一、处理机调度相关基本概念

1、调度方式和调度算法的若干准则

1)面向用户的准则：周转时间短（CPU执行用时Ts、周转时间T=Ts+Tw、带权周转时间W= T/Ts）、响应时间快、均衡性、截止时间的保证、优先权准则

2)面向系统的准则：系统吞吐量高、处理机利用率好、各类资源的平衡利用

3)批处理系统为照顾为数众多的短作业，应采用短作业优先的调度算法；分时系统为保证系统具有合理的响应时间，应采用轮转法进行调度

二、常用调度算法

1、先来先服务调度算法FCFS

（1）按照作业提交，或进程变为就绪状态的先后次序分派CPU；

（2）新作业只有当当前作业或进程执行完或阻塞才获得CPU运行

（3）被唤醒的作业或进程不立即恢复执行，通常等到当前作业或进程出让CPU。（所以，默认即是非抢占方式）

（4）有利于CPU繁忙型的作业，而不利于I/O繁忙的作业（进程）。

2、短作业（进程）优先调度算法SJF（非抢占）/SPF（抢占）

（1）平均周转时间、平均带权周转时间都有明显改善。SJF/SPF调度算法能有效的降低作业的平均等待时间，提高系统吞吐量。

（2）未考虑作业的紧迫程度，因而不能保证紧迫性作业（进程）的及时处理、对长作业的不利、作业（进程）的长短含主观因素，不一定能真正做到短作业优先。

3、高优先权优先调度算法HPF

（1）两种方式：非抢占式优先权算法、抢占式优先权算法（关键点：新作业产生时）

（2）类型:静态优先权：创建进程时确定，整个运行期间保持不变。动态优先权：创建进程时赋予的优先权可随进程的推进或随其等待时间的增加而改变。

（3）高响应比优先调度算法HRRN

HRRN为每个作业引入动态优先权，使作业的优先级随着等待时间的增加而以速率a提高：优先权 =（等待时间+要求服务时间)/要求服务时间= 响应时间 / 要求服务时间。

什么时候计算各进程的响应比优先权？（作业完成时、新作业产生时（抢占、非抢占）、时间片完成时、进程阻塞时）

4、基于时间片的轮转调度算法RR

（1）时间片轮转算法

过程：1、排成一个队列。2、每次调度时将CPU分派给队首进程。3、时间片结束时，发生时钟中断。4、暂停当前进程的执行，将其送到就绪队列的末尾，并通过上下文切换执行当前就绪的队首进程。

说明：1、进程阻塞情况发生时，未用完时间片也要出让CPU。2、能够及时响应，但没有考虑作业长短等问题。3、系统的处理能力和系统的负载状态影响时间片长度。

（2）多级反馈队列算法FB

过程：1、准备调度：先将它放入第一个队列的末尾，按FCFS原则排队等待调度。2、IF时间片内完成，便可准备撤离系统。3、IF时间片内未能完成，调度程序便将该进程转入第二队列的末尾等待再次被调度执行。4、当第一队列中的进程都执行完，系统再按FCFS原则调度第二队列。在第二队列的稍放长些的时间片内仍未完成，再依次将它放入第三队列。5、依次降到第n队列后，在第n队列中便采取按时间片轮转的方式运行。

说明:1、设置多个就绪队列，各队列有不同的优先级,优先级从第一个队列依次降低。2、赋予各队列进程执行时间片大小不同, 优先权越高，时间片越短。3、仅当优先权高的队列（如第一队列）空闲时，调度程序才调度第二队列中的进程运行。4、高优先级抢占时，被抢占的进程放回原就绪队列末尾。