

作业调度算法

崔 帅¹ 楚蓝天² 高 凯²

(1.中国矿业大学环境与测绘学院 江苏 徐州 221116;

2.中国矿业大学化工学院 江苏 徐州 221116)

【摘 要】在多道系统中,对批处理作业需要进行作业调度。作业调度是在资源满足的条件下,将处于后备状态的作业调入内存,同时生成与作业相对应的进程,并为这些进程提供所需要的资源。作业调度程序只能保证被调度的作业有获得处理器的资格,而处理器的分配则需要进程调度才能完成。作业调度需要根据作业控制块中的信息,检查系统是否满足作业的资源需求。只有在满足作业的资源需求的情况下,系统才能进行作业调度。下面是几种常见的作业调度算法:先来先服务、轮转法、多级反馈队列算法、优先算法、短作业优先法以及最高响应比优先法。本文将对这几种算法进行详细的介绍。

【关键词】先来先服务;轮转法;多级反馈队列算法;优先算法;短作业优先法;最高响应比优先法

1.先来先服务

1.1 定义

先来先服务(FCFS, First Come First Serve)是最简单的调度算法,按先后顺序进行调度。按照作业提交或进程变为就绪状态的先后次序,分派 CPU;当前作业或进程占用 CPU,直到执行完或阻塞,才出让 CPU(非抢占方式)。在作业或进程唤醒后(如 I/O 完成),并不立即恢复执行,通常等到当前作业或进程出让 CPU。

1.2 适用场景

比较有利于长作业,而不利短作业。有利于 CPU 繁忙的作业,而不利 I/O 繁忙的作业。

2.轮转法

2.1 定义

轮转法(Round Robin)是让每个进程在就绪队列中的等待时间与享受服务的时间成正比。将系统中所有的就绪进程按照 FCFS 原则,排成一个队列。每次调度时将 CPU 分派给队首进程,让其执行一个时间片。时间片的长度从几个 ms 到几百 ms。在一个时间片结束时,发生时钟中断。调度程序据此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的队首进程。

进程可以未使用完一个时间片,就出让 CPU(如阻塞)。

2.2 时间片长度的确定

时间片长度变化的影响:过长会导致退化为 FCFS 算法,进程在一个时间片内都执行完,响应时间长。过短会导致用户的一次请求需要多个时间片才能处理完,上下文切换次数增加,响应时间长。

对响应时间的要求: $T(\text{响应时间})=N(\text{进程数目}) \times q(\text{时间片})$ 。

就绪进程的数目:数目越多,时间片越小。

系统的处理能力:应当使用户输入通常在一个时间片内能处理完,否则使响应时间,平均周转时间和平均带权周转时间延长。

3.多级反馈队列算法

3.1 定义

多级反馈队列算法(Round Robin with Multiple Feedback)是轮转算法和优先算法的综合和发展。设置多个就绪队列,分别赋予不同的优先级,如逐级降低,队列 1 的优先级最高。每个队列执行时间片的长度也不同,规定优先级越低则时间片越长,如逐级加倍。新进程进入内存后,先投入队列 1 的末尾,按 FCFS 算法调度;若按队列 1 一个时间片未能执行完,则降低投入队列 2 的末尾,同样按 FCFS 算法调度;如此下去,降低到最后的队列,则按“时间片轮转”算法调度直到完成。仅当较高优先级的队列为空,才调度较低优先级的队列中的进程执行。如果进程执行时有新进程进入较高优先级的队列,则抢先执行新进程,并把被抢先的进程投入原队列的末尾。

3.2 优点

为提高系统吞吐量和缩短平均周转时间而照顾短进程。

为获得较好的 I/O 设备利用率和缩短响应时间而照顾 I/O 型进程。

不必估计进程的执行时间,动态调节。

4.优先级法

4.1 静态优先级

优先级算法(Priority Scheduling)是多级队列算法的改进,平衡各进程对响应时间的要求。适用于作业调度和进程调度,可分成抢先式和非抢先式。作业调度中的静态优先级大多按以下原则确定:由用户自己根据作业的紧急程度输入一个适当的优先级。由系统或操作员根据作业类型指定优先级。系统根据作业要求资源情况确定优先级。进程的静态优先级的确定原则:按进程的类型给予不同的优先级。将作业的静态优先级作为它所属进程的优先级。

4.2 动态优先级

进程的动态优先级一般根据以下原则确定:根据进程占用 CPU 时间的长短来决定。根据就绪进程等待 CPU 的时间长短来决定。

5.短作业优先法

5.1 定义

短作业优先(SJF, Shortest Job First)又称为“短进程优先”SPN(Shortest Process Next);这是对 FCFS 算法的改进,其目标是减少平均周转时间。对预计执行时间短的作业(进程)优先分派处理机。通常后来的短作业不抢先正在执行的作业。

5.2SJF 的特点

5.2.1 优点

比 FCFS 改善平均周转时间和平均带权周转时间,缩短作业的等待时间;提高系统的吞吐量;

5.2.2 缺点

对长作业非常不利,可能长时间得不到执行;未能依据作业的紧急程度来划分执行的优先级;难以准确估计作业(进程)的执行时间,从而影响调度性能。

6.最高响应比优先法

最高响应比优先法(HRN, Highest Response_ratio Next)是对 FCFS 方式和 SJF 方式的一种综合平衡。FCFS 方式只考虑每个作业的等待时间而未考虑执行时间的长短,而 SJF 方式只考虑执行时间而未考虑等待时间的长短。因此,这两种调度算法在某些极端情况下会带来某些不便。HRN 调度策略同时考虑每个作业的等待时间长短和估计需要的执行时间长短,从中选出响应比最高的作业投入执行。

响应比 R 定义如下: $R = (W + T) / T = 1 + W / T$ 。其中 T 为该作业估计需要的执行时间, W 为作业在后备状态队列中的等待时间。每当要进行作业调度时,系统计算每个作业的响应比,选择其中 R 最大者投入执行。这样,即使是长作业,随着它等待时间的增加, W / T 也就随着增加,也就有机会获得调度执行。这种算法是介于 FCFS 和 SJF 之间的一种折中算法。由于长作业也有机会投入运行,在同一时间内处理的作业数显然要少于 SJF 法,从而采用 HRN 方式时其吞吐量将小于采用 SJF 法时的吞吐量。另外,由于每次调度前要计算响应比,系统开销也要相应增加。

(上接第 115 页)【参考文献】

[1]曹景鑫,田辉,聂影.一起特大瓦斯窒息伤亡事故的调查[B].职业与健康,

2008.

[2]王金兰.河北省矿难状况实证分析.河北企业[J].2007.10.

[3]李叶宏,惠建利.中国矿难原因及防治对策[A].2006.11.