统的功能分为各个模块，模块间任意调用。效率高、难以扩展和升级。

二、操作系统接口

·系统调用：

系统调用是在操作系统核心设置的一组用于实现各种系统功能的子程序。应用程序有特殊请求时，把请求传给内核，调用相应的内核函数。

·系统调用&普通调用

系统调用在用户态被调用，运行时会切换到内核态；普通调用不会。

系统调用靠异常陷入内核态，进而跳转到对应内核代码；普通调用直接跳转。

系统调用返回时可能重新调度；普通调用直接返回。

三、进程&并发

1、进程&线程

·并发（多个任务分时运行）&并行（多个任务在多个处理机上同时运行）

·进程&程序

程序是指令集；

进程是程序的一次执行，是操作系统进行资源分配调度的单位。

进程具有动态性、并发性、独立性、异步性。

进程=进程控制块+程序+数据。

·进程控制块PCB

一个与动态过程相联系的数据结构，记载了进程的外部特性（名字、状态）以及与其他进程的联系（通信关系），还记录了进程所拥有的各种资源。

进程控制块是进程存在的标志。

组织方式：线性表、链接方式、索引方式。

·进程的三个基本状态

执行状态（进程已获得除处理机外的所需资源，等待分配）；

就绪状态（占用处理机资源）；

阻塞状态（正在执行的进程由于某种事件的发生放弃处理机）。

执行（时间片用完）到就绪；就绪（被调度）到时间片；阻塞（事件发生）到就绪；执行（等待某个事件）到阻塞。

·进程的控制原语

创建（fork、exec）

撤销（kill）

·原语&系统调用

原语：由若干条指令组成的指令序列，来实现某个特定的操作功能。连续不可分割、是操作系统核心组成部分、必须在内核态执行且常驻内存。

系统调用由用户态调用，运行时转到内核态。

·进程&线程

进程是操作系统资源分配调度的单元，包含资源和可执行单元；

可执行单元称为线程。线程是进程内部的活动的并发单元，是轻量级（易创建、撤销）的进程。

目的：线程减小进程切换的开销；提高进程内的并发程度；共享资源。

2、进程调度（计算题）

·概念

从就绪队列中按照一定算法选择进程分配处理机，以实现进程并发。

·高级调度、中级调度、低级调度

对作业进行调度；

对存储器资源调度（外存到内存、内存到外存）；

对CPU资源调度。

·非抢占式、抢占式

进程调用原语、时间片用完、I/O中断等原因进入阻塞才重新调度；

就绪队列一旦有优先级高于当前运行进程的，就进行调度。

·何时调度

创建新进程、进程运行完毕、进程被阻塞（调用原语、时间片用完）、I/O中断、时钟中断（分时系统）

·衡量标准

周转时间（作业从提交到完成所经过的时间）、平均周转时间、带权平均周转时间

吞吐量：单位时间内完成的作业数。

响应时间：进程数\*时间片大小。

响应比：（执行+等待）/执行。

CPU利用率：总用时-CPU空转/总用时。

·调度方法

FCFS（先来先服务）有利于长作业和CPU繁忙作业；不利于短作业和I/O繁忙的作业。

SJF（短作业优先）减少平均周转时间、提高吞吐量；不利于长作业、无优先级、可能饥饿。

Round Robin（时间片轮转）提高进程并发性和响应时间。

优先级（静态优先级不变、动态优先级随时间改变）

3、进程通信

·临界资源&临界区

临界资源是同一时刻只允许一个进程使用的资源；

临界区是由于共享临界资源，必须互斥执行的程序段。

·同步&互斥（解答题）

同步：多个相关进程在执行次序上的协调。

互斥：多道程序环境中，每次只允许一个进程对临界资源进行访问。

同步、互斥的原则：空闲让进、忙则等待、有限等待、让权等待。

经典的同步、互斥问题：

生产者-消费者（同步、互斥）

读者-写者（互斥）

哲学家就餐（互斥）

理发师（同步、互斥）

·信号量&P、V操作

信号量是只能由P、V操作修改的数据结构

S=0表示？？？？？

·管程

4、死锁

·死锁&活锁&饥饿

·四个产生死锁的必要条件

互斥、不可剥夺、请求和保持、环路等待。

·进程-资源图/资源分配图

圆进程、方资源、有向边意义。

·处理死锁的方法

允许发生：无作为；检测（化简进程-资源图）并解除（资源剥夺、撤销进程）；

不允许发生：预防（静态地）破坏四个条件、避免（动态）安全性、银行家算法。

四、存储管理

1、存储管理的功能

内存的分配与回收、存储保护、地址转换、（静态、动态）重定位、存储共享、扩充内存容量。

2、分区存储管理

·固定分区

基本思想

优缺点

·可变分区

基本思想

优缺点

算法

BestFit

WorstFit

FirstFit

NextFit

·覆盖与交换

3、页式存储管理（计算题）

·页面&页框

把每个作业的地址空间分成大小相等的片，即一个个页面。

把主存的地址空间分成和页面大小相等的片，即一个个页框。

·MMU

内存管理单元，负责虚拟地址、物理地址的映射，提供硬件机制和的内存访问授权

·页表

·快表

存在于地址变换机构中的一个由高速寄存器组成的小容量的相联寄存器，构成的一张表。

·页面大小

一个页对应的数据大小。

4、段式存储管理

·基本思想

·地址变换

·分页&分段

·段页式

结合分段和分页的原理，先将用户程序分为若干段，每段再分为若干页。为了实现逻辑地址到物理地址的转换，需同时配置段表和页表。段表内容包括页表起始地址和页表长度。

5、虚拟存储原理&算法（计算题）

·内存访问的局部性原理

·页面置换策略

FIFO

LRU

LFU

第二次机会

工作集

Clock

·缺页率

五、输入、输出系统

1、I/O

·I/O设备

字符设备：提供连续的数据流，应用程序可以顺序读取，通常不支持随机存取；

（键盘打印机）

块设备：应用程序可以随机访问设备数据，程序可自行确定读取数据的位置；

（磁盘）

网络设备：面向数据包的接收和发送。并不对应文件系统的结点，而是由系统分配一个唯一的名字。

·I/O控制技术

程序控制（轮询）测试设备闲忙，不忙则执行输入输出，否则不断直到不忙。

中断驱动：缓冲区满时触发中断，在中断处理时由CPU控制完成数据传送，在每个数据传送完成时中断CPU。

DMA（直接内存访问）以数据块为单位，由DMA控制器完成数据传送，在一批数据全部传送结束时中断CPU。

Channel（通道技术）CPU把数据传输功能下放给通道，通道与CPU分时使用内存，实现CPU和外设的并行工作。

2、缓冲区

·概念

为了匹配外设和CPU的处理速度，减少终端次数和CPU的中断处理时间，使用专用硬件缓冲区或在内存中划一个区域，用来暂时存放输入/输出数据的地方。

·类别

单缓冲区

双缓冲区

环形缓冲区

缓冲池

3、I/O管理软件

·流程

·设备无关性

程序可以通过一组统一的操作来操作设备，这种操作的接口与具体设备无关。

六、文件系统

·文件系统

操作系统中与文件管理有关的那部分软件和被管理的文件以及实施管理所需要的一些数据结构的总体。

·文件

具有符号名（文件名）的数据项集合，包括文件体和文件说明。

文件名是文件的标识符。

·目录

由文件说明索引组成的用于文件检索的特殊文件，内容主要是文件访问和控制信息。

·文件控制块

操作系统为管理文件而设置的数据结构，存放了为管理文件所需的所有有关信息（文件名、文件类型、物理地址、文件大小、最近访问时间、最近修改时间、文件主标识、访问权限等）。文件控制块是文件存在的标志。

·文件的逻辑结构

记录式（有逻辑结构）有记录

流式（无逻辑结构）数据按顺序记录保存

·文件的物理结构

连续、串联、索引（一级索引、多级索引）。

·磁盘空间的管理

空闲表、空闲链表、位示图、成组链接。

七、磁盘管理（计算题）

·磁盘访问时间=寻道时间+旋转延迟时间+传输时间

·磁盘调度算法

SSTF（最短寻道时间优先）访问下一个最近的

FCFS（先来先服务）按输入顺序访问

SCAN（扫描）先往一个方向到最远顺路访问，再往另一个方向到最远顺路访问

CSCAN（循环扫描）先往一个方向到最远顺路访问，再回到原点

·磁盘冗余阵列RAID

RAID：把多块独立的硬盘（物理硬盘）按照不同方式组合起来形成一个硬盘组（逻辑硬盘）。并行提高性能、数据冗余提高可靠性。

成本低、功耗小、传输速率高、容错。

·条带化

一个字节块可能存放在多个数据盘上。

性能好、磁盘负载均衡；可靠性差、不同I/O需要排队。

·校验

数据通过某种运算得出用以检验正确性的结果。

·RAID级别

RAID 0：并行交叉存取，无冗余。

RAID 1：镜像冗余。一倍冗余。

RAID 10：RAID 0+RAID 1。先分块后镜像or先镜像后分块。

RAID 2：海明码纠错（海明校验），按位条带化。对数级冗余。

RAID 3：奇偶校验（按位异或），字节级条带化。一个冗余。

RAID 4：并行处理磁盘阵列。数据块交叉。一个校验盘。

RAID 5：数据和校验都分布在磁盘中，没有专门的校验盘。

RAID 6：双维（相较RAID 5）校验独立存取盘阵列。

七、其他（略）安全、分布式系统

·密码学基本原理

·分布式系统体系结构