

目 录

第 1 章 操作系统概述	1	第 4 章 存储管理	74
第 2 章 进程管理	7	4.1 基本概念	74
2.1 进程及状态转换	7	4.2 逻辑地址和物理地址	81
2.2 进程同步	14	4.3 动态分区管理	85
2.3 信号量和同步	19	4.4 页面置换	86
2.4 信号量和前趋关系	23	第 5 章 设备管理	112
2.5 信号量应用	25	5.1 基本概念	112
第 3 章 处理机调度	50	5.2 磁盘调度算法	125
3.1 调度算法比较	50	第 6 章 文件管理	134
3.2 作业调度	52	6.1 基本概念及原理	134
3.3 死锁基本概念	59	6.2 磁盘管理	138
3.4 银行家算法	60	6.3 访问磁盘次数	145
3.5 死锁检测	63		

第3章 处理机调度

3.1 调度算法比较

重要内容回顾

关于调度算法的几点说明。

(1) 批处理系统、分时系统和实时系统中的主要调度算法：

批处理系统中既设有作业调度，又设有进程调度。批处理系统中的作业调度算法有先来先服务（FCFS）、短作业优先（SJF）、优先级调度（HPF）和高响应比优先（RF）。批处理系统的进程调度算法有：先进先出（FIFO）、短进程优先（SPF）、优先级调度（PRI）和高响应比优先（RF）。

分时系统中只设有进程调度，不设作业调度。其进程调度算法只有轮转法（RR）一种。

实时系统中只设有进程调度，不设作业调度。其进程调度算法有：轮转法（RR）、优先级调度算法（HPF）。后者又可细分为：非抢占式优先级调度、抢占式优先级调度（基于时钟中断的抢占式优先级调度和立即抢占的优先级调度）。

说明：实时系统中不可以使用先进先出（FIFO）和短进程优先算法（SPF）。

(2) 时间片轮转法：分时系统中，多个进程以轮流方式分享CPU，一般与进程的优先级、进程进入就绪队列的时间、进程的长短等无关。

习题精讲

1. 分时系统中的当前运行进程连续获得了两个时间片，原因可能是（ ）。

- A. 该进程的优先级最高
- B. 就绪队列为空
- C. 该进程最早进入就绪队列
- D. 该进程是一个短进程

【解析】对于分时系统，多个进程以轮流方式分享 CPU，一般来说不考虑进程的优先级、进程进入就绪队列的时间、进程的长短等。当前进程运行完一个时间片后回到就绪队列队尾，如果此刻就绪队列为空队列，那么下一个时间片仍然分给该进程。因此应该选择 B。

答案：B。

2. 若进程 P 一旦被唤醒就能够投入运行，则系统可能为（ ）。

- A. 在分时系统中，进程 P 的优先级最高
- B. 抢占调度方式，就绪队列上的所有进程的优先级皆比 P 的低

C. 就绪队列为空队列

D. 抢占调度方式, P 的优先级高于当前运行的进程

【解析】 本题考核考生对一个进程被调度的条件的理解, 相关概念有:

在分时系统中, 进程调度是按照轮转方式进行的。系统并不登记进程的优先级, 也就是说与 P 的优先级高低无关。

在抢占调度方式中, P 的优先级高于就绪队列上的所有进程, 但不一定高于当前的运行进程, 所以也不一定能立即运行。

无论哪种调度方式, 若就绪队列为空队列, P 被唤醒并插入后都会成为该队列的唯一进程。但这并不能说 P 可以立即获得处理机。只有当前运行的进程释放了 CPU, 且就绪队列无其他进程时, P 才可以立即使用 CPU。

在抢占式调度中, 一个进程 P 到来, 若 P 的优先级高于当前进程时, 可以抢占 CPU, 立即投入运行。因此应该选择 D。

答案: D。

3. 下列进程调度算法中, () 可能会出现进程长期得不到调度的情况。

A. 非抢占式静态优先权法

B. 抢占式静态优先权法

C. 时间片轮转调度算法

D. 非抢占式动态优先权法

【解析】 抢占式静态优先权法会出现进程长期得不到调度的情况。

答案: B。

4. 在单处理器的多进程系统中, 进程什么时候占用处理器和能占用多长时间, 取决于 ()。

A. 进程相应的程序段的长度

B. 进程总共需要运行多长时间

C. 进程自身和进程调度策略

D. 进程完成什么功能

【解析】 进程调度的目的是为进程分配处理机, 在进程调度算法中, 对于非抢占方式, 一旦进程获得处理机, 便会让该进程一直执行, 直到该进程完成或因某事件而阻塞才把处理机分配给其他进程; 对于抢占方式, 抢占的原则有时间片原则、优先权原则、短作业优先原则等, 其中都包含对进程占用处理机时间的约束。

答案: C。

5. 驱动调度算法中, () 算法可能会随时改变移动臂的运动方向。

A. 电梯调度

B. 最短寻找时间优先

C. 扫描

D. 单向扫描

【解析】 除了最短寻找时间优先之外的其余三种算法, 在磁头移动到磁道的尽头前都是单向移动。因此答案选择 B。

答案：B。

6. 若某单处理器多进程系统中有多个就绪态进程，则下列关于处理机调度的叙述中，错误的是（ ）。

- A. 在进程结束时能进行处理机调度
- B. 创建新进程后能进行处理机调度
- C. 在进程处于临界区时不能进行处理机调度
- D. 在系统调用完成并返回用户态时能进行处理机调度

【解析】 本题考查考生对处理机调度时机的理解。选项 A、B、D 显然是可以进行处理机调度的情况。对于 C，当进程处于临界区时，说明进程正在占用处理机，只要不破坏临界资源的使用规则，是不会影响处理机调度的。比如，访问的临界资源是慢速外设（如打印机），如果在进程访问打印机时，不能进行处理机调度，则系统的性能将非常低。几种不适合进行处理机调度的情况：①在处理中断的过程中；②进程在操作系统内核程序临界区中；③其他需要完全屏蔽中断的原子操作过程中。

答案：C。

7. 下列选项中，满足短任务优先且不会发生饥饿现象的调度算法是（ ）。

- A. 先来先服务
- B. 高响应比优先
- C. 时间片轮转
- D. 非抢占式短任务优先

【解析】 $\text{响应比} = \text{作业响应时间} / \text{作业执行时间} = (\text{作业执行时间} + \text{作业等待时间}) / \text{作业执行时间}$ 。高响应比算法，在等待时间相同的情况下，作业执行时间越短，响应比越高，优先执行，满足短任务优先。随着等待时间增加，响应比也会变大，执行机会就增大，所以不会产生饥饿现象。先来先服务和时间片轮转不符合短任务优先，非抢占式短任务优先会产生饥饿现象。

答案：B。

3.2 作业调度

重要内容回顾

熟练掌握常见调度算法的应用，并计算平均周转时间、平均加权周转时间等。下面是几点提示。

（1）解题步骤：确定多个任务的调度顺序→计算每个任务完成时间、周转时间

和带权周转时间→计算平均周转时间和带权平均周转时间。

(2) 在确定多个任务的调度顺序时,会受到任务到达时刻以及系统开始调度时刻的影响,常见有以下几种情况。

① 多个任务同时到达并开始调度。这种情况下,所有任务的到达时刻都可以理解为 0。

② 多个任务先后在不同时刻到达,但所有任务全部到达后才开始调度。

③ 多个任务先后在不同时刻到达,不要求所有任务全部到达后才开始调度。在这种情况下,某个作业到达时,如果后备队列为空,则立即得到调度。

习题精讲

1. 有三个作业 A (到达时间 8:50, 执行时间 1.5h)、B (到达时间 9:00, 执行时间 0.4h)、C (到达时间 9:30, 执行时间 1h)。当作业全部到达后,单道批处理系统按照响应比高者优先算法进行调度,则作业被选中的次序是 ()。

- A. (ABC) B. (BAC) C. (BCA)
D. (CBA) E. (CAB) F. (ACB)

【解析】 本题考核的知识点是采用高响应比优先调度算法的作业调度。根据题目,可以明确以下两点:作业全部到达后才开始调度,即作业全部到达后,才第一次计算所有作业的响应比,调度其中最高的;是单道批处理系统,即当前作业运行结束后,再计算剩余作业的响应比,调度其中最高的,直到所有作业运行结束。

对于本题,当作业全部到达后,也就是 9:30,系统开始调度。此刻各作业的等待时间如下: A 为 40 min (0.67h)、B 为 0.5 h、C 为 0 h。其响应比分别为:

$$\text{ack}_A = 1 + 0.67/1.5 = 1.45; \text{ack}_B = 1 + 0.5/0.4 = 2.25; \text{ack}_C = 1 + 0/1 = 1。$$

系统首先选 B 运行,至 9:54 运行结束。此时,作业 A 和 C 的等待时间分别是 1.1h 和 0.4h,其响应比分别修改为:

$$\text{ack}_A = 1 + 1.07/1.5 = 1.71; \text{ack}_C = 1 + 0.4/1 = 1.4。$$

系统再选 A 运行,至 11:24 运行结束。最后选择 C 运行至 12:24 结束。因此,应该选择 B。

答案: B。

2. 实时系统 ()。

- A. 是依赖人为干预的监督和控制系统
B. 必须既要及时响应、快速处理,又要有高可靠性和安全性
C. 强调系统资源的利用率
D. 实质上是批处理系统和分时系统的结合

【解析】 一个实时系统是指计算的正确性不仅取决于程序的逻辑正确性，也取决于结果产生的时间，如果系统的时间约束条件得不到满足，将会发生系统出错。所谓“实时”，是表示“及时”，而实时系统是指系统能及时响应外部事件的请求，在规定的时间内完成对该事件的处理，并控制所有实时任务协调一致的运行。实时系统的正确性不仅依赖系统计算的逻辑结果，还依赖于产生这个结果的时间。实时系统是能够在指定或者确定的时间内完成系统功能和外部或内部、同步或异步时间做出响应的系统。因此实时系统应该具有在事先定义的时间范围内识别和处理离散事件的能力；能够处理和存储控制系统所需要的大量数据。

答案：B。

3. 下面有关选择进程调度算法的准则错误的是（ ）。

- A. 尽量提高处理器利用率
- B. 尽可能提高系统吞吐量
- C. 适当增长进程在就绪队列中的等待时间
- D. 尽快响应交互式用户的请求

【解析】 在选择调度算法时，应考虑以下几个准则：

公平：确保每个进程获得合理的 CPU 份额。

有效：使 CPU 尽可能地繁忙。

响应时间：使交互用户的响应时间尽可能短。

周转时间：使批处理用户等待输出的时间尽可能短。

吞吐量：使单位时间处理的进程数尽可能最多。

答案：C。

4. 某系统采用基于优先权的非抢占式进程调度策略，完成一次进程调度和进程切换的系统时间开销为 $1\mu\text{s}$ 。在 T 时刻就绪队列中有 3 个进程 P1、P2 和 P3，其在就绪队列中的等待时间、需要的 CPU 时间和优先权如表 3.1 所示。【408 2018 年】

表 3.1 进程详情

进 程	等 待 时 间	需要的 CPU 时间	优 先 权
P1	$30\mu\text{s}$	$12\mu\text{s}$	10
P2	$15\mu\text{s}$	$24\mu\text{s}$	30
P3	$18\mu\text{s}$	$36\mu\text{s}$	20

若优先权值大的进程优先获得 CPU，从 T 时刻起系统开始进程调度，则系统的平均周转时间为（ ）。

- A. $54\mu\text{s}$ B. $73\mu\text{s}$ C. $74\mu\text{s}$ D. $75\mu\text{s}$

【解析】 最高优先级调度算法原则上总是调度就绪队列中优先级最高的那个进程。非抢占式和抢占式进程调度都属于最高优先级进程调度。采用非抢占式最高优先级调度算法，当就绪队列中某进程的最高优先级高于正在处理器中运行的进程的最高优先级时，并不会让正在运行的进程退出处理器，而是将高优先数的排在就绪队列的首部。而采用抢占式最高优先级进程调度算法，则高优先数的进程会抢占处理器，让正在处理的进程处于就绪队列。故应为 $1+24+1+36+1+12=75\mu s$ 。

答案：D。

5. 下列调度算法中，不可能导致饥饿现象的是（ ）。

- A. 时间片轮转
- B. 静态优先数调度
- C. 非抢占式短作业优先
- D. 抢占式短作业优先

【解析】 采用静态优先级调度时，当系统总是出现优先级高的任务，则优先级低的任务会因一直得不到处理机而产生饥饿现象；而短作业优先调度无论是抢占式或是非抢占，当系统连续出现新来的短任务时，长任务会因一直得不到处理机，产生饥饿现象，因此 B、C、D 都错误。

答案：A。

6. 有 5 个任务 A、B、C、D、E，它们几乎同时到达，预计运行时间分别为 10min，6 min，2 min，4 min，8min，优先级分别为 3、5、2、1 和 4，其中 5 为最高优先级。对于下列每一种调度算法，计算其平均进程周转时间（进程切换开销可不考虑）。

- (1) 先来先服务（按 A、B、C、D、E）算法。
- (2) 优先级调度算法。
- (3) 时间片轮转算法。

【解析】 本题中，由于 5 个任务同时到达，所以所有任务的到达时刻都可以理解为 0。

(1) 采用先来先服务（FCFS）调度算法时，5 个任务在系统中的执行顺序、完成时间及周转时间如表 3.2 所示。

表 3.2 采用先来先服务（FCFS）调度算法（单位：min）					
执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间
A	0	10	0	10	10
B	0	6	10	16	16
C	0	2	16	18	18
D	0	4	18	22	22
E	0	8	22	30	30

5 个进程的平均周转时间 T 为：

$$T=(10+16+18+22+30)/5=19.2\text{min}$$

(2) 采用最高优先级调度 (HPF) 算法时，5 个任务在系统中的执行顺序、完成时间及周转时间如表 3.3 所示。

表 3.3 采用最高优先级调度算法 (单位: min)

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间
B	0	6	0	6	6
E	0	8	6	14	14
A	0	10	14	24	24
C	0	2	24	26	26
D	0	4	26	30	30

它们的平均周转时间为：

$$T=(6+14+24+26+30)/5=20\text{min}$$

(3) 如果系统采用时间片轮转 (RR) 算法，令时间片为 2min，5 个任务轮流执行的情况为：

- 第 1 轮：就绪队列为 (A, B, C, D, E)，每个任务执行 2min，则 C 在本轮执行完毕，总用时 $T_1=5 \times 2=10$ ，其中 C 在第 $2 \times 3=6$ 时刻结束，故 $T_C=6\text{min}$ 。
- 第 2 轮：就绪队列为 (A, B, D, E)，每个任务执行 2min，则 D 在本轮执行完毕，总用时 $T_2=4 \times 2=8$ ，其中 D 在第 $2 \times 3=6$ 时刻结束，故 $T_D=T_1+6=10+6=16\text{min}$ 。
- 第 3 轮：就绪队列为 (A, B, E)，每个任务执行 2min，则 B 在本轮执行完毕，总用时 $T_3=3 \times 2=6$ ，其中 B 在第 $2 \times 2=4$ 时刻结束，故 $T_B=T_1+T_2+4=10+8+4=22\text{min}$ 。
- 第 4 轮：就绪队列为 (A, E)，每个任务执行 2min，则 E 在本轮执行完毕，总用时 $T_4=2 \times 2=4$ ，其中 E 在第 $2 \times 2=4$ 时刻结束，故 $T_E=T_1+T_2+T_3+4=10+8+6+4=28\text{min}$ 。
- 第 5 轮：就绪队列为 (A)，执行 2min，A 在本轮执行完毕，总用时 $T_5=1 \times 2=2$ ，其中 A 在第 $2 \times 1=2$ 时刻结束，故 $T_A=T_1+T_2+T_3+T_4+2=10+8+6+4+2=30\text{min}$ 。

平均周转时间 $T=(T_A+T_B+T_C+T_D+T_E)/5=(30+22+6+16+28)/5=20.4\text{min}$ 。

7. 在一单道批处理系统中，一组作业的提交时间和运行时间如表 3.4 所示。试计算以下 3 种作业调度算法的平均周转时间 T 和平均带权周转时间 W。

【西北大学 1998 年】

表 3.4 作业提交时间和运行时间表

作 业	提 交 时 间	运 行 时 间
1	8.0	1.0
2	8.5	0.5
3	9.0	0.2
4	9.1	0.1

- (1) 先来先服务。
- (2) 短作业优先。
- (3) 响应比高者优先。

【解析】 本题中，多个任务先后在不同时刻到达且不要求任务全部到达后才开始调度。也就是说，某个作业到达时，如果后备队列为空，则立即得到调度。

(1) 采用先来先服务（FCFS）调度算法时，4 个作业在系统中的调度顺序、完成时间、周转时间和带权周转时间如表 3.5 所示。

表 3.5 采用先来先服务（FCFS）调度算法

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
1	8.0	1.0	8.0	9.0	1.0	1.0
2	8.5	0.5	9.0	9.5	1.0	2.0
3	9.0	0.2	9.5	9.7	0.7	3.5
4	9.1	0.1	9.7	9.8	0.7	7.0

4 个作业的平均周转时间 T 为：

$$T=(1.0+1.0+0.7+0.7)/4=0.85$$

4 个作业的带权平均周转时间 W 为：

$$W=(1.0+2.0+3.5+7.0)/4=3.375$$

(2) 采用短作业优先调度算法时，4 个作业的调度顺序是这样的：作业 1 在时刻 8.0 到达，由于此时只有作业 1，所以作业 1 得到 CPU；在时刻 9.0，作业 1 完成，此时，作业 2 和作业 3 都已到达，由于作业 3 比作业 2 服务时间短，所以作业 3 得到 CPU；在时刻 9.2，作业 3 完成，此时，作业 4 也已经到达，由于作业 4 比作业 2 服务时间短，所以作业 4 得到 CPU；最后调度作业 2。

4 个作业在系统中的调度顺序、完成时间、周转时间和带权周转时间如表 3.6 所示。

表 3.6 短作业优先调度算法

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
1	8.0	1.0	8.0	9.0	1.0	1.0
3	9.0	0.2	9.0	9.2	0.2	1.0
4	9.1	0.1	9.2	9.3	0.2	2.0
2	8.5	0.5	9.3	9.8	1.3	2.6

4 个作业的平均周转时间 T 为：

$$T=(1.0+1.3+0.2+0.2)/4=0.675$$

4 个作业的带权平均周转时间 W 为：

$$W=(1.0+2.6+1.0+2.0)/4=1.65$$

（3）采用响应比高者优先调度算法时，4 个作业的调度顺序是这样的：作业 1 在时刻 8.0 到达，由于此时只有作业 1，所以作业 1 得到 CPU；在时刻 9.0，作业 1 完成，此时，作业 2 和作业 3 都已到达，计算他们的响应比，作业 2 的响应比为 $(0.5+0.5)/0.5=2$ ，作业 3 的响应比为 $(0+0.2)/0.2=1$ ，由于作业 2 的响应比比作业 3 高，所以作业 2 得到 CPU；在时刻 9.5，作业 2 完成，此时，作业 4 也已经到达，目前有作业 3 和作业 4，计算它们的响应比，作业 3 的响应比为 $(0.5+0.2)/0.2=3.5$ ，作业 4 的响应比为 $(0.4+0.1)/0.1=5$ ，由于作业 4 的响应比比作业 3 高，所以作业 4 得到 CPU；最后调度作业 3。

4 个作业在系统中的调度顺序、完成时间、周转时间和带权周转时间如表 3.7 所示。

表 3.7 短作业优先调度算法

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
1	8.0	1.0	8.0	9.0	1.0	1.0
2	8.5	0.5	9.0	9.5	1.0	2.0
4	9.1	0.1	9.5	9.6	0.5	5.0
3	9.0	0.2	9.6	9.8	0.8	4.0

4 个作业的平均周转时间 T 为：

$$T=(1.0+1.0+0.8+0.5)/4=0.825$$

4 个作业的带权平均周转时间 W 为：

$$W=(1.0+2.0+4.0+5.0)/4=3.0$$

8. 某进程调度程序采用基于优先数（priority）的调度策略，即选择优先数最小的进程运行，进程创建时由用户指定一个 nice 作为静态优先数。为了动态调整优先数，引入运行时间 cpuTime 和等待时间 waitTime，初值均为 0。进程处于执行态时，cpuTime 定时加 1，且 waitTime 置 0；进程处于就绪态时，cpuTime 置 0，waitTime 定时加 1。请回答下列问题。

（1）若调度程序只将 nice 的值作为进程的优先数，即 $\text{priority}=\text{nice}$ ，则可能会出现饥饿现象，为什么？

（2）使用 nice、cpuTime 和 waitTime 设计一种动态优先数计算方法，以避免产生饥饿现象，并说明 waitTime 的作用。

【解析】

（1）由于采用了静态优先数，只要就绪队列中存在优先数较小的进程，则优先数较大的进程一直没有机会运行，因而会出现饥饿现象。

（2）优先数 priority 的计算公式为：

$\text{priority}=\text{nice}+k_1 \times \text{cpuTime}-k_2 \times \text{waitTime}$ ，其中 $k_1>0$ ， $k_2>0$ ，用来分别调整 cpuTime 和 waitTime 在 priority 中所占的比例。waitTime 可使长时间等待的进程优先数减小，从而避免出现饥饿现象。

3.3 死锁基本概念

重要内容回顾

1. 资源竞争与死锁

设系统拥有某类资源的数量为 M ，现有 L 个进程，每个进程要求的该类资源数量为 N_i ($i=1, 2, \dots, L$)。在最坏的情况下，各进程都占用了 N_i-1 台，而且都在请求自己所需的最后一台，如果此时系统尚有多余的一台，那么就可以满足其中一个进程运行完毕，当该进程运行完毕释放出它所占有的该类资源，又可以进一步满足其他进程，系统就不会出现死锁。因此，如果 $M \geq \sum_{i=1}^L (N_i - 1) + 1$ ，则绝对不会出现死锁。

2. 区分死锁的预防、避免、检测与解除

（1）死锁预防：破坏“请求和保持”条件（或称为“资源静态分配法”）、破坏“不剥夺”条件、破坏“环路等待”（或称为“按序分配法”）条件。

（2）死锁避免：银行家算法。

（3）死锁检测：资源分配图简化法。

（4）死锁解除：剥夺资源（或称为“剥夺资源法”）、撤销进程（或称为“消进程法”）。

习题精讲

1. 一个计算机系统中拥有 6 台打印机，现有 N 个进程竞争使用，每个进程要求两台，试问， N 的值如何选取时系统绝对不会出现死锁？

【解析】当 $6 \geq N+1$ 时，绝对不会出现死锁，即 $N \leq 5$ 。

答案： $N \leq 5$ 。

2. 在下列选项中，属于检测死锁的方法是（ ）。

- A. 银行家算法
- B. 消进程法
- C. 资源静态分配法
- D. 资源分配图简化法

【解析】银行家算法用于死锁预防，消进程法用于死锁解除，资源静态分配法用于死锁避免，资源分配图简化法用于死锁检测。因此，应该选择 D。

答案：D。

3.4 银行家算法

重要内容回顾

银行家算法：

1. 判断某时刻是否为安全状态，如果是，其安全序列可能不唯一。
2. 利用银行家算法避免死锁：
 - （1） P_i 请求资源 $\leq P_i$ 资源需求量吗？如果是，继续步骤（2）；否则不能分配。
 - （2） P_i 请求资源 $\leq$ 剩余资源数吗？如果是，继续步骤（3）；否则不能分配。
 - （3）试分配并修改相应数据结构，继续步骤（4）。
 - （4）再利用安全性算法检查系统是否安全。如果是，则分配；否则不能分配。

利用银行家算法避免死锁时，算法不难，但容易发生计算错误，所以，规范计算过程很重要。

习题精讲

设系统中有 3 种类型的资源（A，B，C）和 5 个进程 P_1 、 P_2 、 P_3 、 P_4 、 P_5 ，A 资源的数量为 17，B 资源的数量为 5，C 资源的数量为 20。在 T_0 时刻系统状态如表 3.8 所示。系统采用银行家算法实施死锁避免策略。 【北京大学 1997 年】

表 3.8 T0 时刻系统状态

进 程	最大资源需求量			已分配资源数量		
	A	B	C	A	B	C
P1	5	5	9	2	1	2
P2	5	3	6	4	0	2
P3	4	0	11	4	0	5
P4	4	2	5	2	0	4
P5	4	2	4	3	1	4

- (1) T0 时刻是否为安全状态？若是，请给出安全序列。
- (2) T0 时刻若进程 P2 请求资源 (0, 3, 4)，是否能实施资源分配？为什么？
- (3) 在 (2) 基础上，若进程 P4 请求资源 (2, 0, 1)，是否能实施资源分配？为什么？
- (4) 在 (3) 基础上，若进程 P1 请求资源 (0, 2, 0)，是否能实施资源分配？为什么？

【解析】

(1) T0 时刻各类资源剩余资源数量为 (2, 3, 3)。利用银行家算法对 T0 时刻的资源分配情况进行分析，可得 T0 时刻的安全性分析情况，如表 3.9 所示。

表 3.9 T0 时刻的安全性分析情况

进 程	Work			Need			Allocation			Work + Allocation			Finish
	A	B	C	A	B	C	A	B	C	A	B	C	
P5	2	3	3	1	1	0	3	1	4	5	4	7	TRUE
P4	5	4	7	2	2	1	2	0	4	7	4	11	TRUE
P3	7	4	11	0	0	6	4	0	5	11	4	16	TRUE
P2	11	4	16	1	3	4	4	0	2	15	4	18	TRUE
P1	15	4	18	3	4	7	2	1	2	17	5	20	TRUE

- 此时存在一个安全序列 {P5, P4, P3, P2, P1}，故该状态是安全的。
- (2) 在 T0 时刻，若进程 P2 请求资源 (0, 3, 4)，因请求资源数 (0, 3, 4) > 剩余资源数 (2, 2, 3)，所以不能分配。
- (3) 在 (2) 基础上，若进程 P4 请求资源 (2, 0, 1)，按银行家算法进行检查：
P4 请求资源 (2, 0, 1) ≤ P4 资源需求量 (2, 2, 1)
P4 请求资源 (2, 0, 1) ≤ 剩余资源数 (2, 3, 3)
试分配并修改相应数据结构，资源分配情况如表 3.10 所示。

表 3.10 试分配 P4 所请求资源后的资源分配情况

进 程	Max			Allocation			Need			Available		
	A	B	C	A	B	C	A	B	C	A	B	C
P1	5	5	9	2	1	2	3	4	7	0	3	2
P2	5	3	6	4	0	2	1	3	4			
P3	4	0	11	4	0	5	0	0	6			
P4	4	2	5	4	0	5	0	2	0			
P5	4	2	4	3	1	4	1	1	0			

再利用安全性算法检查系统是否安全，可得到此刻的安全性分析情况，如表 3.11 所示。

表 3.11 安全性分析情况

进 程	Work			Need			Allocation			Work + Allocation			Finish
	A	B	C	A	B	C	A	B	C	A	B	C	
P4	0	3	2	0	2	0	4	0	5	4	3	7	TRUE
P5	4	3	7	1	1	0	3	1	4	7	4	11	TRUE
P3	7	4	11	0	0	6	4	0	5	11	4	16	TRUE
P2	11	4	16	1	3	4	4	0	2	15	4	18	TRUE
P1	15	4	18	3	4	7	2	1	2	17	5	20	TRUE

此时，存在一个安全序列{P4, P5, P3, P2, P1}，故该状态是安全的，可以立即将 P4 所申请的资源分配给它。

(4) 在 (3) 基础上，若进程 P1 请求资源 (0, 2, 0)，按银行家算法检查：

P1 请求资源 (0, 2, 0) ≤ P1 资源需求量 (3, 4, 7)，

P1 请求资源 (0, 2, 0) ≤ 剩余资源数 (0, 3, 2)，

试分配并修改相应数据结构，资源分配情况如表 3.12 所示。

表 3.12 试分配 P1 所请求资源后的资源分配情况

进 程	Max			Allocation			Need			Available		
	A	B	C	A	B	C	A	B	C	A	B	C
P1	5	5	9	2	3	2	3	2	7	0	1	2
P2	5	3	6	4	0	2	1	3	4			
P3	4	0	11	4	0	5	0	0	6			
P4	4	2	5	4	0	5	0	2	0			
P5	4	2	4	3	1	4	1	1	0			

再利用安全性算法检查系统是否安全。此时 Available 为 (0, 1, 2)，所以，Work 初始化为 (0, 1, 2)。显然，此时的 Work 已不能满足任何进程的资源需求，故系统进入不安全状态。因此系统不能将资源分配给 P1。

3.5 死 锁 检 测

重要内容回顾

死锁检测：利用资源分配图加以简化的方法，来检测系统处于某状态时是否为死锁。

(1) 从资源分配图中找出一个既不阻塞又非独立的进程节点 P_i ，在顺利的情况下运行完毕，释放其占有的全部资源，消去 P_i 的请求边和分配边。由于 P_i 释放了资源，这样便能使其他被阻塞的进程获得资源继续运行。

(2) 重复 (1)，经过一系列简化后若能消去图中节点的所有边，使所有进程节点均孤立，则称该图可完全简化，反之则不可完全简化。

结论：系统为死锁状态的充分条件为当且仅当系统的当前资源分配图不可完全简化。

习题精讲

1. 简化如图 3.1 所示的进程-资源图，并利用死锁定理给出相应结论。

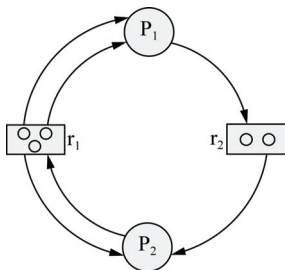


图 3.1 进程-资源图

【解析】如图 3.2 所示，其中：

(1) 查找图 3.2 (a) 中既不阻塞又非独立的进程结点：

① P_1 拥有 1 个 r_1 ，同时请求 1 个 r_2 。系统共有 2 个 r_2 且分配出去 1 个。因此 P_1 满足既不阻塞又非独立。

② P_2 拥有 1 个 r_1 和 1 个 r_2 ，同时请求 1 个 r_1 。系统共有 3 个 r_1 且全部分配

出去。因此 P_2 阻塞。

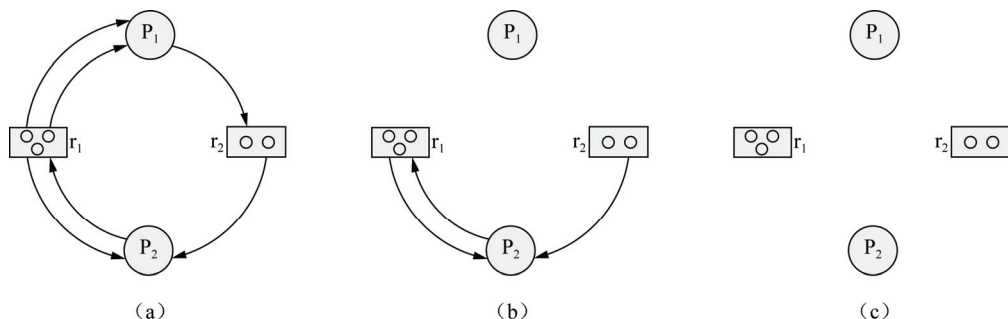


图 3.2 进程-资源图的简化

所以， P_1 为既不阻塞又非独立的进程节点，消去其所有相关边，如图 3.2 (b) 所示。

(2) P_1 结束运行，释放占有的全部资源。

(3) P_2 获得所需资源 r_1 ，继续运行。

(4) P_2 为既不阻塞又非独立的进程结点，消去其所有相关边，如图 3.2 (c) 所示。

(5) P_2 结束运行，释放占有的全部资源。

所以，该图可完全简化，没有发生死锁。

2. 当前运行的进程 ()，将引发系统进行进程调度。

- A. 执行了一条转移指令
- B. 要求增加主存空间，经系统调用银行家算法进行测算认为是安全的
- C. 执行了一条 I/O 指令
- D. 执行程序期间发生了 I/O 完成中断

【解析】 本题考核考生对进程调度时机的理解，相关的概念如下：

(1) 进程执行转移指令表示 CPU 将转到一个新程序段，并不是转到一个新进程，因而不会重新分配 CPU。

(2) 当前进程提出主存请求时，若系统认为分配是安全的，则可以立即使进程的请求得到满足，不会造成进程阻塞。因此不会分配 CPU。

(3) 当前进程执行了 I/O 指令，提出了输入/输出请求。由于 I/O 是低速的，不能让 CPU 等待 I/O 完成。因此需要阻塞当前进程，重新分配 CPU。

(4) 当前进程运行程序期间发生了 I/O 完成中断，说明有一个处于阻塞队列上的进程，正等待此 I/O 事件的出现。因此可在中断处理程序中查出等待的进程，将

其唤醒。然后返回到当前进程。只要恢复当前进程的现场信息就可以了，不会重新分配 CPU。

答案：C。

3. 为照顾紧迫型作业，应采用（ ）。

- A. 先来服务调度算法 B. 短作业优先调度算法
C. 时间片轮转调度算法 D. 优先权调度算法

【解析】考虑到紧迫型作业进入系统后能得到优先处理，引入最高优先级优先调度算法。

答案：D。

4. 一个作业 8:00 到达系统，估计运行时间为 1h。若 10:00 开始执行该作业，其响应比是（ ）。

- A. 2 B. 1 C. 3 D. 4

【解析】 本题考查考生对作业调度相关知识的理解。

响应比=作业响应时间/作业执行时间=(作业执行时间+作业等待时间)/作业执行时间。已知作业 8:00 到达系统，10:00 开始执行，那么等待时间为 2h，而估计运行时间为 1h，所以响应比=(1+2)/1=3。

答案：C。

5. 现有 3 个同时到达的作业 J1、J2 和 J3，它们的执行时间分别是 T1、T2 和 T3，且 $T_1 < T_2 < T_3$ 。系统按单道方式运行且采用短作业优先算法，则平均周转时间是（ ）。

【西安电子科技大学 2002 年】

- A. $T_1+T_2+T_3$ B. $(T_1+T_2+T_3)/3$
C. $(3T_1+2T_2+T_3)/3$ D. $(T_1+2T_2+3T_3)/3$

【解析】按照短作业优先算法，执行顺序为 J1、J2、J3，三个作业完成时间分别是 T1、T1+T2、T1+T2+T3，所以平均周转时间是 $(T_1+T_1+T_2+T_1+T_2+T_3)/3$ 。

答案：C。

6. 关于多级队列调度和多级反馈队列调度的叙述，不正确的是（ ）。

- A. 多级反馈队列调度中就绪队列的设置不是像多级队列调度一样按作业性质划分，而是按时间片的大小划分
B. 多级队列调度用到优先权，而多级反馈队列调度中没有用到优先权
C. 多级队列调度中的进程固定在某一个队列中，而多级反馈队列调度中的进程不固定
D. 多级队列调度中每个队列按作业性质不同采用不同的调度算法，而多级反馈队列调度中除了个别队列外，均采用相同的调度算法

【解析】 多级队列：该算法将系统中的进程就绪队列从一个拆分为若干个，将不同类型或性质的进程固定分配在不同的就绪队列，不同的就绪队列采用不同的调度算法。一个就绪队列中的进程可以设置不同的优先级，不同的就绪队列本身也可以设置不同的优先级。

多级队列调度算法由于设置多个就绪队列，因此对每个就绪队列可以实施不同的调度算法。系统针对不同用户进程的需求，很容易提供多种调度策略。

多级反馈队列：

(1) 设置多个就绪队列。在系统中设置多个就绪队列，并为每个队列赋予不同的优先级。第一个队列的优先级最高，第二个次之，其余的优先级逐个降低。该算法为不同的队列中的进程所赋予的执行时间片的大小也各不相同，在优先级愈高的队列，其时间片就愈小。

(2) 每个队列都采用 FCFS 算法。

(3) 按队列优先级调度。

答案：BCD。

7. 在操作系统中，死锁出现指的是（ ）。 【北京交通大学 2000 年】

- A. 计算机系统发生重大故障
- B. 资源数目远远小于进程数
- C. 若干进程因竞争资源而无限等待其他进程释放已占有的资源
- D. 进程同时申请的资源数超过资源总数

【解析】 死锁指两个或两个以上的进程在执行过程中，因争夺资源而造成的一种互相等待现象，若无外力作用，它们都将无法推进下去。此时称系统处于死锁状态或系统产生了死锁。由于资源占用是互斥的，当某个进程提出申请资源后，会使得有关进程在无外力协助下，因永远分配不到必需的资源而无法继续运行，这就产生了一种特殊现象：死锁。

答案：C。

8. 某系统采用了银行家算法，则下列叙述正确的是（ ）。

- A. 系统处于不安全状态时一定会发生死锁
- B. 系统处于不安全状态时可能会发生死锁
- C. 系统处于安全状态时可能会发生死锁
- D. 系统处于安全状态时一定会发生死锁

【解析】 银行家算法指系统在进行资源分配之前，面临多个进程的分配请求时，先计算此次分配资源的安全性。假如分配资源后不会导致系统进入不安全状态，则分配，否则等待。系统处于安全状态时，一定不会发生死锁；系统处于不安全状态

时，不一定会发生死锁。

答案：B。

9. 在下列选项中，属于解除死锁的方法是（ ）。

- A. 剥夺资源法
- B. 资源分配图简化法
- C. 银行家算法
- D. 资源静态分配法

【解析】 本题考查考生对解除死锁方法的掌握情况。解除死锁的方法有撤销进程法、剥夺资源法等。

答案：A。

10. 死锁的避免是根据（ ）采取措施实现的。【北京理工大学 2002 年】

- A. 配置足够的系统资源
- B. 使进程的推进顺序合理
- C. 破坏死锁的四个必要条件之一
- D. 防止系统进入不安全状态

【解析】 本题考查考生对避免死锁的理解。避免死锁的基本思路是在资源的动态分配过程中，用某种方法去防止系统进入不安全状态。因此应该选择 D。

答案：D。

11. 某系统中有 3 个并发过程都需要 4 个同类资源，该系统不会发生死锁的最少资源是（ ）。

- A. 9
- B. 10
- C. 11
- D. 12

【解析】 可能会发生死锁的上限资源数为每个进程均只差一个资源的情况，共为 9 个，只要再加一个资源就不可能发生死锁了，所以不会发生死锁的最少资源数为 10。

答案：B。

12. 下面关于检测死锁的叙述错误的是（ ）。

- A. 检测死锁方法对系统资源的分配不加限制，只要有则可以分配
- B. 检测死锁中系统需要反复检测各进程资源申请和分配情况
- C. 检测死锁是为了预防系统卷入死锁
- D. 检测死锁只能发现死锁，而不能消除死锁

【解析】 本题考查考生对检测死锁的理解。检测死锁对系统资源的分配不加限制，只要有资源即可进行分配。允许系统在运行过程中发生死锁，但可通过系统设置的检测机构，反复检测各进程资源申请和分配情况，及时检测出死锁的发生，并确定与死锁有关的进程和资源，然后通过解除死锁的措施，将已发生的死锁清除掉。检测死锁不是预防系统进入死锁。

答案：C。

13. 某计算机系统有 8 台打印机，由 K 个进程竞争使用，每个进程最多需要 3 台打印机。该系统可能会发生死锁的 K 的最小值是（ ）。

- A. 2 B. 3 C. 4 D. 5

【解析】 两个进程时，每个进程分配 3 台打印机，不会产生死锁；对于三个进程，可以有两个进程分别获得 3 台打印机，使其执行完并释放后让第三个进程获得 3 台，所以也不会产生死锁；对于四个进程，假若每个进程各获得 2 台而同时需要另外一台，产生了死锁，所以产生死锁的最小值是 4。

类似题型：假设现在有 P 个进程，每个进程最多需要 m 个资源，并且有 r 个资源可用。什么样的条件可以保证死锁不会发生。

如果一个进程有 m 个资源它能够结束，不会使自己陷入死锁中。因此最差情况是每个进程有 m-1 个资源并且需要另外一个资源。如果留下有一个资源可用，那么其中某个进程就能够结束并释放它的所有资源，使其他进程也能够结束。所以避免死锁的条件是： $r \geq p(m-1)+1$ 。

由此条件解上题： $r=8, m=3$ ，带入公式得 $2p \leq 7$ 。即当 p 小于等于 3 时才可保证死锁不会发生，所以可能会产生死锁的最小值是 4。

答案：C。

14. 假设有四个作业，它们的提交、运行时间如表 3.13 所示。若采用响应比高者优先调度算法，试问平均周转时间和平均带权周转时间为多少？

表 3.13 作业提交、运行时间
(时间单位：h，以十进制进行计算)

作 业 号	到 达 时 间	运 行 时 间
1	8.0	2.0
2	8.3	0.5
3	8.5	0.1
4	9.0	0.4

【解析】 响应比 = 响应时间 / 要求服务时间 = 1 + 等待时间 / 要求服务时间

在 8:00 时，由于只有作业 1 到达，因此系统将作业 1 投入运行。作业 1 运行 2h，于 10:00 完成。作业 1 完成后，剩余 3 个作业的响应比为：

$$r_2 = 1 + (10.0 - 8.3) / 0.5 = 4.4$$

$$r_3 = 1 + (10.0 - 8.5) / 0.1 = 16$$

$$r_4 = 1 + (10.0 - 9.0) / 0.4 = 3.5$$

r3 最大，所以让作业 3 运行。作业 3 运行 0.1h，在 10:1 完成，此时作业 2、作业 4 的响应比为：

$$r2=1+(10.1-8.3)/0.5=4.6$$

$$r4=1+(10.1-9.0)/0.4=3.75$$

r2 最大，所以让作业 2 运行。作业 2 完成后，运行作业 4。

所以，四个作业的调度次序为：作业 1、作业 3、作业 2、作业 4。

作业的执行顺序、完成时间、周转时间和带权周转时间如表 3.14 所示。

表 3.14 作业执行顺序与时间 (时间单位：h)

作业号	到达时间	运行时间	开始时间	完成时间	周转时间	带权周转时间
1	8.0	2.0	8.0	10.0	2.0	1.0
3	8.5	0.1	10.0	10.1	1.6	16.0
2	8.3	0.5	10.1	10.6	2.3	4.6
4	9.0	0.4	10.6	11.0	2.0	5.0

平均周转时间：T=(2.0+2.3+1.6+2.0)/4=1.975h

平均带权周转时间：W=(1+4.6+16+5)/4=6.65h

15. 假定要在一台处理机上执行下列作业，且假定这些作业在时刻 0 以 1、2、3、4、5 的顺序到达（参见表 3.15）。【西北工业大学 2000 年】

表 3.15 作业执行时间和优先级 (时间单位：h)

作 业	执 行 时 间	优 先 级
1	10	3
2	1	1
3	2	3
4	1	4
5	5	2

(1) 给出分别使用先来先服务算法（FCFS）、时间片轮转算法（RR，时间片=1）、最短作业优先算法（SJF）以及非抢占式优先调度算法（1 的优先级最高）时，这些作业的执行顺序。

(2) 针对上述每种调度算法，分别给出平均周转时间和平均带权周转时间。

【解析】 (1) 采用 FCFS（先来先服务算法）时的执行顺序、完成时间、周转时间和带权周转时间如表 3.16 所示。

表 3.16 采用 FCFS（先来先服务算法）时的执行顺序、完成时间、周转时间和带权周转时间（时间单位: h）

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
1	0	10	0	10	10	1
2	0	1	10	11	11	11
3	0	2	11	13	13	6.5
4	0	1	13	14	14	14
5	0	5	14	19	19	3.8

5 个进程的平均周转时间 T 为：

$$T=(10+11+13+14+19)/5=13.4h$$

5 个进程的平均带权周转时间 W 为：

$$W=(1+11+6.5+14+3.8)/5=7.26h$$

(2) 采用时间片轮转算法（RR，时间片=1）时的执行情况：

- 第 1 轮：（1， 2， 3， 4， 5）
- 第 2 轮：（1， 3， 5）
- 第 3 轮：（1， 5）
- 第 4 轮：（1， 5）
- 第 5 轮：（1， 5）
- 第 6 轮：（1）
- 第 7 轮：（1）
- 第 8 轮：（1）
- 第 9 轮：（1）
- 第 10 轮：（1）

5 个进程的周转时间为：T1=19、T2=2、T3=7、T4=4、T5=14。5 个进程的平均周转时间 T 为：

$$T=(19+2+7+4+14)/5=9.2h$$

5 个进程的带权周转时间为：W1=1.9、 W2=2、 W3=3.5、W4=4、W5=2.8。5 个进程的平均带权周转时间 W 为：

$$W=(1.9+2+3.5+4+2.8)/5=2.84h$$

(3) 采用 SJF（最短作业优先算法）时的执行顺序、完成时间、周转时间和带

权周转时间如表 3.17 所示。

表 3.17 采用 SJF（最短作业优先算法）时的执行顺序、完成时间、周转时间和带权周转时间

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
2	0	1	0	1	1	1
4	0	1	1	2	2	2
3	0	2	2	4	4	2
5	0	5	4	9	9	1.8
1	0	10	9	19	19	1.9

5 个进程的平均周转时间为：

$$T=(19+1+4+2+9)/5= 7$$

5 个进程的平均带权周转时间 W 为：

$$W=(1.9+1+2+2+1.8)/5=1.74$$

（4）采用非抢占式优先调度算法时的执行顺序、完成时间、周转时间和带权周转时间如表 3.18 所示。

表 3.18 采用非抢占式优先调度算法时的执行顺序、完成时间、周转时间和带权周转时间

执行次序	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
2	0	1	0	1	1	1
5	0	5	1	6	6	1.2
1	0	10	6	16	16	1.6
3	0	2	16	18	18	9
4	0	1	18	19	19	19

5 个进程的平均周转时间为：

$$T=(16+1+18+19+6)/5= 12$$

5 个进程的平均带权周转时间 W 为：

$$W=(1.6+1+9+19+1.2)/5=6.36$$

16. 某系统有 12 个磁带机，3 个进程 P0、P1 和 P2。每个进程对磁带机的最

大需求量以及当前持有磁带机的数量如表 3.19 所示。 【山东科技大学 2006 年】

表 3.19 每个进程对磁带机的最大需求量以及当前持有磁带机的数量

进 程	最大需求量 (Maximum Needs)	当前持有量 (Current Holdings)	可 用
P0	10	5	3
P1	4	2	
P2	9	2	

请问系统当前是否处于安全状态，并给出解释。

【解析】 在 T0 时刻，每个进程的资源需求都可在动态分配的过程中得到满足，存在一个安全序列 (p1, p0, p2) 或 (p1, p2, p0)，所以系统是安全的。

17. 试化简如图 3.3 所示的进程-资源图，并利用死锁定理给出相应的结论。

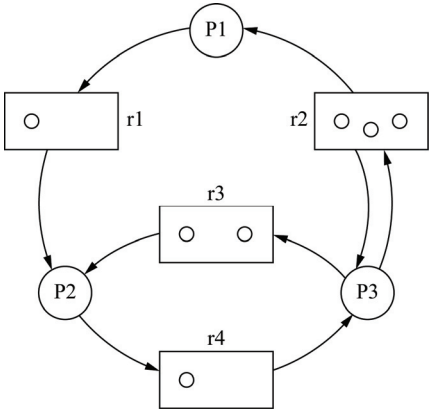


图 3.3 进程-资源图

【解析】 简化方法如下：

在图 3.4 (a) 中，找出一个既不阻塞又非独立的进程结点：P₁ 拥有一个 r₂ 并请求一个 r₁，但 r₁ 只有一个并且已经分配给了 P₂，因此 P₁ 阻塞；P₂ 拥有一个 r₁ 和一个 r₃ 并请求一个 r₄，但 r₄ 只有一个并且已经分配给了 P₃，因此 P₂ 阻塞；P₃ 拥有一个 r₂ 和一个 r₄ 并请求一个 r₂ 和一个 r₃，系统中 r₂ 总共有三个且只分配出去两个，系统中 r₃ 总共有两个且只分配出去一个，因此 P₃ 满足既不阻塞又非独立。因此选择 P₃，消去 P₃ 的请求边和分配边，如图 3.4 (b) 所示。

在图 3.4 (b) 中，P₂ 可以获得 r₄ 而不再阻塞。因此选择 P₂，消去 P₂ 的请求边和分配边，如图 3.4 (c) 所示。

在图 3.4 (c) 中，P₁ 可以获得 r₁ 而继续运行直至完成。选择 P₁，消去 P₁ 的请求边和分配边，如图 3.4 (d) 所示。

结论：该图是可完全简化的，没有发生死锁。

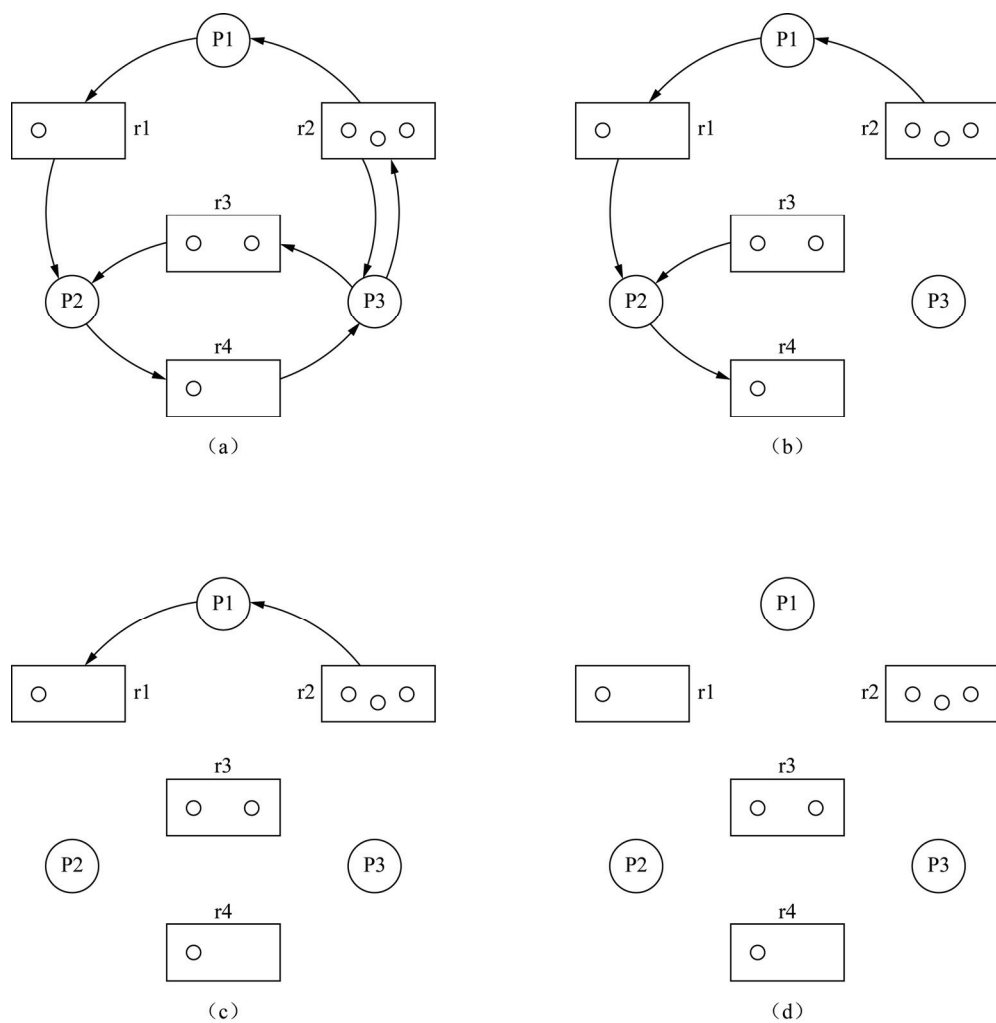


图 3.4 进程-资源图简化

第5章 设备管理

5.1 基本概念

重要内容回顾

(1) 共享设备：“一段时间内”允许多个进程同时访问的设备，任一时刻只能有一个进程访问。

(2) 虚拟设备：利用软件方法，比如 SPOOLing 系统，将一台独占的物理设备虚拟为若干台逻辑设备，供若干个用户/进程同时使用。

(3) I/O 通道：计算机上配置的专门用于输入/输出的处理机，可执行简单的 I/O 指令。通道没有独立内存，它所执行的通道程序存放在主机的内存中。

(4) 不同的 I/O 控制方式。

- 中断方式：通过执行中断服务程序完成 I/O 操作。
- DMA 方式：通过接管系统总线完成外设和内存之间的 I/O 操作。
- I/O 通道方式：通过执行通道程序完成 I/O 操作。
- 中断方式、DMA 方式和 I/O 通道方式都实现了 CPU 与 I/O 设备、I/O 设备之间的并行。
- DMA 方式和 I/O 通道方式实现了数据在设备和内存间的直接传输。
- I/O 通道方式的通道程序由系统组织，其传送过程需要 CPU 的干预最少。

(5) 引入缓冲技术后系统处理某数据所用总时间：设从磁盘将数据传送到缓冲区所用时间为 T ，将缓冲区数据传送到用户区所用时间为 M ，CPU 处理数据所用时间为 C 。则：

- 单缓冲： $\max(C, T) + M$ 。
- 双缓冲： $\max(C, T)$ 。

(6) 缓冲技术的实现方法。

- 硬件方法：适于速度和安全性要求较高的重要场合，费用很高。
- 软件方法：在内存中开辟出专门的存储区域当作缓冲区。

(7) 设备分配涉及的数据结构。

- 设备控制表 (DCT)：每个设备有一个 DCT，记录设备信息，如设备类型、设备标识符、设备状态、指向设备控制器表的指针、重复执行次数或时间、设备队列的队首指针等。

- 控制器控制表 (COCT): 每个控制器有一个 COCT, 记录控制器的信息, 如控制器标识符、控制器状态、与控制器连接的通道表指针、控制器队列的队首指针、控制器队列的队尾指针等。
- 通道控制表 (CHCT): 每个通道有一个 CHCT, 记录通道信息, 如通道标识符、通道状态、与通道连接的控制器表首址、通道队列的队首指针、通道队列的队尾指针等。
- 系统设备表 (SDT): 记录系统中全部设备信息的表格, 每个设备有一个表目, 包括设备类型、设备标识符、设备控制表、设备驱动程序入口地址等。

(8) 设备分配算法。

- 先来先服务。
- 优先级高者优先。
- 时间片轮转算法。该算法不能用于 I/O 调度, 主要原因如下:
 - 在 I/O 操作中, 大部分外部设备都是独占设备, 这种设备一旦被某进程占用, 直到使用完才能被释放。
 - I/O 设备的速度普遍比 CPU 慢, I/O 设备间来回切换的开销很大, 如果采用时间片轮转法, 会导致大量时间浪费在设备的启动和切换上。
 - 由于各 I/O 设备的数据传输速率差别较大, 时间片的大小不好确定。
 - 在通道方式中, 通道程序的执行是不受中断影响的, 时间片中断信号不能中断通道程序的操作。

(9) SPOOLing 系统: 又称“假脱机 I/O 系统”, 通过软件方法实现输入/输出。

其核心技术是使系统中的一台或少数几台独占设备“变成”多台可并行使用的虚拟设备。用一道程序模拟脱机输入时的外围控制机功能, 从而在主机的直接控制下, 实现脱机输入。用一道程序模拟脱机输出时的外围控制机功能, 从而在主机的直接控制下, 实现脱机输出。

习题精讲

1. 下面设备中属于共享设备的是 ()。 【南京理工大学 2001 年】

- A. 打印机 B. 磁带机 C. 磁盘 D. 磁带机和磁盘

【解析】本题考核的知识点是共享设备的概念和常用设备特征, 理解概念即可。

(1) 共享设备可寻址、可随机访问。

(2) 常见 I/O 设备。

- 典型的共享设备: 磁盘。
- 典型的独占设备: 打印机、磁带机、鼠标、键盘、语音输入输出设备等。

答案: C。

2. 有关设备管理概念的下列叙述中, () 是不正确的。

【西安电子科技大学 2002 年】

- A. 通道是处理输入/输出的软件
- B. 所有外围设备的启动工作都由系统统一来做
- C. 来自通道的 I/O 中断事件由设备管理负责处理
- D. 编制好的通道程序是存放在主存储器中的

【解析】 本题考核的知识点是有关设备管理的基本概念, 理解概念即可。

(1) 通道是计算机系统配置的专门用于输入/输出的设备, 为 I/O 系统的硬件组成部分。

(2) 一个外部设备对应一个设备驱动程序, 设备的驱动和输入/输出均由系统统一管理。

(3) 设备管理模块的底层软件配有专门处理设备中断的服务程序, 通道中断属于设备中断的一种。

(4) 通道没有独立内存, 通道程序存放在主机的内存中。

答案: A。

3. 利用通道实现了 () 之间数据的快速传输。

- A. CPU 和外设
- B. 内存和 CPU
- C. 内存和外设
- D. 外设和外设

【解析】 本题考核的知识点是通道的作用。

(1) 设置通道的目的: 原来由 CPU 处理的输入/输出操作交给通道完成, 把 CPU 从繁杂的输入/输出操作中解脱出来, 提高 CPU 效率。

(2) 通道与 CPU 之间的连接, 用于接收来自 CPU 的参数设置、启动外设的命令, 以及向 CPU 报告操作状态等。

(3) 通道与内存之间的连接, 主要实现内存和外设之间的数据快速传输。

(4) 通道程序来自内存。

答案: C。

4. 设备驱动程序是系统提供的一种通道程序, 它专门用于在请求 I/O 的进程与设备控制器之间传输信息。下面的选项中不是设备驱动程序功能的是 ()。

- A. 检查用户 I/O 请求的合法性
- B. 及时响应由控制器或通道发来的中断请求
- C. 控制 I/O 设备的 I/O 操作
- D. 了解 I/O 设备的状态, 传送有关参数, 设置设备的工作方式

【解析】 设备驱动程序的功能包括:

- (1) 将接收到的抽象要求转换为具体要求。
- (2) 检查用户 I/O 的合法性, 了解设备状态, 传递有关参数, 设置设备的工作方式。
- (3) 发出 I/O 操作命令, 启动 I/O 设备, 完成 I/O 操作。
- (4) 响应通道发来的中断请求, 根据中断类型调用相应中断处理程序。
- (5) 构造通道程序。

答案: C。

5. 假脱机技术中, 对打印机的操作实际上是用磁盘存储实现的。用以替代打印机的部分是指 ()。

- A. 共享设备 B. 独占设备 C. 虚拟设备 D. 物理设备

【解析】 本题考核的知识点是虚拟设备和 SPOOLing 概念。

SPOOLing 系统实现了将一台独占设备变换为若干台逻辑设备, 供若干个用户/进程同时使用的技术。这若干台逻辑设备称为虚拟设备。

答案: C。

6. 设从磁盘将一块数据传送到缓冲区所用时间为 $80\mu\text{s}$, 将缓冲区中数据传送到用户区所用时间为 $40\mu\text{s}$, CPU 处理数据所用时间为 $30\mu\text{s}$, 则处理该数据, 采用单缓冲传送某磁盘数据, 系统所用总时间为 ()。

- A. $120\mu\text{s}$ B. $110\mu\text{s}$ C. $150\mu\text{s}$ D. $70\mu\text{s}$

【解析】 本题考核的知识点是引入缓冲后系统处理某数据所用总时间的计算方法。

单缓冲的总时间为: $\max(C, T) + M = \max(30\mu\text{s}, 80\mu\text{s}) + 40\mu\text{s} = 120\mu\text{s}$

答案: A。

7. 下列关于通道、设备、设备控制器三者之间的关系叙述中正确的是 ()。

- A. 设备控制器和通道可以分别控制设备
- B. 设备控制器控制通道和设备一起工作
- C. 通道控制设备控制器, 设备控制器控制设备
- D. 设备控制器控制通道, 通道控制设备

【解析】 本题考核的知识点为设备与设备控制器、设备控制器与通道之间的关系。

(1) 设备控制器: CPU 与 I/O 设备之间的接口, 用于接收从 CPU 发来的命令, 并控制 I/O 设备工作。

(2) 通道: CPU 和设备控制器之间的设备, 通过执行通道程序, 与设备控制

器共同实现对 I/O 设备的控制和管理。

(3) 通道控制设备控制器，设备控制器控制设备。

(4) 通道不能直接控制 I/O 设备。

答案：C。

8. 在操作系统中，下列选项不属于软件机制的是（ ）。

- A. 缓冲池
- B. 通道技术
- C. 覆盖技术
- D. SPOOLing 技术

【解析】 本题考核的知识点为设备管理的基本方式。

(1) 通道是计算机上配置的、专门用于输入/输出的设备，属于硬件。

(2) 缓冲池、覆盖以及 SPOOLing 系统都是以软件方法实现的技术。

答案：B。

9. 表 5.1 所示是一段通道程序，四个选项中叙述不正确的是（ ）。

表 5.1 通道程序示例

操 作	P	R	计数 (B)	内 存 地 址
WRITE	0	1	90	743
WRITE	0	1	100	250
READ	0	1	230	1200
WRITE	0	0	120	400
WRITE	0	1	120	350
READ	1	1	70	2000

- A. 该段通道程序包括 6 条、两类通道指令
- B. 这些指令涉及的数据内存地址有相邻接的地方
- C. 该段通道程序共处理了 5 条记录
- D. 单记录最大为 230B

【解析】 本题考核的知识点为通道指令的格式和功能。

(1) 每一行为一条通道指令。

(2) 通道指令格式。

- 操作码：规定通道指令完成的操作。
- 通道程序结束标记 P：P=1 表示本通道指令是通道程序的最后一条指令。
- 记录结束标记 R：R=1 表示本通道指令是处理某记录的最后一条指令，R=0 表示本通道指令与下一条通道指令所处理的数据属于同一记录。
- 计数：本条指令所要读写的字节数。

● 内存地址：读写的内存区域首地址。

(3) 该段通道程序有 6 条指令。

① 第 1 条通道指令实现将内存 743~832 单元的 90 个字符写成 1 个记录。

② 第 2 条通道指令实现将内存 250~349 单元的 100 个字符写成 1 个记录。

③ 第 3 条通道指令读取内存 1200~1429 单元的 2300 个字符组成的 1 个记录。

④ 第 4、第 5 条通道指令实现将内存 400~519 单元的 120 个字符和 350~469 单元的 120 个字符共 240 个字符写成 1 个记录。

⑤ 第 6 条通道指令读取内存 2000~2069 单元的 70 个字符组成的 1 个记录。
通道程序结束。

(4) 该段通道程序出现了 WRITE 和 READ 两类通道指令。

(5) 第 2 条通道指令访问的内存区间 250~349 和第 5 条通道指令访问的内存区间 350~469 邻接。

(6) 题目所给通道程序涉及 5 条记录。

● 第 1~3 条指令和第 6 条指令各处理一条记录，共 4 条。

● 第 4 条指令和第 5 条指令共同处理一条记录，为 1 条。

● 第 4 条指令和第 5 条指令共同处理的记录长度为 240 (120+120) 个字符，占 240B 空间。

答案：D。

10. 对于速率为 9.6kb/s 的数据通信来说，如果设置一个具有 8 位的缓冲寄存器，则 CPU 中断时间和响应时间大约分别为 ()。

A. 0.8ms, 0.8ms

B. 8ms, 1ms

C. 0.8ms, 0.1ms

D. 0.1ms, 0.1ms

【解析】 本题考核的知识点为中断时间和响应时间的计算。

CPU 中断时间：对于传输速率为 9.6kb/s 的数据通信，产生中断的频率为 9.6kHz。如果设置了 8 位的缓冲寄存器可使中断频率降为 1/8，即， $9.6 \times 1024 / 8$ 。因此，设 CPU 的中断时间为 T_1 ，则 $T_1 = 8 / (9.6 \times 1024) \approx 0.8\text{ms}$ 。

CPU 的响应时间：设 CPU 的响应时间为 T_2 ，则 $T_2 = 1 / (9.6 \times 1024) \approx 0.1\text{ms}$ 。

注意：若再增设一个 8 位的缓冲寄存器，响应时间也可放宽到 0.8ms。

答案：C。

11. 某操作系统中，采用中断驱动 I/O 控制方式。设中断时，CPU 用 1ms 来处理中断请求，其他时间 CPU 完全用来计算，若系统时钟中断频率为 100Hz，则 CPU 的利用率为 ()。

A. 60%

B. 90%

C. 80%

D. 70%

【解析】 本题考核的内容是时钟中断处理时 CPU 的利用率。

首先，根据中断频率确定两次中断的间隔时间 $T=1/100=0.01s=10ms$ 。

其次，在 10ms 时间间隔内，CPU 需要花费 1ms 去处理中断，其他时间 ($10ms-1ms=9ms$) 用来进行计算。

最后，计算 CPU 的利用率 $p=9ms/10ms=90\%$ 。

答案：B。

12. 一个快速 SCSI-II 总线上的磁盘转速为 7200RPM，每磁道 160 个扇区，每扇区 512B。那么，理想状态下，其数据传输率为（ ）。【浙江大学 2006 年】

- A. $7200 \times 160kB/s$ B. 7200kB/s
C. 9600kB/s D. 19200kB/s

【解析】 本题考核内容为计算磁盘访问时的数据传输速率。磁盘数据传输速率的计算公式：磁盘每秒所转圈数 $\times$ 转一圈所能访问到的数据量。

题目给出的磁盘转速为 7200RPM，也就是 7200r/min，即 120r/s。

一圈访问一个磁道，一个磁道有 160 个扇区，而每个扇区有 512B。所以转一圈可以访问到的数据量为 $160 \times 512B=80KB$ 。

数据传输速率为 $120r \times 80KB=9600KB$ 。

答案：C。

13. 在下面的 I/O 控制方式中，需要 CPU 干预最少的方式是（ ）。

- A. 程序 I/O 方式 B. 中断驱动 I/O 控制方式
C. 直接存储器访问 DMA 控制方式 D. I/O 通道控制方式

【解析】 程序 I/O 方式，采用 busy-waiting 方式，即 CPU 采用轮询的方式来询问数据是否传送结束，效果最差。

中断 I/O 方式指设备控制器每取出一个数据后向 CPU 发送一个中断请求，然后 CPU 将数据从控制器中取到 CPU 寄存器中，然后再转移到内存中。中断方式下 CPU 以字节方式工作。

DMA 方式指 CPU 通过向 DMA 控制器设定若干参数，然后 DMA 打开一条内存到设备的通道，这样，设备和内存中的数据传送可以不通过 CPU 进行。

I/O 通道相当于一个简单的处理机，有自己的指令，数据传送流程为 CPU 向 I/O 通道发出指令，I/O 通道自动进行数据的获取。I/O 通道方式需要的 CPU 干预最少。

答案：D。

14. 下列哪一条不是磁盘设备的特点？（ ）

- A. 传输速率较高，以数据块为传输单位
B. 一段时间内只允许一个用户（进程）访问

- C. I/O 控制方式常采用 DMA 方式
- D. 可以寻址, 随机地读/写任意数据块

【解析】 磁盘设备传输速率较高, 以数据块为传输单位, I/O 控制方式常采用 DMA 方式; 可直接访问, 即随机读/写任意数据块; 允许多用户同时访问不同的数据块。

答案: B。

15. 下列存储设备中, 适合作为共享设备的是 ()。

- A. 语音输入/输出设备
- B. 打印机
- C. 鼠标
- D. 磁盘

【解析】 独享设备: 在一个用户作业未完成或退出之前, 此设备不能分配给其他作业用。所有字符设备都是独享设备, 如输入机、磁带机、打印机等, 需要安装驱动。

共享设备: 多个用户作业或多个进程可以“同时”从这些设备上存取信息。软硬盘、光盘等块设备都是共享设备——无须驱动。

虚拟设备: 通过软件技术将独享设备改造成共享设备。例如: 通过 SPOOLing 技术将一台打印机虚拟成多台打印机——实质上还是独享设备 (需要驱动)。

共享设备和独享设备之间的区别:

打印机、输入机和磁带机等都是独享设备, 在一个用户或者进程正在占用并且没有完成时, 不能释放这些独享设备, 也不能分配给其他作业使用 (需要驱动)。

硬盘、光盘等都是共享设备, 允许多用户作业和进程可以“同时”从这些设备上存取信息 (无须驱动)。

答案: D。

16. 系统中的通道数量较少, 可能会产生“瓶颈”问题。() 不是解决此问题的有效方法。

- A. 在结构上增加一些连线, 以增加数据传输通路
- B. 在数据传输线路上多增设一些缓冲区
- C. 提高 CPU 的速度
- D. 采用虚拟设备技术

【解析】 提高 CPU 的速度不能改变瓶颈问题。

答案: C。

17. I/O 系统硬件结构分为 4 级: 1. 设备控制器; 2. I/O 设备; 3. 计算机; 4. I/O 通道。按级别由高到低的顺序是 ()。

- A. 2—4—1—3
- B. 3—1—4—2
- C. 2—1—4—3
- D. 3—4—1—2

【解析】 I/O 系统硬件结构由高到低分别是：计算机、I/O 通道、设备控制器、I/O 设备。

答案：D。

18. 通过软件的功能扩充，把原来独占的设备改造成若干个用户可共享的设备，这种设备称为（ ）。 **【南京理工大学 2001 年】**

A. 存储设备 B. 系统设备 C. 虚拟设备 D. 用户设备

【解析】 采用虚拟技术，将独占设备改造成若干用户可共享的设备，这种设备是虚拟设备。

答案：C。

19. CPU 输出数据的速度远远高于打印机的打印速度，为解决这一矛盾，可采用（ ）。 **【南京理工大学 2001 年】**

A. 并行技术 B. 缓冲技术 C. 虚存技术 D. 覆盖技术

【解析】 缓冲技术应用在凡是数据到达速度和离去速度不匹配的地方，比如 CPU 和外设的速度不匹配时。

答案：B。

20. 关于 I/O 中断使用中不正确的描述是（ ）。

- A. I/O 中断是中央处理器和通道协调工作的一种手段
- B. 当设备故障时可形成操作异常 I/O 中断
- C. I/O 中断可用于表示输入/输出操作正常结束
- D. 通道根据 I/O 中断了解输入/输出操作的执行情况

【解析】 I/O 中断：中央处理器和通道协调工作的一种手段。通道借助 I/O 中断请求 CPU 进行干预，CPU 根据产生的 I/O 中断事件了解输入/输出操作的执行情况。I/O 中断事件是由于通道程序的执行或其他外界原因引起的。对通道操作而言，当操作正常结束或异常结束（如设备故障、设备特殊情况引起异常结束）就会形成 I/O 中断，由 CPU 根据相应情况分别处理。

答案：D。

21. 一个正在访问临界资源的进程由于申请等待 I/O 操作而被中断时（ ）。

- A. 可以允许其他进程进入与该进程相关的临界区
- B. 不允许其他进程进入任何临界区
- C. 可以允许其他就绪进程抢占处理器，继续运行
- D. 不允许任何进程抢占处理器

【解析】 进程进入临界区必须满足互斥条件，当进程进入临界区但是尚未离开时就被迫进入阻塞是可以的，系统中经常有这样的情形。在此状态下，只要其他进

程在运行过程中不寻求进入该进程的临界区,就应该允许其运行。该进程所锁定的临界区是不允许其他进程访问的,其他进程若要访问,必定会在临界区的“锁”上阻塞,期待该进程下次运行时可以离开并将临界区交给它。

答案: C。

22. 在采用 SPOOLing 技术的系统中,用户的打印数据首先被送到()。

【西北工业大学 1999 年】

A. 磁盘固定区域

B. 内存固定区域

C. 终端

D. 打印机

【解析】 SPOOLing 技术即假脱机技术,是使独占设备变成多台虚拟设备的一种技术。SPOOLing 系统是由“预输入程序”“缓输出程序”和“井管理程序”以及输入和输出井组成的。其中,输入井和输出井是为存放从输入设备输入的信息以及作业执行的结果,系统在辅助存储器上开辟的存储区域。SPOOLing 的工作过程是操作系统初启后激活 SPOOLing 预输入程序使其处于捕获输入请求的状态,一旦有输入请求,SPOOLing 输入程序立即执行,把装在输入设备上的作业输入到硬盘输入井中;当作业需要输出数据时,可以先将数据送到输出井,当输出设备空闲时,由 SPOOLing 输出程序把硬盘上输出井的数据送到慢速的输出设备上。

答案: A。

23. 设文件 F1 的当前引用计数值为 1,先建立 F1 的符号链接(软链接)文件 F2,再建立 F1 的硬链接文件 F3,然后删除 F1。此时, F2 和 F3 的引用计数值分别是()。

A. 0、1

B. 1、1

C. 1、2

D. 2、1

【解析】 本题考查的知识点为软/硬链接建立的属性。

建立符号链接(软链接)时,引用计数值直接复制。

建立硬链接时,引用计数值加 1。

删除文件时,删除操作对于符号链接是不可见的,这并不影响文件系统,当以后再通过符号链接访问时,发现文件不存在,可直接删除符号链接。但是对于硬链接则不可以直接删除,当引用计数值减 1,若值不为 0 时,不能删除此文件,因为还有其他硬链接指向此文件。

答案: B。

24. 程序员利用系统调用打开 I/O 设备时,通常使用的设备标识是()。

A. 逻辑设备名

B. 物理设备名

C. 主设备名

D. 从设备名

【解析】 本题考查的知识点为系统调用的设备标识。

用户程序对 I/O 设备的请求采用逻辑设备名,而在程序实际执行时使用物理设

备名。

答案：A。

25. 下列关于中断 I/O 方式和 DMA 方式比较的叙述中，错误的是（ ）。

- A. 中断 I/O 方式请求的是 CPU 处理时间，DMA 方式请求的是总线使用权
- B. 中断响应发生在一条指令执行结束后，DMA 响应发生在一个总线事务完成后
- C. 中断 I/O 方式下数据传输通过软件完成，DMA 方式下数据传输由硬件完成
- D. 中断 I/O 方式适用于所有外部设备，DMA 方式仅适用于快速外部设备

【解析】 中断处理方式：在 I/O 设备输入每个数据的过程中，由于无须 CPU 干预，因而可使 CPU 与 I/O 设备并行工作。仅当输完一个数据时，才需 CPU 花费极短的时间进行中断处理。因此中断方式请求的是 CPU 处理时间，中断响应在当前指令执行完毕后进行，数据在软件的控制下完成传输。

DMA 方式：数据传输的基本单位是数据块，即在 CPU 与 I/O 设备之间，每次传输至少一个数据块；DMA 方式每次申请的是总线的使用权，所传输的数据从设备直接送入内存的，或者相反；仅在传输一个或多个数据块的开始和结束时，才需 CPU 干预，整块数据的传输在 DMA 控制器的控制下完成的。

答案：D。

26. 下列关于外部 I/O 中断的叙述中，正确的是（ ）。 【408 2018 年】

- A. 中断控制器按所接收中断请求的先后次序进行中断优先级排队
- B. CPU 响应中断时，通过执行中断隐指令完成通用寄存器的保护
- C. CPU 只有在处于中断允许状态时，才能响应外部设备的中断请求
- D. 有中断请求时，CPU 立即暂停当前指令执行，转去执行中断服务程序

【解析】 外部中断是实时处理外部事件的一种内部机制。当某种外部事件发生时，中断系统将迫使 CPU 暂停正在执行的程序，转而去进行中断事件的处理；中断处理完毕后，又返回被中断的程序处继续执行。中断响应需要一些条件，比如当前指令执行完毕、中断允许等。中断优先级是按照中断事件的紧急状态设定的。

答案：C。

27. 下列有关 I/O 接口的叙述中，错误的是（ ）。

- A. 状态端口和控制端口可以合用同一个寄存器
- B. I/O 接口中 CPU 可访问的寄存器称为 I/O 端口

C. 采用独立编址方式时, I/O 端口地址和主存地址可能相同

D. 采用统一编址方式时, CPU 不能用访存指令访问 I/O 端口

【解析】采用统一编址时, CPU 访存和访问 I/O 端口用的是一样的指令, 所以访存指令可以访问 I/O 端口, D 选项错误, 其他三个选项均为正确陈述。

答案: D。

28. 若某设备中断请求的响应和处理时间为 100ns, 每 400ns 发出一次中断请求, 中断响应所允许的最长延迟时间为 50ns。则在设备持续工作过程中, CPU 用于该设备的 I/O 时间占整个 CPU 时间的百分比至少是 ()。

A. 12.5% B. 25% C. 37.5% D. 50%

【解析】每 400ns 发出一次中断请求, 而响应和处理时间为 100ns, 其中容许的延迟为干扰信息。因为在 50ns 内, 无论怎么延迟, 每 400ns 还是要花费 100ns 处理中断的。所以该设备的 I/O 时间占整个 CPU 时间的百分比为 $100\text{ns}/400\text{ns}=25\%$ 。

答案: B。

29. 某同步总线的时钟频率为 100MHz, 宽度为 32 位, 地址/数据线复用, 每传输一个地址或数据占用一个时钟周期。若该总线支持突发(猝发)传输方式, 则一次“主存写”总线事务传输 128 位数据所需要的时间至少是 ()。

A. 20ns B. 40ns C. 50ns D. 80ns

【解析】本题考查的知识点为总线传输速度的计算。总线频率为 100MHz, 则时钟周期为 10ns。总线宽度与存储字长都是 32 位, 故每次传送一个 32 位的字。猝发式发送可以连续传送地址连续的数据。故总的传送时间为: 传送地址 10ns, 传送 128 位数据 40ns, 共需 50ns。

答案: C。

30. 下列关于 USB 总线特性的描述中, 错误的是 ()。

- A. 可实现外设的即插即用和热插拔
- B. 可通过级联方式连接多台外设
- C. 是一种通信总线, 连接不同外设
- D. 同时可传输 2 位数据, 数据传输率高

【解析】本题考查的知识点是 USB 总线的特性。USB 总线(通用串行总线)的特点有: ①即插即用; ②热插拔; ③很强的连接能力, 采用菊花链形式将所有外设连接起来, 且不损失带宽; ④很好的可扩充性, 一个 USB 控制器可扩充高达 127 个外部周边 USB 设备; ⑤高速传输, 速度可达 480Mb/s。所以 A、B、C 都符合 USB 总线的特点。对于选项 D, USB 是串行总线, 不能同时传输两位数据。

答案: D。

31. 下列选项中, 在 I/O 总线的数据线上传输的信息包括 ()。

- I. I/O 接口中的命令字
- II. I/O 接口中的状态字
- III. 中断类型号

A. 仅 I、II B. 仅 I、III C. 仅 II、III D. I、II、III

【解析】 本题考查的知识点是 I/O 总线的特点。I/O 接口与 CPU 之间的 I/O 总线有数据线、命令线和地址线。命令线和地址线都是单向传输的, 从 CPU 传送给 I/O 接口, 而 I/O 接口中的命令字、状态字以及中断类型号均是由 I/O 接口发往 CPU。

答案: D。

32. 程序员利用系统调用打开 I/O 设备时, 通常使用的设备标识是 ()。

- A. 逻辑设备名 B. 物理设备名 C. 主设备号 D. 从设备号

【解析】 本题考查的知识点为系统调用的设备标识。用户程序对 I/O 设备的请求采用逻辑设备名, 而在程序实际执行时使用物理设备名。

答案: A。

33. 假定计算机的主频为 500MHz, CPI 为 4。现有设备 A 和 B, 其数据传输率分别为 2MB/s 和 40MB/s, 对应 I/O 接口中各有一个 32 位数据缓冲寄存器。请回答下列问题, 要求给出计算过程。 **【408 2018 年】**

(1) 若设备 A 采用定时查询 I/O 方式, 每次输入/输出都至少执行 10 条指令。设备 A 最多间隔多长时间查询一次才能不丢失数据? CPU 用于设备 A 输入/输出的时间占 CPU 总时间的百分比至少是多少?

(2) 在中断 I/O 方式下, 若每次中断响应和中断处理的总时钟周期数至少为 400 个, 则设备 B 能否采用中断 I/O 方式? 为什么?

(3) 若设备 B 采用 DMA 方式, 每次 DMA 传送的数据块大小为 1000B, CPU 用于 DMA 预处理和后处理的总时钟周期数为 500, 则 CPU 用于设备 B 输入/输出的时间占 CPU 总时间的百分比最多是多少?

【解析】

(1) 程序定时向缓存端口查询数据, 以防止部分数据因没有被及时读取而丢失。设备 A 准备 32 位数据所用时间为 $4B/2MB=2\mu s$, 所以最多每隔 $2\mu s$ 必须查询一次, 每秒的查询次数至少是 $1s/2\mu s=5\times 10^5$ 次, 每秒 CPU 用于设备 A 输入/输出的时间至少为 $5\times 10^5\times 10\times 4=2\times 10^7$ 个时钟周期, 占整个 CPU 时间的百分比至少是 $2\times 10^7/500M=4\%$ 。

(2) 中断响应和中断处理的时间为 $400\times(1/500M)=0.8\mu s$, 这时只需判断设备 B 准备 32 位数据需要的时间, 如果准备数据的时间小于中断响应和中断处理的时间,

那么数据就会被刷新、造成丢失。经过计算,设备 B 准备 32 位数据所用时间为 $4\text{B}/40\text{MB}=0.1\mu\text{s}$,因此,设备 B 不适合采用中断 I/O 方式。

(3) 在 DMA 方式中,只有预处理和后处理需要 CPU 参与,数据的传送过程由 DMA 控制。设备 B 每秒的 DMA 次数最多为 $40\text{MB}/1000\text{B}=40000$ 次,CPU 用于设备 B 输入/输出的时间最多为 $40000 \times 500 = 2 \times 10^7$ 个时钟周期,占 CPU 总时间的百分比最多为 $2 \times 10^7 / 500\text{M} = 4\%$ 。

5.2 磁盘调度算法

重要内容回顾

(1) 要牢记每种磁盘调度算法。先按照相应磁盘调度算法确定调度序列,再根据调度序列计算柱面移动次数。

(2) 计算磁盘访问时间。

寻道时间 T_s : 将磁臂从当前位置移到指定磁道上所经历的时间。按照相应磁盘调度算法确定调度序列并计算出柱面移动次数,将柱面移动次数转换为寻道时间。

旋转延迟时间 T_r : 指定扇区移动到磁头下面所经历的时间。

说明: 有效地减少延迟时间是提高磁盘传输效率的重要因素。磁盘是连续旋转的设备,磁盘机读/写一个物理块后,需要短暂的处理时间才能开始读/写下一块。

传输时间 T_t : 数据从磁盘读出或向磁盘写入数据所经历的时间。

习题精讲

1. 在磁盘调度算法中叙述正确的是 ()。

- A. SSTF 算法性能最好
- B. 磁盘调度算法的性能与队列中请求服务的数目无关
- C. 对 I/O 队列中的服务请求进行预处理有利于提高 C-LOOK 算法的性能
- D. C-SCAN 算法适用于磁盘负载较小的系统

【解析】 SSTF 算法的性能不是最好的,在某些情况下, C-SCAN 算法的性能优于 SSTF; 任何磁盘调度算法的性能都与队列中请求服务的数目有关,若队列中常常只有一个服务请求,则所有调度算法的效率几乎等价; 如果预先安排服务请求,使得服务请求按照旋转路径排列,那么 C-LOOK 算法可以得到最好性能; C-SCAN 算法在磁盘负载较大的情况下可以获得较高的性能,因为它几乎可以产

生 0 寻找时间。

答案：C。

2. 下列算法中用于磁盘移臂调度的是（ ）。【西安电子科技大学 2002 年】

- A. 时间片轮转法
- B. LRU 算法
- C. 最短寻找时间优先算法
- D. 优先级高者优先算法

【解析】 A、B、D 都是进程调度算法。因此答案选择 C。

答案：C。

3. 在下列有关旋转延迟的叙述中，不正确的是（ ）。

- A. 旋转延迟的大小与磁盘调度算法无关
- B. 旋转延迟的大小取决于磁盘空闲空间的分配程序
- C. 旋转延迟的大小与文件的物理结构有关
- D. 扇区数据的处理时间对旋转延迟的影响较大

【解析】 旋转延迟：从磁盘寻道结束开始，直到磁头旋转到 I/O 请求的起始数据块位置，其中的时间间隔称为旋转延迟。特定 I/O 操作请求旋转延迟的精确值，只能通过细致的磁盘驱动器仿真或者测量方法获得。一般将磁盘旋转周期的一半作为旋转延迟的近似值。实际应用中普遍采用该方法。因此，一般都将磁盘旋转周期的一半定义为旋转延迟。

寻道时间：磁头从当前位置移动到数据所在磁道所需要的时间。寻道时间越短，I/O 操作越快。

旋转延迟：盘片旋转将请求数据所在扇区移至读写磁头下方所需要的时间，旋转延迟取决于磁盘转速。

答案：D。

4. 有关设备的管理中，下列叙述中不正确的是（ ）。

- A. 计算机系统为每台设备确定一个绝对号
- B. 每台设备都应该有一个唯一的相对号
- C. 申请设备时指定绝对号可以提高设备的使用率
- D. 申请设备时指定设备相对号可使设备分配的灵活性更强

【解析】 设备的绝对号是其在计算机里的独一无二的身份标志，但在申请设备时使用相对设备号。在启动设备时则必须指出设备的绝对编号以唯一、不可更换地确定待启动设备。

答案：B。

5. 将系统调用参数翻译成设备操作命令的工作由（ ）完成。

- A. 用户层 I/O
- B. 设备无关的操作系统软件

【解析】 先按照相应磁盘调度算法确定调度序列，再根据调度序列计算柱面移动次数。

采用最短查找时间优先（最短寻道时间优先）算法的调度序列为 $15 \rightarrow 16 \rightarrow 13 \rightarrow 9 \rightarrow 20 \rightarrow 24 \rightarrow 29$ ，柱面移动次数为 $1+3+4+11+4+5=28$ 。

电梯调度算法的调度序列为 $15 \rightarrow 16 \rightarrow 20 \rightarrow 24 \rightarrow 29 \rightarrow 13 \rightarrow 9$ ，柱面移动次数为 $1+4+4+5+16+4=34$ 。

对于本例，最短查找时间优先算法（SSTF）的效率更高些。但是由于 SSTF 调度算法并不一定是效率最高的算法，而且会导致饥饿现象，所以通常情况下，操作系统并不采用最短查找时间优先算法。

8. 旋转型存储设备上信息的优化分布能减少若干输入/输出服务的总时间。例如，有 10 个记录 A, B, ..., J 存放在某磁盘的某一磁道上，假定这个磁道划分成 10 个扇区，每个扇区存放一个记录，安排如表 5.3 所示。现在要从该磁道上顺序地将 A~J 的 10 个记录读出，如果磁盘旋转一周需花费 20ms，处理程序每读出一个记录后花 4ms 进行处理。试问处理完 10 个记录的总时间是多少（从找到 A 记录开始计算）为了缩短处理时间，应该进行分布优化，试问应如何安排这些记录并计算优化后处理的总时间（从找到 A 记录开始计算）。

表 5.3 扇区及其对应记录号

扇 区	1	2	3	4	5	6	7	8	9	10
记录号	A	B	C	D	E	F	G	H	I	J

【解析】 这是一类考查磁盘调度方法的题目，主要通过对旋转延迟的理解来解答。问题的关键在于扇区数据的处理时间与旋转速度之间的差异，这个差异决定了扇区与记录之间的对应关系。题目中记录处理时间与扇区读取时间正好是倍数关系，因此 B~J 记录的旋转延迟都为 0，如果它们之间不是倍数关系，那么还应该计算旋转延迟。

根据磁盘旋转一周所花费的时间得知读取一个扇区所花费的时间是 $20(\text{ms}/r)/10(\text{扇区}/r)=2(\text{ms}/\text{扇区})$ 。而处理扇区 1 的记录 A 需要 4ms，因此当处理完扇区 1 的数据之后，磁头已经转过了 2 个扇区，到达扇区 4，因此还需要等待磁头旋转到扇区 2（相隔 8 个扇区）。这样读取扇区 2 上记录 B 的旋转延迟为 $8 \times 2 \text{ ms} = 16\text{ms}$ ，读取时间为 2ms，记录 B 的处理时间为 $16\text{ms} + 2\text{ms} + 4\text{ms} = 22\text{ms}$ 。依此类推，B, C, ..., J 记录的处理时间均为 22ms，因此总的处理时间为 $6\text{ms} + 9 \times 22\text{ms} = 204\text{ms}$ 。

为了降低旋转延迟，应该采取如表 5.4 所示的记录安排方式。

表 5.4 扇区及其对应记录号

扇 区	1	2	3	4	5	6	7	8	9	10
记录号	A	H	E	B	I	F	C	J	G	D

此时因为不存在旋转延迟，因此总的处理时间为 $10 \text{ 扇区} \times (2+4)\text{ms}=60\text{ms}$ 。

9. 假设计算机系统采用 CSCAN（循环扫描）磁盘调度策略，使用 2KB 的内存空间记录 16 384 个磁盘块的空闲状态。 【408 2010 年】

（1）请说明在上述条件下如何进行磁盘块空闲状态的管理。

（2）设某单面磁盘旋转速度为 6000r/min，每个磁道有 100 个扇区，相邻磁道间的平均移动时间为 1ms。若在某时刻，磁头位于 100 号磁道处，并沿着磁道号增大的方向移动（如图 5.1 所示），磁道号请求队列为 50，90，30，120，对请求队列中的每个磁道需读取 1 个随机分布的扇区，则读完这 4 个扇区总共需要多少时间？要求给出计算过程。

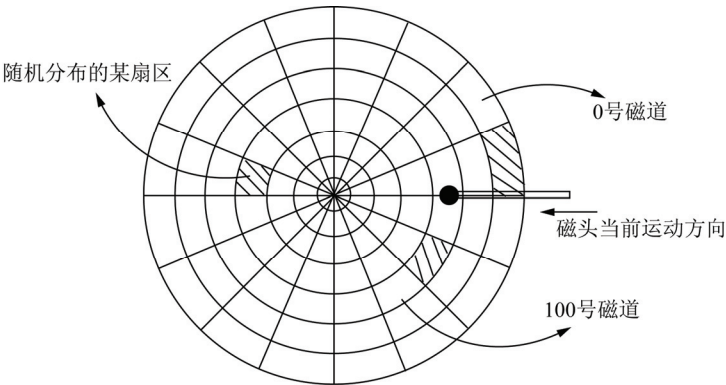


图 5.1 磁头运动方向示意图

（3）如果将磁盘替换为随机访问 Flash 半导体存储器（如 U 盘、SSD 等），是否有比 CSCAN 更高效的磁盘调度策略？若有，给出磁盘调度策略的名称并说明理由；若无，说明理由。

【解析】对于本题中的第 2 问，读完这 4 个扇区总共需要的时间由三部分组成：总寻道时间+4 个扇区总旋转延迟时间+4 个扇区总传输时间。其中，总寻道时间=柱面总移动次数×相邻磁道间的平均移动时间（已知为 1ms）；对于总旋转延迟时间，由于是读取随机分布的扇区，因此，应该使用平均旋转延迟时间，读取 1 个随机分布的扇区的平均旋转延迟时间=(0+转 1 圈所用时间)/2，因此，总旋转延迟时间=读取 1 个扇区的平均旋转延迟时间×4；对于总传输时间，总传输时间=1 个扇区

的传输时间 $\times 4$ ，1个扇区的传输时间=转1圈所用时间/每个磁道的扇区数量。

(1) $2\text{KB}=2\times 1024\times 8\text{b}=16\,384\text{b}$ 。因此可以使用位图法进行磁盘块空闲状态管理，每1位表示一个磁盘块是否空闲。

(2) 根据 CSCAN 算法，被访问的磁道号顺序为 100、120、30、50、90，因此，寻道用去的总时间为 $(20+90+20+40)\times 1\text{ms}=170\text{ms}$ ；由于每分钟 6000 转，因此转 1 圈的时间为 0.01s，一个扇区的读取时间为 $0.01\text{s}/100=0.0001\text{s}$ ，一个扇区的平均旋转延迟时间为 $(0+0.01)/2\text{s}$ ，总共要随机读取 4 个扇区，总共用去的旋转延迟时间和传输时间为 $(0.01\times 0.5+0.0001)\times 4=0.0204\text{s}=20.4\text{ms}$ ，所以，读完这 4 个扇区总共需要 $170\text{ms}+20.4\text{ms}=190.4\text{ms}$ 。

(3) 若将磁盘换为 U 盘等，则有比 CSCAN 更高效的磁盘调度策略。U 盘的存储介质为电可擦除只读存储器 (EEPROM) 的变种，其寻址方式类似于内存的寻址方式，不涉及磁盘的寻道、寻扇区等操作，因此可采用先来先服务、优先级高者优先等算法。

10. 设磁盘组共有 100 个柱面、8 个磁头，每个磁道划分成 8 个扇区，某文件记录存放到 3681 号逻辑磁盘块，请问该记录存放到哪个柱面的哪个磁道的哪个扇区？（柱面号、磁头号、扇区号、逻辑磁盘块号均从 0 开始编号）

【解析】 本题根据硬盘逻辑地址向物理地址转换的相关公式进行计算即可。具体如下：

C1: 起始扇区的柱面号。

H1: 起始扇区的磁头号。

S1: 起始扇区的扇区号。

NS: 每磁道的扇区数。

NH: 硬盘每柱面磁道数。

C、H、S 分别表示硬盘当前的柱面号、磁头号 and 扇区号。

$$C=((Ls \div NS) \div NH) + C1$$

$$H=((Ls \div NS) \bmod NH) + H1$$

$$S=(Ls \bmod NS) + S1$$

已知 $C1=0$ ， $H1=0$ ， $S1=0$ ， $NS=8$ ， $NH=8$ 。

$$\text{柱面号 } C=((3681 \div 8) \div 8) + 0 = 57$$

$$\text{磁头号 } H=((3681 \div 8) \bmod 8) + 0 = 4$$

$$\text{扇区号 } S=(3681 \bmod 8) + 0 = 1$$

11. 若磁头的当前位置为 100 磁道，磁头正向磁道号增加方向移动。现有一磁盘读写请求队列 23, 376, 205, 132, 19, 61, 190, 398, 29, 4, 18, 40，若采

用先来先服务、最短寻道时间优先和扫描算法，试计算出平均寻道长度各为多少？

【解析】

(1) 采用先来先服务磁盘调度算法，进行调度的情况如表 5.5 所示。

表 5.5 从 100 磁道开始进行调度的情况

下 一 磁 道	移动磁道数
23	77
376	353
205	171
132	73
19	113
61	42
190	129
398	208
29	369
4	25
18	14
40	22

移动磁道总数为 1596，平均寻道长度为 133。

(2) 采用最短寻道时间优先磁盘调度算法，进行调度的情况为如表 5.6 所示。

表 5.6 从 100 磁道开始进行调度的情况

下 一 磁 道	移动磁道数
132	32
190	58
205	15
61	144
40	21
29	11
23	6
19	4
18	1
4	14
376	372
398	22

移动磁道总数为 700，平均寻道长度为 58.3。

(3) 采用磁盘扫描调度算法，进行调度的情况如表 5.7 所示。

表 5.7 从 100 磁道开始进行调度的情况

下 一 磁 道	移动磁道数
132	32
190	58
205	15
376	171
398	22
61	337
40	21
29	11
23	6
19	4
18	1
4	14

移动磁道总数为 692，平均寻道长度为 57.7。

12. 假定磁盘块的大小为 1KB，对于 540MB 的硬盘，其文件分配表 FAT 需要占用多少存储空间？当硬盘容量为 1.2GB 时，FAT 需要占用多少空间？

【解析】 由题目所给条件可知，硬盘大小为 540MB，磁盘块的大小为 1KB，所以该硬盘共有盘块： $540\text{MB}/1\text{KB}=540\text{KB}$ 个。

又因为 $512\text{KB} < 540\text{KB} < 1024\text{KB}$ ，故 540K 个盘块号要用 20 位二进制表示，即文件分配表的每个表目为 2.5B。FAT 占用的存储空间总数为 $2.5\text{B} \times 540\text{K}=1350\text{KB}$ 。

当硬盘大小为 1.2GB 时，硬盘共有盘块 $1.2\text{GB}/1\text{KB} = 1.2\text{M}$ 个。

又因为 $1\text{M} < 1.2\text{M} < 2\text{M}$ ，故 1.2M 个盘块号需要 31 位二进制表示。为方便文件分配表的存取，每个表目用 32 位二进制表示，即 FAT 的每个表目大小为 4B。

FAT 要占用的存储空间总数为： $4 \times 1.2\text{MB} = 4.8\text{MB}$ 。

13. 有 5 个记录 A、B、C、D、E，存放在某磁盘的某磁道上，假定这个磁道划分成 5 块，每块存放 1 个记录，安排如表 5.8 所示。现在要顺序处理这 5 个记录，如果磁盘旋转 1 周需 20ms，处理程序每读出 1 个记录后要花 6ms 进行处理，试问：

(1) 处理完这 5 个记录所需的总时间是多少？

(2) 为减少磁盘旋转的周数，应如何安排这 5 个记录，并计算所需的总时间。

表 5.8 块号及其对应记录号

块 号	1	2	3	4	5
记录号	A	B	C	D	E

【解析】 本题是一个旋转调度问题。

(1) 所需的总时间 $= (20 \times 5 + 6) \cdot \text{ms} = 106\text{ms}$ 。因为，每转过 1 个记录需要 4ms，每读 1 个记录后要有 6ms 的处理时间，处理完毕后读取下一个记录时，需要再等一周。当读出第 5 个记录后还需增加 6ms 的处理时间。

(2) 将记录安排改进为：块号 1~5 分别存放记录 A、C、E、B、D。改进后，所需总时间为 $20 \times 3 \cdot \text{ms} = 60\text{ms}$ 。