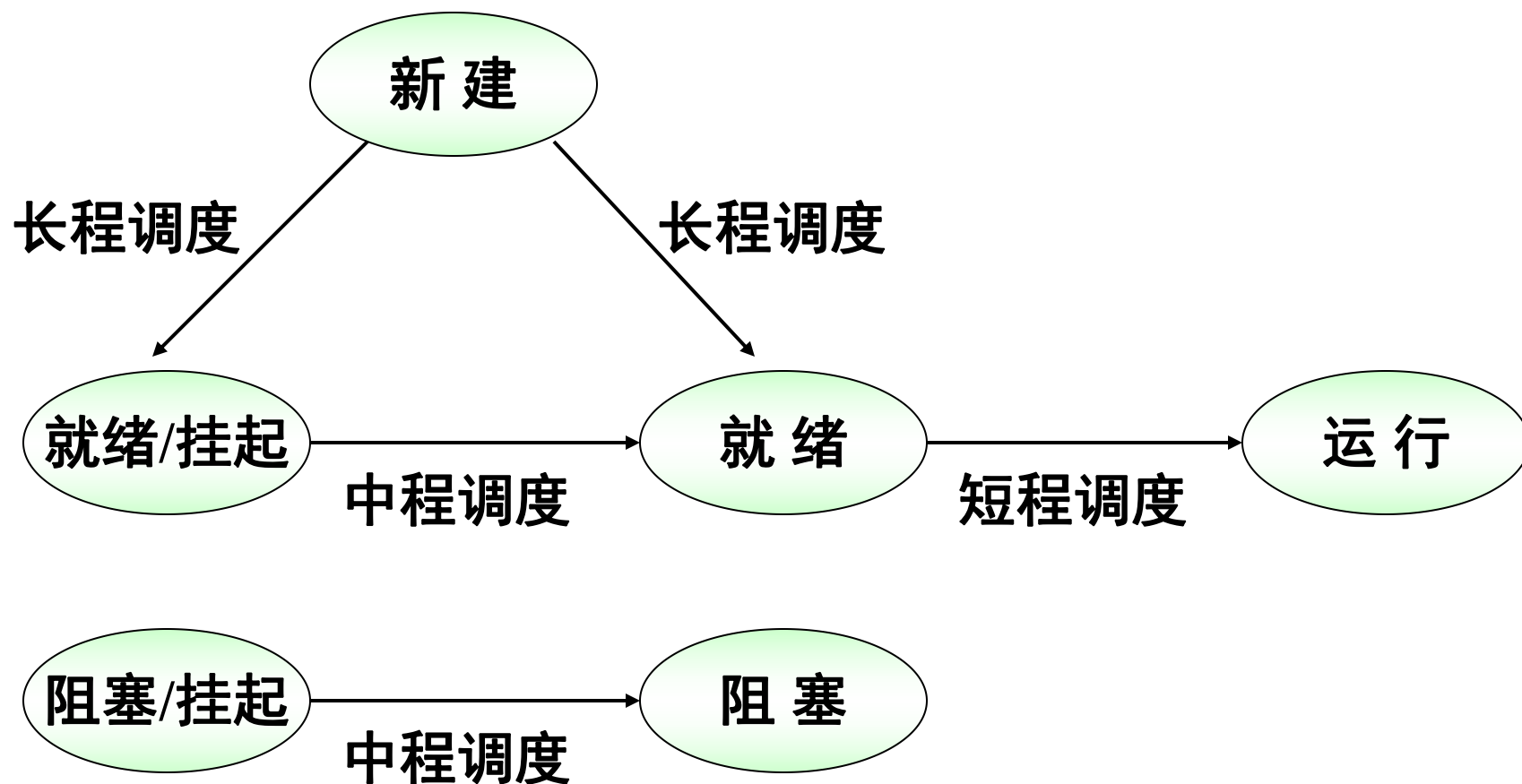


第9章 单处理器调度

- 主要内容

- 9.1 处理器调度的类型 ★★☆☆☆
- 9.2 调度算法 ★★☆☆☆
- 9.3 传统的UNIX调度 ★★☆☆☆

9.1 处理器调度的类型 ★★☆☆☆



9.1 处理器调度的类型 ★★★★★

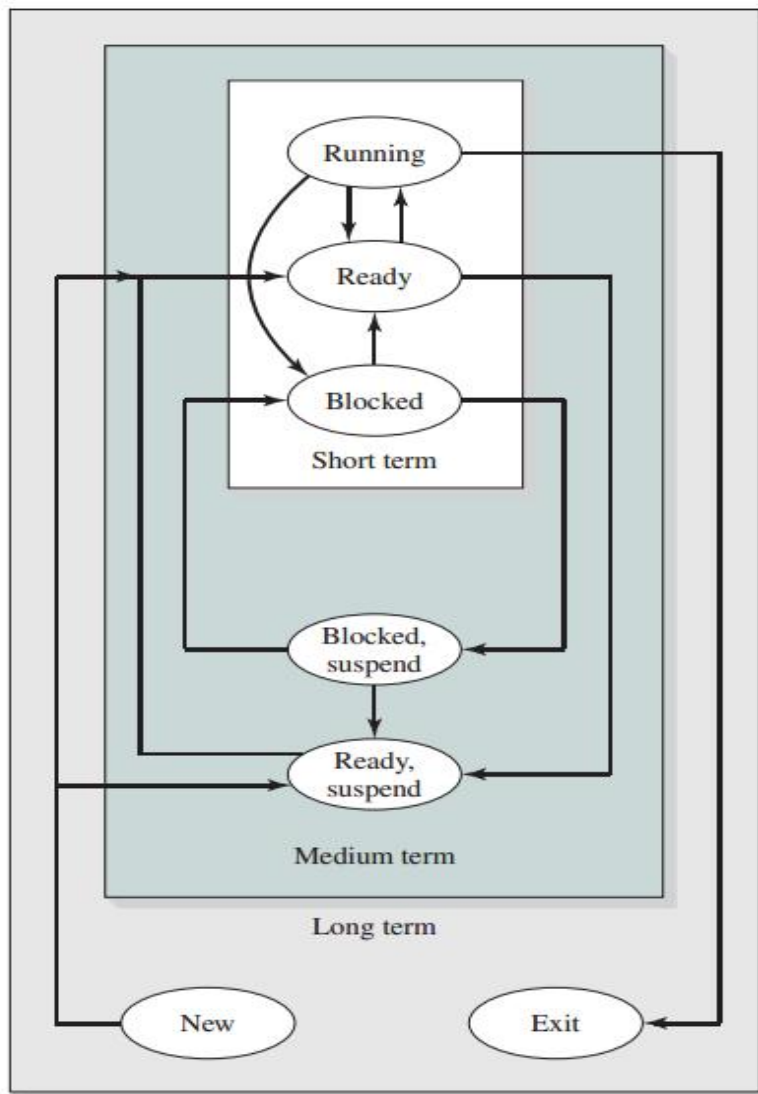


Figure 9.2 Levels of Scheduling

9.1.1 长程调度

- 长程调度程序决定哪个程序可以进入系统中处理，控制着系统的并发度。
- 何时调度？
 - 有作业终止时
 - 处理器的空余时间片超过了一定的阈值
- 调度哪个？
 - 先来先服务
 - 优先级
 - 实时性
 - I/O需求

9.1.2 中程调度

- 换入：取决于管理系统并发度的要求；
- 换出：进程的存储需求。

9.1.3 短程调度

- 长程调度程序执行频率较低；
- 中程调度程序执行频率稍高；
- 短程调度程序执行频率最高。
 - 当可能导致当前进程阻塞或可能抢占当前运行进程的事件发生时，调用短程调度程序。这类事件包括：时间中断；操作系统调用；I/O中断；信号（如信号量）
 - 分派程序，精确地决定下一次执行哪一个进程。

9.2 调度算法

9.2.1 短程调度准则

- 面向用户的准则
 - 单个用户或进程感知到的系统行为
 - 例：响应时间
- 面向系统的准则
 - 处理器使用的效果和效率
 - 例：吞吐量
- 与性能相关的准则
 - 是定量的，可以很容易地度量
 - 例：响应时间和吞吐量
- 与性能无关的准则
 - 或者本质上是定性的，或者不容易测量和分析
 - 例：可预测性

提问

1. 中程调度决定哪个程序可以进入系统处理中，因此它控制了系统的并发度。

A. 正确

B. 错误

2. 导致当前进程阻塞或抢占当前运行进程的事件发生时，会调用（）调度程序。

A. 长程

B. 中程

C. 短程

3. 在执行的频繁程度方面，（）调度执行的频率最高。

A. 长程

B. 中程

C. 短程

4. 进程由阻塞挂起态变为阻塞态，该转变由（）调度负责。

A. 长程

B. 中程

C. 短程

调度准则

1、面向用户，与性能相关

- 周转时间
 - 从提交到完成之间的时间间隔
- 响应时间
 - 从提交到开始接收响应之间的时间间隔
- 最后期限
 - 进程完成的最后期限

2、面向用户，与性能无关

- 可预测性
 - 希望提供给用户的服务能够随着时间的流逝展现给用户一贯相同的特性，而与系统执行的其他工作无关。

3、面向系统，与性能相关

- 吞吐量
 - 单位时间内完成的进程数目
- 处理器利用率
 - 处理器处于忙的状态的时间百分比

4、面向系统，与性能无关

- 公平性
 - 进程被平等对待
- 强制优先级
 - 进程被指定优先级，调度策略优先选择高优先级进程。
- 平衡资源
 - 保持系统中所有资源处于繁忙状态，较少适用紧缺资源的进程应该受到照顾。

9.2.2 优先级的使用

- 每个进程被指定一个优先级，调度程序总是选择具有较高优先级的进程。
 - 纯粹的优先级调度方案可能会导致低优先级进程长时间处于饥饿状态；
 - 一个进程的优先级可随着它的时间或执行历史而变化。

9.2.3 选择调度策略

- 选择函数
 - 确定在就绪进程中选择哪一个进程在下一次执行
 - 基于优先级？资源需求？进程执行特性？
- 决策模式
 - 非抢占
 - 抢占
- 调度准则
 - 周转时间 T_r
 - 归一化周转时间 T_r/T_s （周转时间与服务时间的比率）

9.2.3 选择调度策略

- 关键点
 - 调度时机（什么时候调度？）
 - 算法（怎么调度？）
 - 就绪队列（去哪里调度？）

Table 9.3 Characteristics of Various Scheduling Policies

	FCFS	Round Robin	SPN	SRT	HRRN	Feedback
Selection Function	$\max[w]$	constant	$\min[s]$	$\min[s - e]$	$\max\left(\frac{w + s}{s}\right)$	(see text)
Decision Mode	Non-preemptive	Preemptive (at time quantum)	Non-preemptive	Preemptive (at arrival)	Non-preemptive	Preemptive (at time quantum)
Throughput	Not emphasized	May be low if quantum is too small	High	High	High	Not emphasized
Response Time	May be high, especially if there is a large variance in process execution times	Provides good response time for short processes	Provides good response time for short processes	Provides good response time	Provides good response time	Not emphasized
Overhead	Minimum	Minimum	Can be high	Can be high	Can be high	Can be high
Effect on Processes	Penalizes short processes; penalizes I/O bound processes	Fair treatment	Penalizes long processes	Penalizes long processes	Good balance	May favor I/O bound processes
Starvation	No	No	Possible	Possible	No	Possible

1、先来先服务 FCFS

进 程	到达时间	服务时间
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2



0 3 9 13 18 20

平均周转时间: $(3+7+9+12+12)/5=8.6$

归一化周转时间: $(3/3, 7/6, 9/4, 12/5, 12/2)$

平均归一化周转时间: $(1+1.17+2.25+2.4+6)/5=2.56$

2、轮转 RR (1) $q=1$

进 程	到达时间	服务时间
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

- 时间片长度
- 其他处理器开销
- 基于时钟抢占

	A	A	B	A	B	C	B	D	C	B	E	D	C	B	E	D	C	B	D	D
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
			A	B	C	B	DC	CB	BED	EDC	DCB	CBE	BED	EDC	DCB	CB	BD	D		

就绪队列

平均周转时间: $(4+16+13+14+7)/5=10.8$

平均归一化周转时间: $(1.33+2.67+3.25+2.8+3.5)/5=2.71$

2、轮转 RR (1) $q=4$

进 程	到达时间	服务时间
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

	A	A	A	B	B	B	B	C	C	C	C	D	D	D	D	B	B	E	E	D
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
			B		C		CD	DB	DBE	DBE	DBE	BE	BE	BE	BE	ED	ED	D	D	

平均周转时间: $(3+15+7+14+11)/5=10$

平均归一化周转时间: $(1+2.5+1.75+2.8+5.5)/5=2.71$

3、最短进程优先 SPN

进 程	到达时间	服务时间
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

- 减少FCFS固有的对长进程的偏向
- 非抢占
- 某个进程结束后再选择！

	A	A	A	B	B	B	B	B	B	E	E	C	C	C	C	D	D	D	D	D
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
			B		C	C	CD	CD	CDE	CD	CD	D	D	D	D					

平均周转时间： $(3+7+11+14+3)/5=7.6$

平均归一化周转时间： $(1+1.17+2.75+2.8+1.5)/5=1.84$

4、最短剩余时间 SRT

进 程	到达时间	服务时间
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

- 对SPN增加抢占机制。

	A	A	A	B	C	C	C	C	E	E	B	B	B	B	B	D	D	D	D	D
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
			B		B		BD	BD	BD	BD	D	D	D	D	D					

B:5 B:5 B:5

C:4 C:2 D:5

D:5 E:2

平均周转时间: $(3+13+4+14+2)/5=7.2$

平均归一化周转时间: $(1+2.17+2+2.8+1)/5=1.59$

5、最高响应比优先 HRRN

- $R = (w + s) / s$
 - R: 响应比
 - w: 等待处理器的时间
 - s: 预计的服务时间

5、最高响应比优先 HRRN

进 程	到达时间	服务时间
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

B运行完，计算CDE的响应比：

$$R_C = (5+4) / 4 = 2.25$$

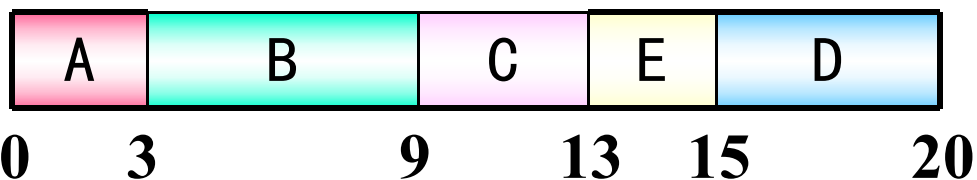
$$R_D = (3+5) / 5 = 1.6$$

$$R_E = (1+2) / 2 = 1.5$$

C运行完，计算DE的响应比：

$$R_D = (7+5) / 5 = 2.4$$

$$R_E = (5+2) / 2 = 3.5$$



- 进程完成时或阻塞
计算响应比？

平均周转时间： $(3+7+9+14+7)/5=8$

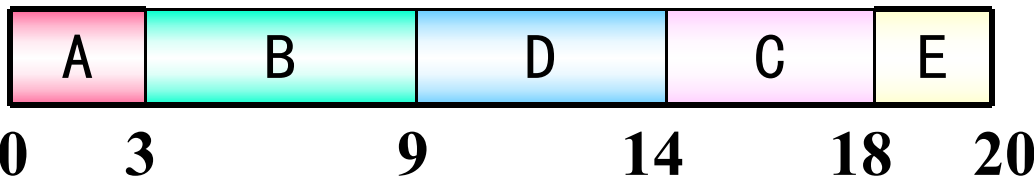
平均归一化周转时间： $(1+1.17+2.25+2.8+3.5)/5=2.14$

6、优先级

- 静态优先级
- 动态优先级
- 抢占
- 非抢占
- 约定：数小代表优先级高

非抢占优先级

进 程	到达时间	服务时间	优先级
A	0	3	3
B	2	6	2
C	4	4	4
D	6	5	1
E	8	2	5



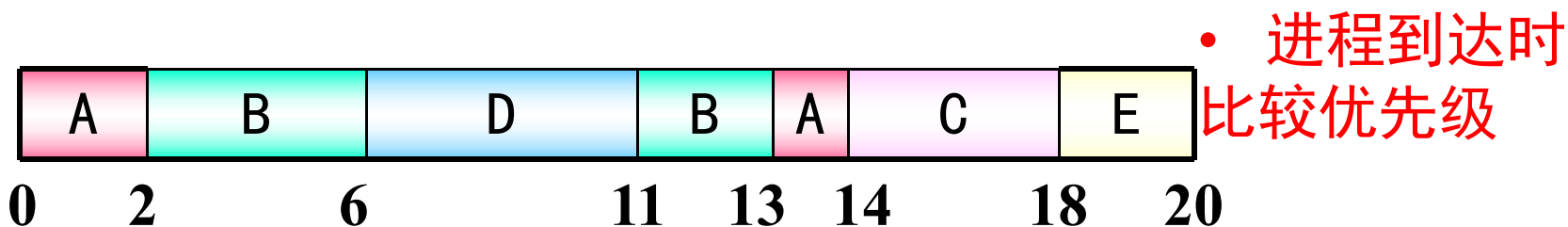
- 进程完成时再比较优先级

平均周转时间: $(3+7+14+8+12)/5=8.8$

平均归一化周转时间: $(1+1.17+3.5+1.6+6)/5=2.65$

抢占优先级

进 程	到达时间	服务时间	优先级
A	0	3	3
B	2	6	2
C	4	4	4
D	6	5	1
E	8	2	5



平均周转时间: $(14+11+14+5+12)/5=11.2$

平均归一化周转时间: $(4.67+1.83+3.5+1+6)/5=3.4$

7、反馈法

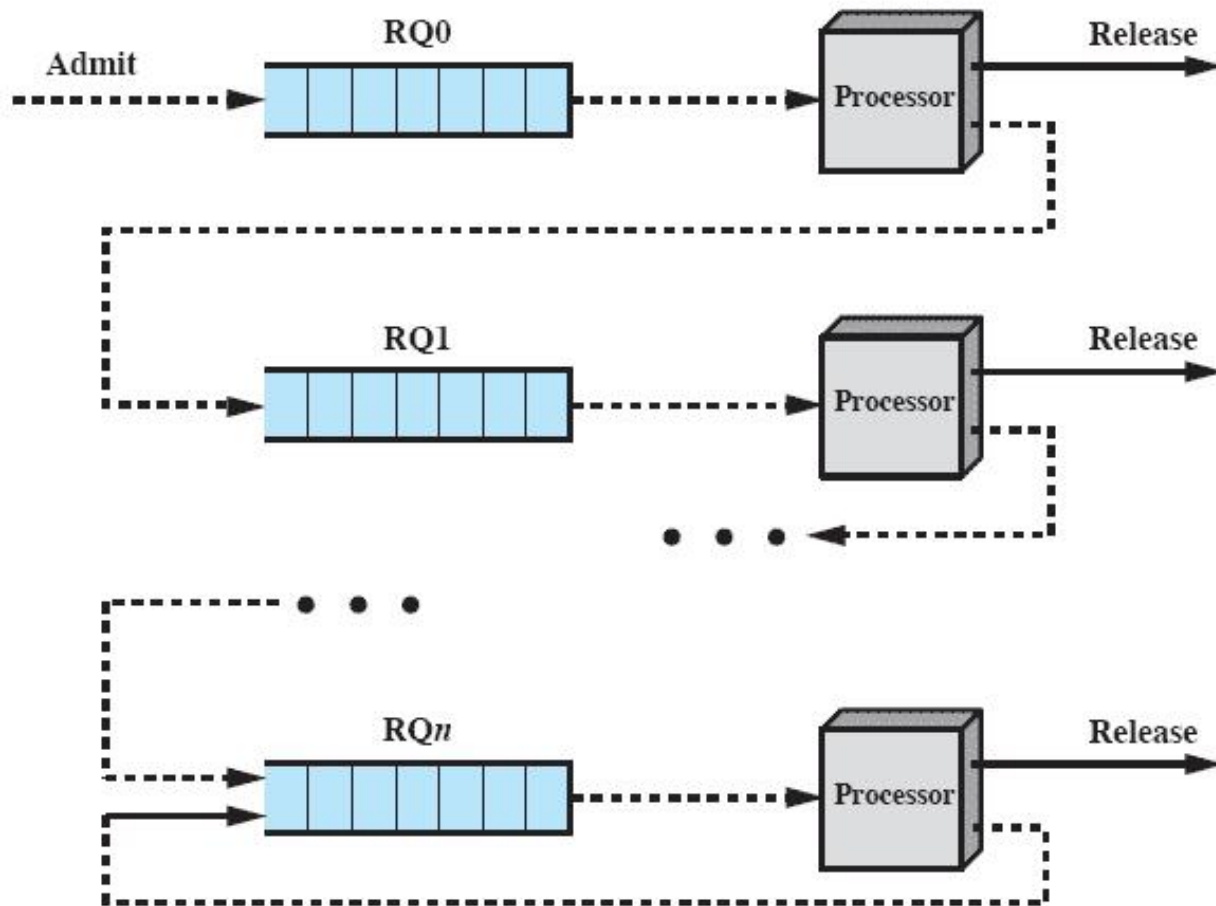


Figure 9.10 Feedback Scheduling

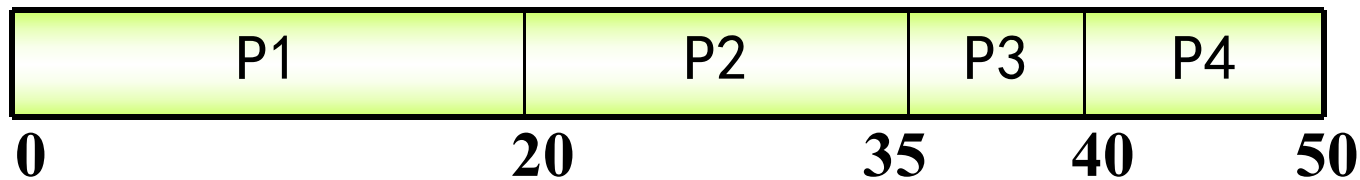
➤ 综合练习题

进程	到达时间	服务时间	优先级
P1	0	20	2
P2	5	15	3
P3	10	5	1
P4	15	10	4

FCFS、SPN、SRT、优先级（抢占、非抢占）、RR（时间片=5）、最高响应比优先

计算平均周转时间

1、FCFS



平均周转时间:

$$[20 + (35 - 5) + (40 - 10) + (50 - 15)] / 4 = 115 / 4 = 28.75$$

进程 到达时间 服务时间 优先级

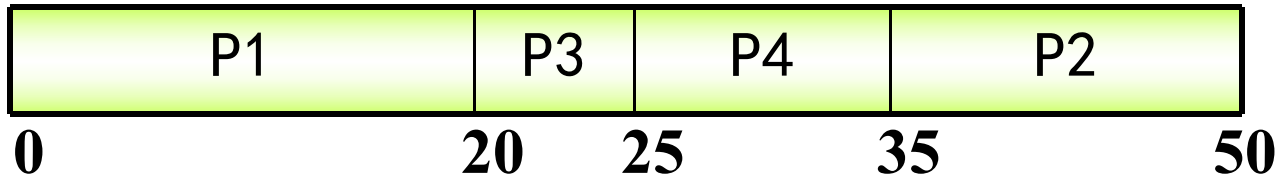
P1 **0** **20** **2**

P2 **5** **15** **3**

P3 **10** **5** **1**

P4 **15** **10** **4**

2、SPN



平均周转时间:

$$[20 + (50 - 5) + (25 - 10) + (35 - 15)] / 4 = 100 / 4 = 25$$

进程	到达时间	服务时间	优先级
----	------	------	-----

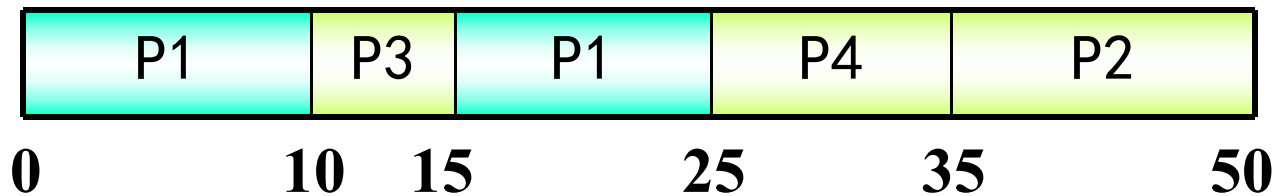
P1	0	20	2
----	---	----	---

P2	5	15	3
----	---	----	---

P3	10	5	1
----	----	---	---

P4	15	10	4
----	----	----	---

2、SRT



平均周转时间:

$$[25 + (50-5) + (15-10) + (35-15)] = 95/4 = 23.75$$

进程	到达时间	服务时间	优先级
----	------	------	-----

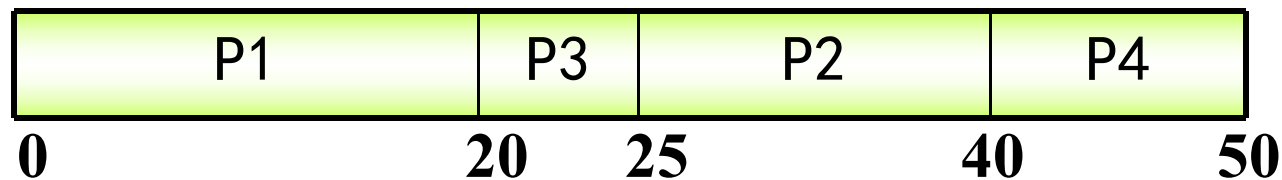
P1	0	20	2
----	---	----	---

P2	5	15	3
----	---	----	---

P3	10	5	1
----	----	---	---

P4	15	10	4
----	----	----	---

3、优先级（非抢占）



平均周转时间:

$$[20+(40-5) +(25-10)+(50-15)]/4=105/4=26.25$$

进程	到达时间	服务时间	优先级
----	------	------	-----

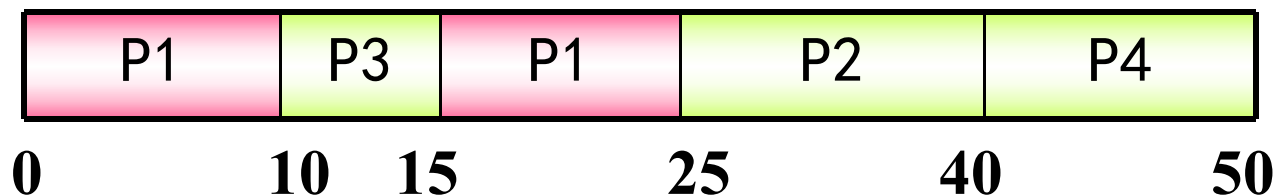
P1	0	20	2
----	---	----	---

P2	5	15	3
----	---	----	---

P3	10	5	1
----	----	---	---

P4	15	10	4
----	----	----	---

3、优先级（抢占）



平均周转时间:

$$[25 + (40-5) + (15-10) + (50-15)] = 100/4 = 25$$

进程	到达时间	服务时间	优先级
----	------	------	-----

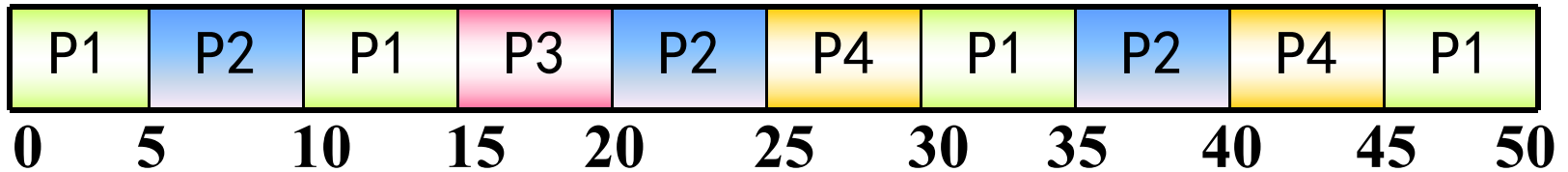
P1	0	20	2
----	---	----	---

P2	5	15	3
----	---	----	---

P3	10	5	1
----	----	---	---

P4	15	10	4
----	----	----	---

4、RR (5)



平均周转时间:

$$[50 + (40 - 5) + (20 - 10) + (45 - 15)] / 4 = 125 / 4 = 31.25$$

进程	到达时间	服务时间	优先级
----	------	------	-----

P1	0	20	2
----	---	----	---

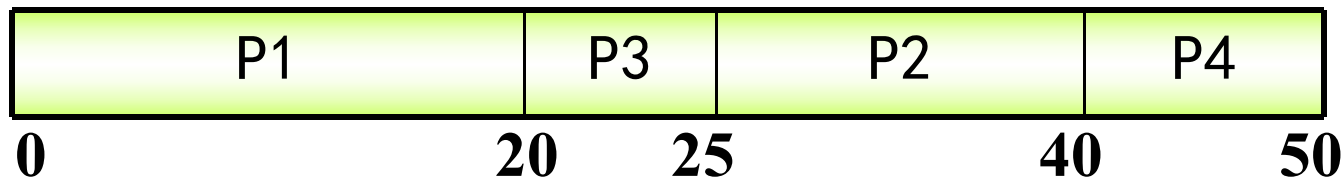
P2	5	15	3
----	---	----	---

P3	10	5	1
----	----	---	---

P4	15	10	4
----	----	----	---

5、最高响应比优先

3、最高响应比优先



平均周转时间:

$$[20+(40-5) + (25-10) + (50-15)]/4 = 105/4 = 26.25$$

进程	到达时间	服务时间	优先级
----	------	------	-----

P1	0	20	2
----	---	----	---

P2	5	15	3
----	---	----	---

P3	10	5	1
----	----	---	---

P4	15	10	4
----	----	----	---

P1运行完，计算P2P3P4的响应比：

$$R_{P2} = (15+15)/15 = 2$$

$$R_{P3} = (10+5)/5 = 3$$

$$R_{P4} = (5+10)/10 = 1.5$$

P3运行完，计算P2P4的响应比：

$$R_{P2} = (20+15)/15 = 2.3$$

$$R_{P4} = (10+10)/10 = 2$$

9.3 传统的UNIX调度

- 多级反馈
 - 每个优先级队列中使用轮转方法
 - 使用1秒抢占方式
 - 每秒都会重新计算每个进程的优先级

作业

- 复习题 9.1 9.5
- 习题 9.16 （要求计算平均周转时间和平均归一化周转时间）