

# 任务优先级调度策略性能分析

王远振<sup>1</sup>, 赵坚<sup>2</sup>, 聂成<sup>1</sup>

(1. 空军工程大学 导弹学院, 陕西 三原 713800; 2. 北京跟踪与通信技术研究所 北京 100094)

**摘要:**针对任务优先级调度策略,对任务调度成功概率及其影响因素进行了分析。然后分析了具有柔性起始时间任务调度的性能,讨论了设备时间轴上剩余空隙的分布对调度成功率的影响。最后提出了提高任务调度成功概率的冲突解决策略。

**关键词:**调度策略; 任务调度; 评价指标; 冲突解决策略

**中图分类号:**V556.8 **文献标识码:**A **文章编号:**1009-3516(2003)03-0031-05

在多任务系统中,系统能否在指定时间内响应各种事件,完成相应处理任务是管理者和决策者关心的重要问题。当构成系统的硬件设备性能确定后,任务的调度策略便是影响系统服务质量的主要因素。不同的调度策略下有不同的冲突避免和冲突解决方法,最终形成的调度方案亦不同<sup>[1,4]</sup>。

优先级调度方法一般首先对申请使用同一设备的待调度任务进行优先级排序,对于同一优先级中的各任务按任务所需时间长短排序。然后依据任务的优先级顺序在该设备时间轴上为每个任务寻找一个可占用的设备时间。若当前正在进行调度分配的任务起始时间与已经调度的任务所占用的设备时间发生冲突,则该任务被放弃。然后,接着对已经排好序了的任务表中下一任务进行调度。

对调度策略进行评价的前提是建立相应的评价指标<sup>[1-2]</sup>。本文主要从用户(任务)角度通过申请任务调度成功概率的分析来评价优先级调度策略的性能及影响因素。第一部分分别对基于一个和多个设备的单个任务被成功调度的概率进行了分析。然后在更一般情况下,给出了任意多个任务被调度成功的概率的分析过程。接着给出了任务的起始时间具有柔性特性对调度成功率的影响情况。最后,讨论了设备时间轴上剩余空隙的分布对调度成功率的影响及其改进策略。

## 1 单任务调度成功概率分析

可通过分析在一个有若干已调度任务的设备时间轴上再安排另一任务的可能情况来给出调度成功概率。如图1有一个随机安排了 $n$ 个任务的设备时间轴,这 $n$ 个任务的起始时间和服务时间长度分别为

$$\{s_i, d_i\}, i=1, \dots, n,$$

并有: $s_i + d_i \leq s_{i+1}$ 。

考虑一个新的任务,它的服务时间长度为 $\delta$ ,起始时间 $\sigma$ 是随机的,那么,在下列情况下,该任务都将因为设备冲突而被放弃:

$$s_i \leq \sigma \leq s_i + d \text{ 或 } \sigma \leq s_i \leq \sigma \leq \delta.$$

既然 $\sigma$ 与已经调度的各任务互相独立,那么该任务与前面 $n$ 个任务不发生冲突的概率 $P_\sigma$ 由下式给出:

$$P_\sigma = t_i / t_a$$

收稿日期:2002-12-19

基金项目:国家“863”高技术(863-701-299-02)、国防预研基金(11301-01-01)资助项目

作者简介:王远振(1975-),男,安徽宿州人,博士生,主要从事军事优化理论与作战决策分析研究;

赵坚(1964-),男,四川通江人,高级工程师,国家“863-701”主题专家组专家,主要从事航天装备总体与发展战略研究;

聂成(1942-),男,安徽淮南人,教授,博士生导师,从事军事优化理论与作战决策分析研究。

这里,  $t_a$  是整个时间段的长度,  $t_r = t_a - \sum_{i=1}^n d_i$  是没有被前  $n$  个任务占用的时间段之和。

至于前  $n$  个任务的起始时间都不与当前任务冲突的概率可以如下考虑: 把已经分配给前  $n$  个任务的所有时间段删除掉, 这样原来的设备时间轴长度  $t_a$  被压缩为  $t_r$ , 在新设备时间轴  $t_r$  上原来已经调度了的每个任务的服务时间长度  $d_i$  被压缩为 0。假设  $\delta \ll t_r$  并且  $d_i \ll t_a$ , 于是前面已经调度了的某个任务的起始时间  $s_i$  不与当前正在等待调度的任务发生冲突的概率  $P_{s_i}$  可由下式给出:

$$P_{s_i} = \delta / t_r$$

综合考虑  $\sigma$  不与前  $n$  个

任务冲突的概率  $P_\sigma$  和每个  $s_i$  不与  $\sigma$  发生冲突的概率  $P_{s_i}$  可以得到任务  $\sigma$  被调度成功的概率  $P_1$ :

$$P_1 = \left( \frac{t_r}{t_a} \right) \left( 1 - \frac{\delta}{t_r} \right)^n$$

上式当  $n$  趋于无穷大时有:

$$P_1 \rightarrow \lim_{n \rightarrow \infty} P_1 = (1 - L) \times e^{-[i \frac{L}{1-L} \times \frac{\delta}{t_a}]}$$

式中  $L = \frac{\sum_{i=1}^n d_i}{t_a} = 1 - \frac{t_r}{t_a}$  可看作设备时间轴利用率,  $\langle d \rangle = \frac{1}{n} \sum_{i=1}^n d_i = \frac{L t_a}{n}$  是已经调度任务的平均持续时间长度。

若每一任务均可在  $m$  个功能相当的设备中的任一个上得到服务, 则,  $P_1$  就是当前任务可在  $m$  个设备中的每一个上得到成功调度的概率。当前任务可在  $m$  个设备中的任何一个上得到成功调度的概率  $P_m$  为

$$P_m = 1 - (1 - P_1)^m$$

图 2(图中曲线上数字表示设备数量) 给出了不同设备数量情况下, 具有平均持续时间长度的任务得到成功调度的概率与设备时间轴利用率  $L$  之间的函数关系。从图中可以看出当  $L$  值较小时, 调度成功的概率值  $P_m \approx 1 - L^m$ , 当  $L$  值较大时,  $P_m$  迅速减小。对于单设备情况, 当  $L = 0.3$  时,  $P_m$  即  $P_1$  降到 0.46, 而对于有四个相同设备的情况, 当  $L = 0.5$  时仍能保证  $P_m = P_4 \geq 0.5$ 。

## 2 所有任务的综合调度成功概率分析

考虑有  $m$  个同类型的设备可以为  $N$  个具有随机起始时间的任务提供服务的情形。对于按优先级排好序的任务表中前  $m$  个任务, 可以无冲突地获得调度成功, 每一个任务被分配到一套设备。因此获得成功调度的任务数及设备时间轴利用率分别为

$$n(m) = m, L(m) = \sum_{j=1}^m d_j / m t_a$$

对于任务  $j = m + 1, \dots, N$ , 由于第  $j$  个任务被成功调度的概率  $P_m(j) = P_m$  可以计算得到, 所以有:

$$n(j) = n(j-1) + P_m(j), L(j) = L(j-1) + P_m(j) d_j / m t_a$$

所有任务的平均调度情况可用对  $N$  个任务调度成功率  $n(N)/N$  表示或用成功调度时间比

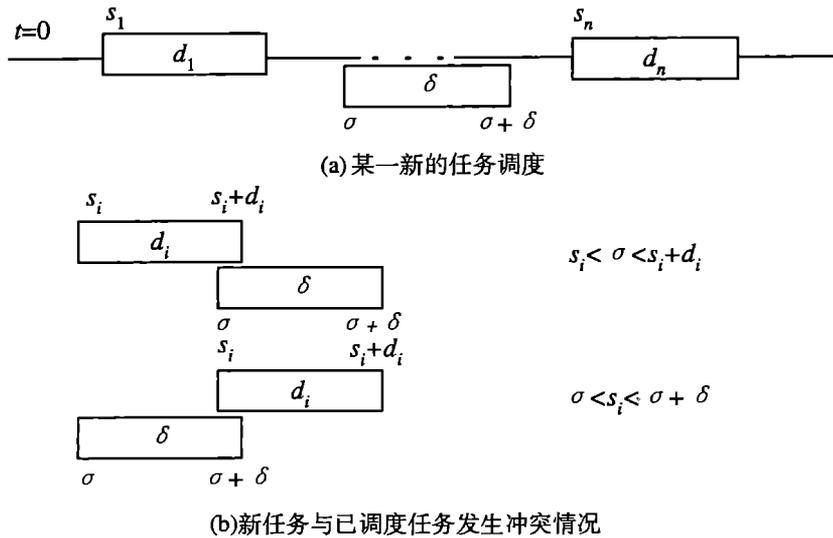


图 1 单任务调度分析示意

$mt_a L(N) / \sum_{j=1}^N d_j$ 来表示。当所有任务的持续时间长度都相等时,这两种表示方法是一样的。图3给出了不同设备数量情况下,具有同样持续时间长度的  $N$  个任务的平均调度成功率随任务需求时间与设备可用时间比的变化曲线(图中曲线上数字表示设备数量)。

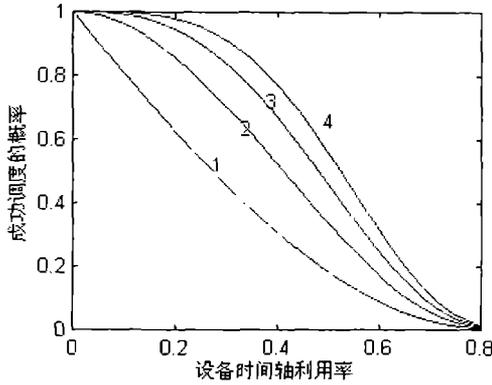


图2 单任务调度成功概率分析

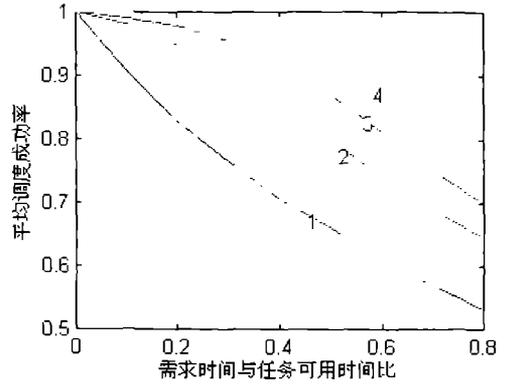


图3 多任务综合调度成功概率分析

### 3 柔性起始时间对调度的影响

有一些调度任务的起始时间具有柔性,即可在一定的范围内调整。当某一任务的起始时间的柔性为  $\tau$  时,只要能够在  $\sigma \leq t \leq \sigma + \tau$  内对该任务安排服务,都能满足该任务的设备使用需求。对于柔性起始时间对调度的影响可考虑如图4所示的情况。

在图4中任务  $\sigma$  与已调度好的任务  $s_i$  发生设备冲突,有  $s_i \leq \sigma \leq s_i + d_i$ 。如果  $\tau \geq s_i + d_i - \sigma$ ,那么,可以通过把任务  $\sigma$  的起始时间调整到  $s_i + d_i$ 来解决与任务  $s_i$ 的冲突。这一冲突能够得到解决的概率为

$$\begin{cases} \tau/d_i, & \text{当 } \tau < d_i \text{ 时} \\ 1, & \text{当 } \tau \geq d_i \text{ 时} \end{cases}$$

因此,任务  $\sigma$  与前面已调度的  $n$  个任务发生调度冲突时能够被解决的概率为

$$\sum_{i=1}^n \left( \frac{d_i}{t_a} \right) \left( \frac{\tau}{d_i} \right) = \frac{n\tau}{t_a} = L \frac{\tau}{<d>} = Lf$$

这里,  $f = \tau / <d>$  称为规格化起始时间柔性。

考虑柔性起始时间带来的冲突解决概率,有  $P_1 = (1 - L + Lf) \times e^{-[L/(1-L) \times \frac{\delta}{d_i}]}$ 。

图5给出了当起始时间柔性  $f = 1$  时,具有平均持续时间长度的任务调度成功的概率与设备时间轴利用率之间的函数关系曲线。同图2比较,柔性起始时间的存在能够在很大程度上增加任务调度成功概率。从图5可以看出:对于单个设备的情况,当设备时间轴的利用率  $L = 0.3$  时,调度成功率  $P_1$  从0.46增加至0.65;对于4个设备的情况,直到  $L = 0.65$  仍然有0.5的调度成功率。

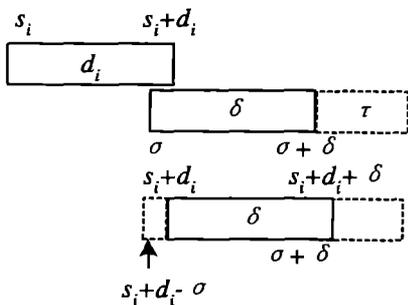


图4 柔性起始时间对调度的影响

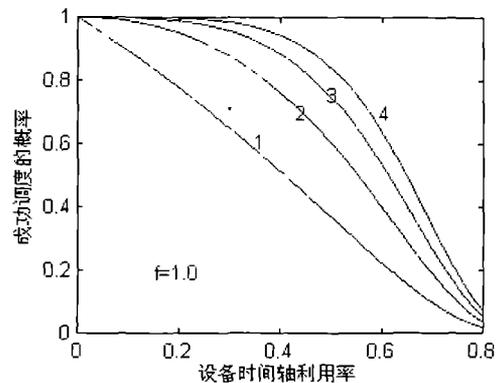


图5 柔性起始时间下单任务调度成功概率分析

随着起始时间柔性的提高,即随着  $f$  的增大,  $P_1 = (1 - L + Lf) \times e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$  增加。当某一待调度任务的  $\tau > d_i$  时,它的起始时间  $\sigma$  与任何已调度任务的冲突都可以通过将该任务的起始时间延迟到相冲突任务结束点后来解决。这时  $P_1 = (1 - L + Lf) \times e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$  中的线性部分  $(1 - L + Lf)$  被略去,于是有:  $P_1 = e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$ ,事实上,这一概率是单设备任务调度成功概率的一个下界。因为当已调度任务的起始时间  $s_i$  与当前待调度任务发生冲突且  $\sigma + \tau > s_i + d_i$  时,当前的调度冲突仍可以解决。

若  $\tau > d_i + d_{i+1} + \delta$  (对  $\delta = \langle d \rangle$  的情况,这个  $\tau$  值相当于  $f > 3$ ),任务调度成功概率将进一步增加,详见图 6。

图中 6,若待调度任务起始时间  $\sigma$  后的已调度任务的第一个任务间间隔  $g_i$  小于待调度任务持续时间长度  $\delta$ ,即  $g_i < \delta$ ,则该待调度任务在间隔  $g_i$  中是不可调度的,这一事件的概率是  $1 - e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$ 。同样待调度任务的第二个任务间间隔  $g_{i+1}$  也可能因为太小而不能安排该待调度任务。因此,该待调度任务能够被安排在  $g_i$  或  $g_{i+1}$  中的概率为  $P_1 = 1 - (1 - e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]})^2$ ,随着  $\tau$  的进一步增加,  $P_1$  进一步增大。

上述这些结论是基于如下假设的:即被调度任务是在它们的起始时间柔性范围内随机安排的。但我们可以使用很多方法来优选被调度任务的起始时间以优化设备的使用效果。这些方法通过有效减小小间隔的时间长度,增大大间隔的时间长度,进而增加调度成功机会。

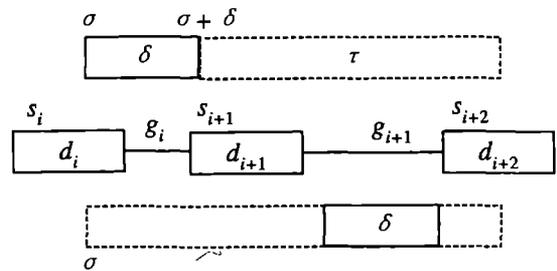


图 6 高柔性起始时间对调度的影响

#### 4 设备时间轴上剩余时间空隙长度分布对调度的影响

设备时间轴利用率太高和设备时间轴上剩余时间空隙太小是导致当设备时间轴利用率  $L > 0.6$  时,对新任务的调度越来越困难的主要原因。

某一空隙的长度大于  $g$  的概率  $v(g)$  可以通过考虑没有其它任务可被安排在从所选被调度任务结束点时间  $(s_i + d_i)$  开始长度为  $g$  的空隙中的概率得到。 $v(g)$  恰恰等于我们前面得到的无起始时间冲突地调度某一任务的概率,即:  $v(g) = e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$ 。

$$\text{剩余时间空隙长度分布由在 } g \sim g + dg \text{ 之间存在空隙概率给出: } - \langle d \rangle > \frac{dv(g)}{dg} = \frac{L}{1-L} e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$$

随着设备时间轴被更多地分配出去,剩余时间空隙同任务的平均持续时间长度相比变得越来越小,这使得这些空隙对一般持续时间长度的任务来说越来越没有使用价值。长度大于  $g$  的空隙中所剩时间总和  $T(g)$  为

$$T(g) = \int_g^\infty ng \frac{dv(g)}{dg} dg = n \left( g + \frac{(1-L) \times \langle d \rangle}{L} \right) e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$$

$T(g)$  与  $t_a$  以及  $t_r$  的比率分别为

$$\frac{T(g)}{t_a} = \left( \frac{Lg}{\langle d \rangle} + (1-L) \right) e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]} \text{ 和 } \frac{T(g)}{t_r} = \left( 1 + \frac{Lg}{(1-L) \langle d \rangle} \right) e^{-[\frac{L}{1-L} \times \frac{\delta}{\langle d \rangle}]}$$

当  $g = \langle d \rangle$  时,上面两式分别变为

$$\frac{T(g)}{t_a} = e^{-\frac{L}{1-L}} \text{ 和 } \frac{T(g)}{t_r} = \frac{L}{1-L} e^{-\frac{L}{1-L}}$$

图 7 给出了设备时间轴上长度大于平均任务持续时间长度的空隙中所剩时间量随设备时间轴利用率  $L$  变化的函数曲线。从图上可以看出:当设备时间轴利用率  $L$  达到 50% 时,只有 37% 的剩余部分是由大于平均任务持续时间长度的空隙构成的;当设备时间轴利用率  $L$  达到 70% 时,只有 10% 的剩余部分是由大于平均任务持续时间长度的空隙构成的,其余 20% 的剩余部分都是由长度太小而不能在其中安排具有平均持

续时间长度的任务的空隙构成的。可以通过调整被调度任务起始时间的办法减小这些无法利用的设备时间,进而增加大间隙部分的比重。

## 5 结论

通过以上分析,给出了具有柔性起始时间的任务在单个和多个设备上调度分配成功概率的计算公式。提出了确定综合调度成功概率的叠代技术。讨论了对于具有柔性特征任务的调度存在很大的改进余地。同时给出了设备时间轴上剩余时间空隙长度分布,并证明了随着设备时间轴被更多地分配出去,设备时间轴上所剩大间隙急剧减少是造成调度冲突的主要原因。

### 参考文献:

- [ 1 ] Surender D. Reddy. Generating and Evaluating Alternate Schedules[ A]. Network Technology Conference 1994, N94 - 21336 [ C]. 1994, 163 - 180.
- [ 2 ] Fredric Messing. Predicting Scheduling Success[ A]. Network Technology Conference 1994, N94 - 21334 [ C]. 1994, 137 - 150.
- [ 3 ] Ted Kaplan, Andrew Musliner, David Wampler. The Automated Conflict Resolution System[ A]. Network Technology Conference 1994, N94 - 21341 [ C]. 1994, 225 - 233.
- [ 4 ] Jeffrey S. Wike. Automating the Conflict Resolution Process[ A]. Space Network Conference on Resource Allocation Concepts and Approaches, 1992, N92 - 11061 [ C]. 1992, 265 - 269.
- [ 5 ] Upendra Belhe, Andrew Kusiak. Dynamic Scheduling of Design Activities with Resource Constraints[ J]. IEEE TRANS. SYS, MAN AND CYBER - PART A 1997, 27(1): 105 - 111.
- [ 6 ] 林 闯, 样士强. 多服务器多队列系统调度方案的性能分析[ J]. 电子学报, 2000, (5): 17 - 20.
- [ 7 ] 林 闯, 样士强. 多服务器多队列系统模型的分解和近似分析[ J]. 软件学报, 1999, (6): 6 - 12.
- [ 8 ] 孙志峻, 朱剑英. 具有柔性加工路径的作业车间智能优化调度[ J]. 机械科学与技术, 2001, (6): 931 - 935.
- [ 9 ] 舒忠梅, 胡金柱, 左亚尧. 实时系统中任务可调度性研究. 计算机应用研究[ J]. 2000, (1): 35 - 36.
- [ 10 ] James R Wertz, Wiley J Larson. 航天任务的分析与设计(下册)[ M]. 王长龙. 北京: 航空工业出版社, 1992.

(编辑: 田新华)

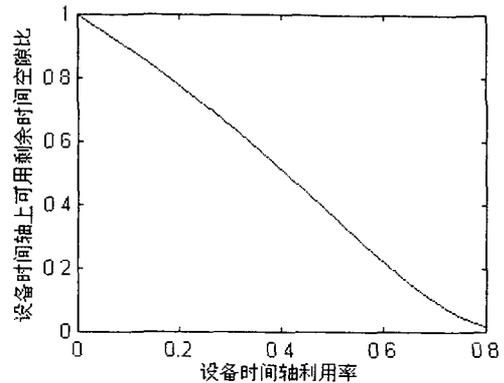


图7 设备时间轴上可用剩余时间空隙分析

## Performance Analysis of Activity Priority Scheduling Approach

WANG Yuan - zhen<sup>1</sup>, ZHAO Jian<sup>2</sup>, NIE Cheng<sup>1</sup>

(1. The Missile Institute, Air Force Engineering University, Sanyuan, Shaanxi 713800, China; 2. Beijing Institute of Tracking & Telecommunications Technology, Beijing 100094, China)

**Abstract:** Aimed at the activity priority scheduling approach, in this paper the probability of activity scheduling success and its influence factors are analysed, then the benefits to improved scheduling success from strat - time flexibility are included and the distribution of gaps remaining in the resource time line is discussed. Finally, conflict resolving policies are presented.

**Keyword:** scheduling policy; activity scheduling; evaluation indexes; conflict resolving approach