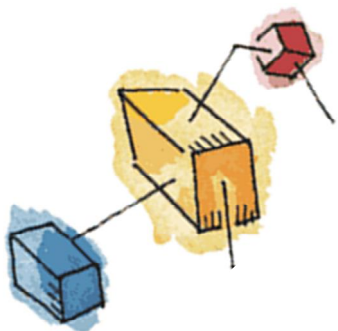


به نام خدا

فصل نهم:

زمان بندی تک پردازنده ای فرآیندها

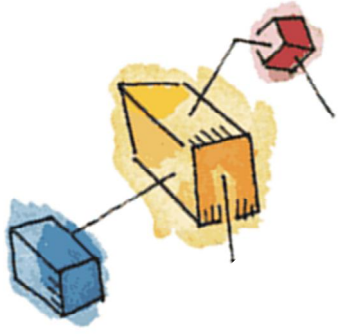
Uniprocessor Scheduling



سرفصل مطالب

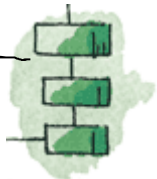
- انواع زمانبندی پردازنده
- الگوریتم های زمانبندی

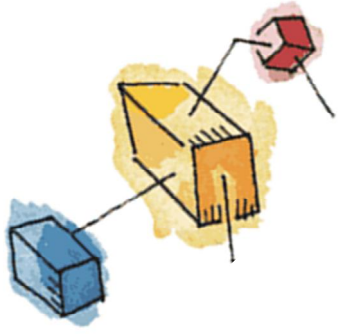




سرفصل مطالب

- انواع زمانبندی پردازنده
- الگوریتم های زمانبندی

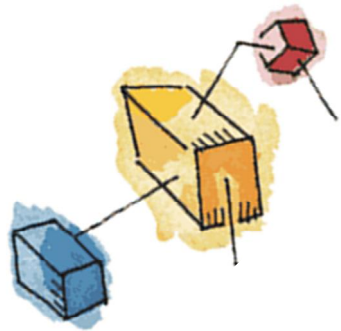




اهداف زمانبندی

- واگذاری فرآیندها به پردازنده(ها) برای اجرا.
- کاهش زمان پاسخ (Response time)
- بالا بردن توان عملیاتی (Throughput)
- افزایش بهره وری پردازنده (Processor efficiency)





انواع زمانبندی

- بلند مدت
- میان مدت
- کوتاه مدت
- ورودی/خروجی

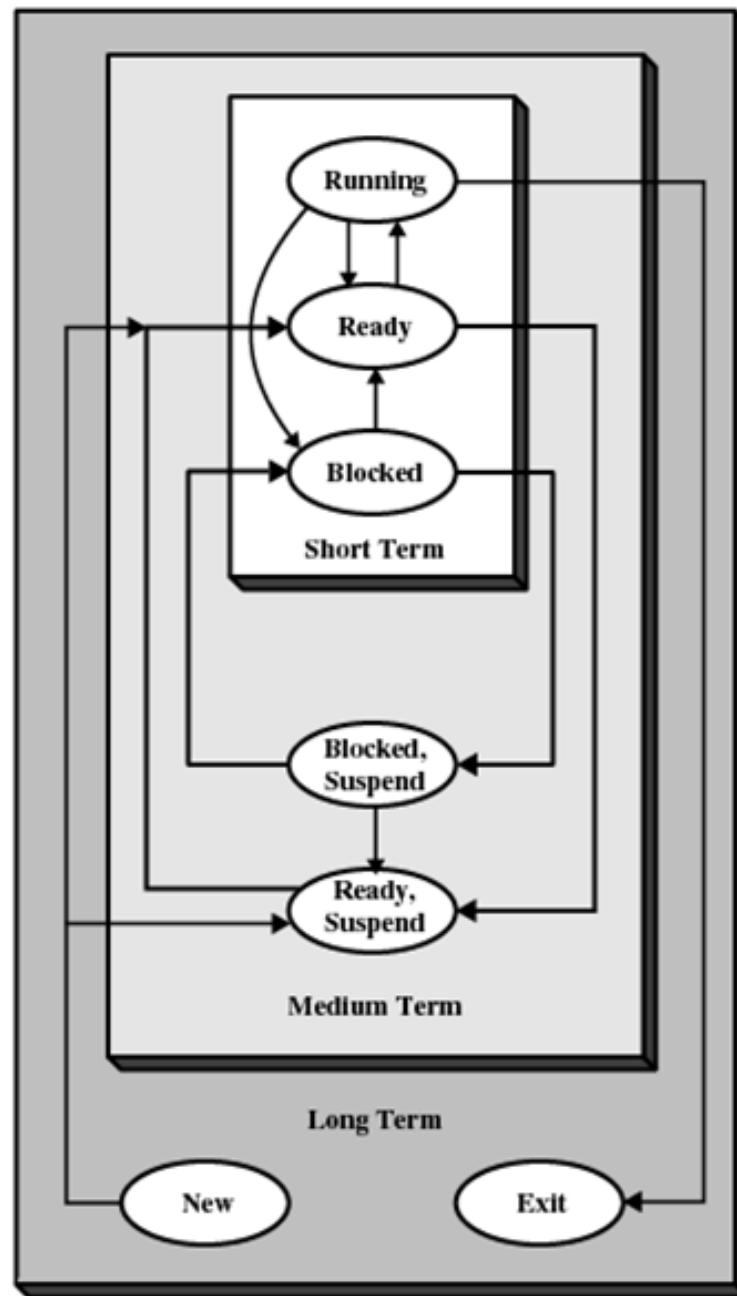


Figure 9.2 Levels of Scheduling



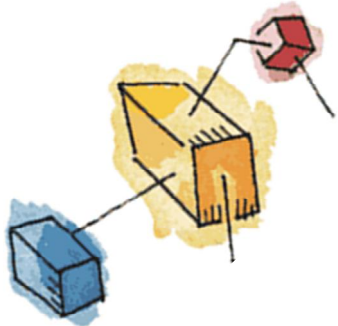
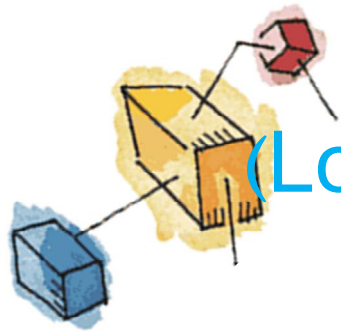


Table 9.1 Types of Scheduling

Long-term scheduling	The decision to add to the pool of processes to be executed
Medium-term scheduling	The decision to add to the number of processes that are partially or fully in main memory
Short-term scheduling	The decision as to which available process will be executed by the processor
I/O scheduling	The decision as to which process's pending I/O request shall be handled by an available I/O device

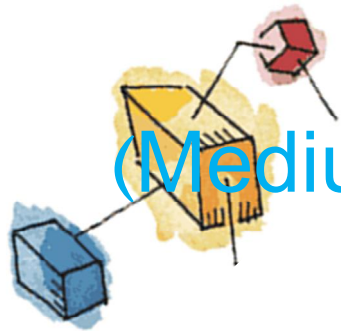




زمانبندی بلند مدت (Long-Term Scheduling)

- زمانبند کار (Job Scheduler)
- تعیین می کند کدام برنامه ها برای اجرا توسط سیستم پذیرفته شده اند.
- زمانبند بلند مدت درجه چند برنامه را در سیستم کنترل می کند.
- $\text{درجه چند برنامه} = \text{تعداد فرآیندهای موجود در حافظه اصلی}$
- اگر تعداد فرآیندها در حافظه اصلی زیاد باشد، هر فرآیند درصد کمتری از زمان را اجرا می شود.

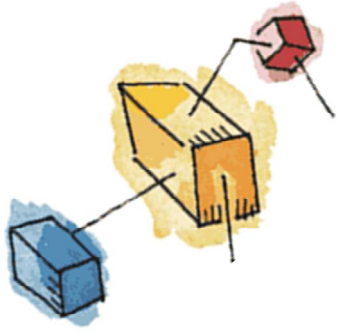




زمانبندی میان مدت (Medium-Term Scheduling)

- بخشی از عملیات مبادله (Swapping) است.
- مبادله فرآیندها بین حافظه اصلی و جانبی.
- بر اساس نیاز به مدیریت درجه چندبرنامگی.





زمانبندی کوتاه مدت (Short-Term Scheduling) یا زمانبندی پردازنده

- انتخاب فرایندها از صف آماده و تحویل به پردازنده
- زمانبند کوتاه مدت یا توزیع کننده (Dispatcher)
- بیش از سایر زمانبندها اجرا می شود
- بایستی بسیار سریع باشد
- در صورت وقوع هر یک از رویدادهای زیر احضار می شود:
 - وقفه های ساعت
 - وقفه های ورودی/خروجی
 - فراخوانی های سیستم عامل



و



صف های زمانبندی

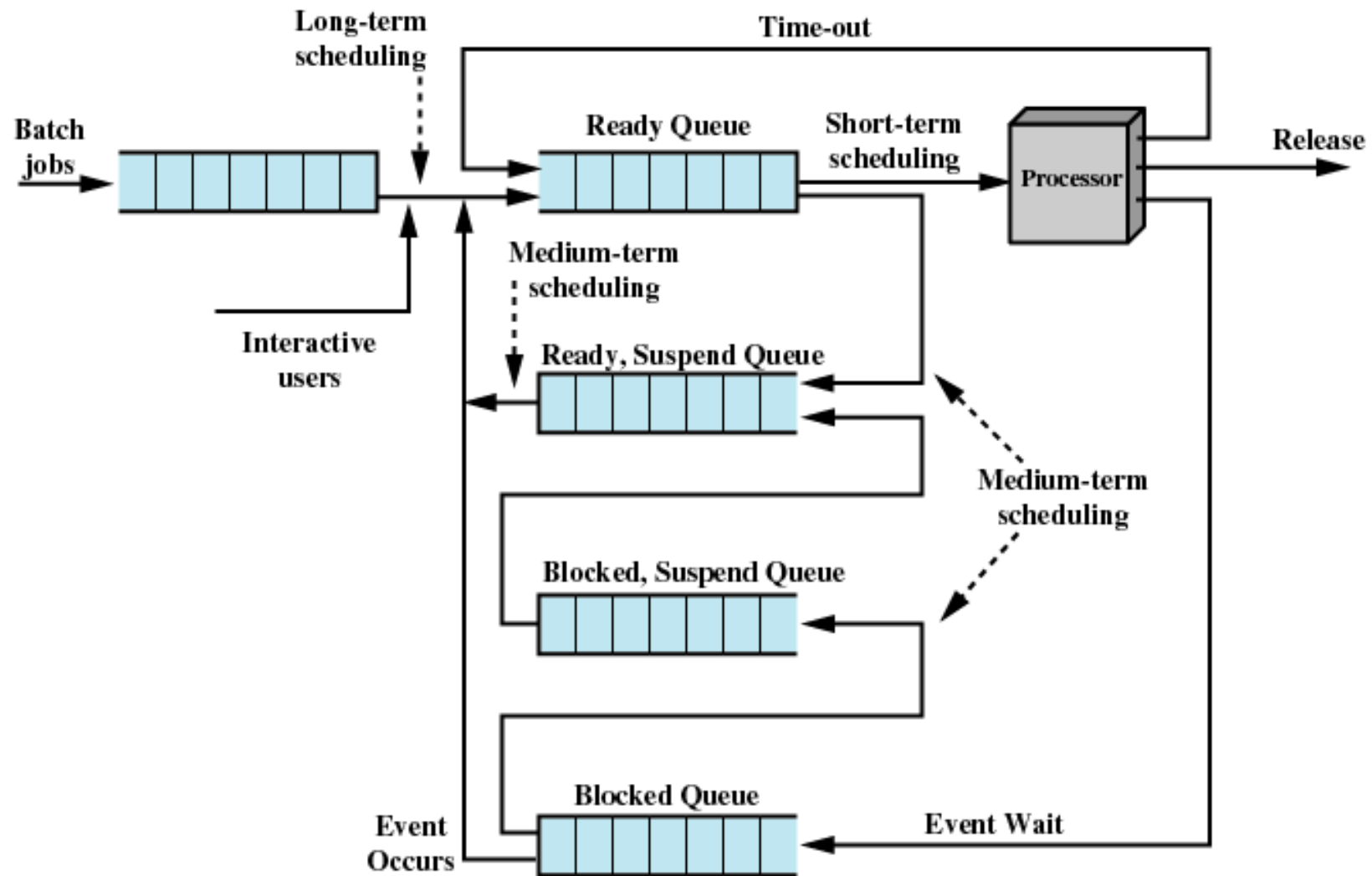
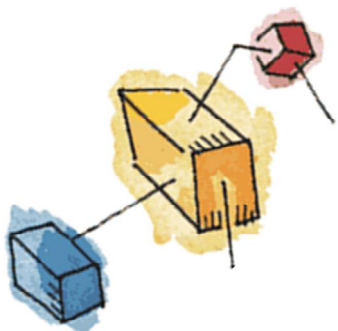


Figure 9.3 Queuing Diagram for Scheduling



معیارهای زمانبندی کوتاه مدت

- مبتنی بر کاربر

- زمان پاسخ

- زمان سپری شده از لحظه ارائه درخواست تا لحظه شروع پاسخ آن

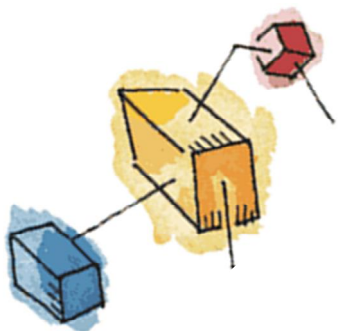
- زمان انتظار

- مدت زمانی که فرآیند در صف فرآیندهای آماده منتظر می ماند.

- زمان بازگشت (Turnaround)

- فاصله زمانی بین لحظه تحویل کار تا لحظه تکمیل آن





معیارهای زمانبندی کوتاه مدت

- مبتنی بر سیستم

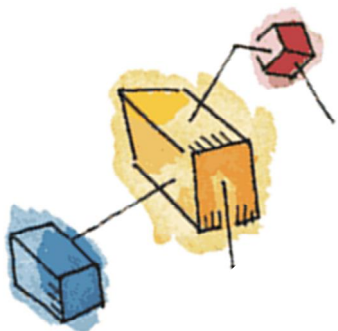
- بهره‌وری پردازنده

- درصد زمانی که پردازنده مشغول است.
- در سیستم‌های چند برنامه‌ای معیار مهمی است.

- توان عملیاتی

- مقدار کار انجام شده (به پایان رسیده) در واحد زمان
- وابسته به متوسط طول فرآیندها

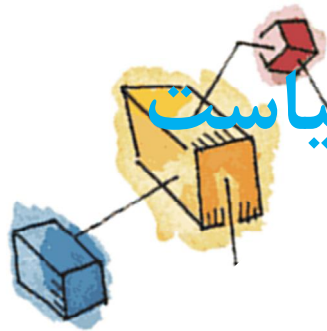




سرفصل مطالب

- انواع زمانبندی پردازنده
- الگوریتم های زمانبندی پردازنده





نوع تصمیم: انحصاری یا غیر انحصاری بودن سیاست زمانبندی فرآیندها

• انحصاری (non-preemptive)

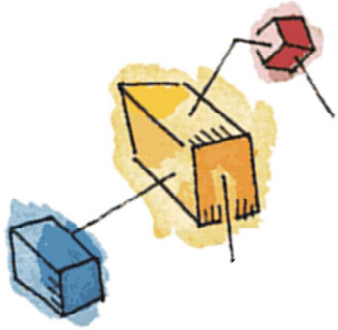
– هرگاه فرآیندی در وضعیت اجرا قرار گیرد، تا زمانی که خاتمه یافته و یا برای I/O خودش را بلاک (مسدود) کند در این حالت باقی خواهد ماند.

• غیر انحصاری (preemptive)

– فرآیندی که در حال اجرا می باشد، ممکن است توسط سیستم عامل دچار وقفه شده و به حالت آماده منتقل شود.

– امکان سرویس دهی بهتر را فراهم می آورد، چرا که یک فرآیند نمی تواند برای مدتی طولانی بصورت انحصاری وقت پردازنده را به خود اختصاص دهد.

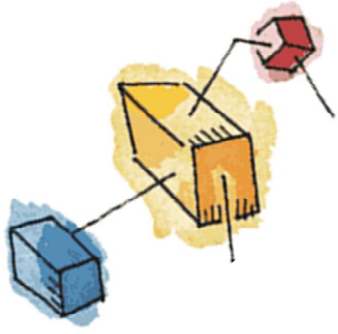




الگوریتم های زمانبندی پردازنده

- مبتنی بر اولویت (Priority)
- سرویس دهی به ترتیب ورود (First-Come-First-Served) یا FIFO یا FCFS
- نوبت گردشی (Round-Robin) یا RR
- کوتاه ترین زمان اجرا (Shortest Process Next) یا SPN یا SJF
- کمترین زمان اجرای باقیمانده (Shortest Remaining Time) یا SRT
- بالاترین نسبت زمان پاسخ دهی (Highest Response Ratio Next) یا HRRN
- بازخورد (Feedback)



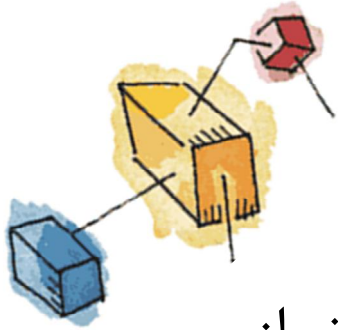


مثالی برای زمانبندی فرآیندها

Table 9.4 Process Scheduling Example

Process	Arrival Time	Service Time
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2





الگوریتم مبتنی بر اولویت (Priority)

- زمانبند، فرآیندی را برای اجرا انتخاب می کند که اولویتی بالاتر از فرآیندهای دیگر دارد.
- در این الگوریتم، چندین صف آماده وجود دارد که سطوح مختلف اولویت را نشان می دهند.
- فرآیندهای با اولویت پایین تر، ممکن است دچار مشکل قحطی زدگی شوند.
- برای مدیریت این مشکل، اولویت فرآیند می تواند بر اساس سن فرآیند (روش aging یا سالخوردگی) یا سابقه اجرای آن افزایش یابد.



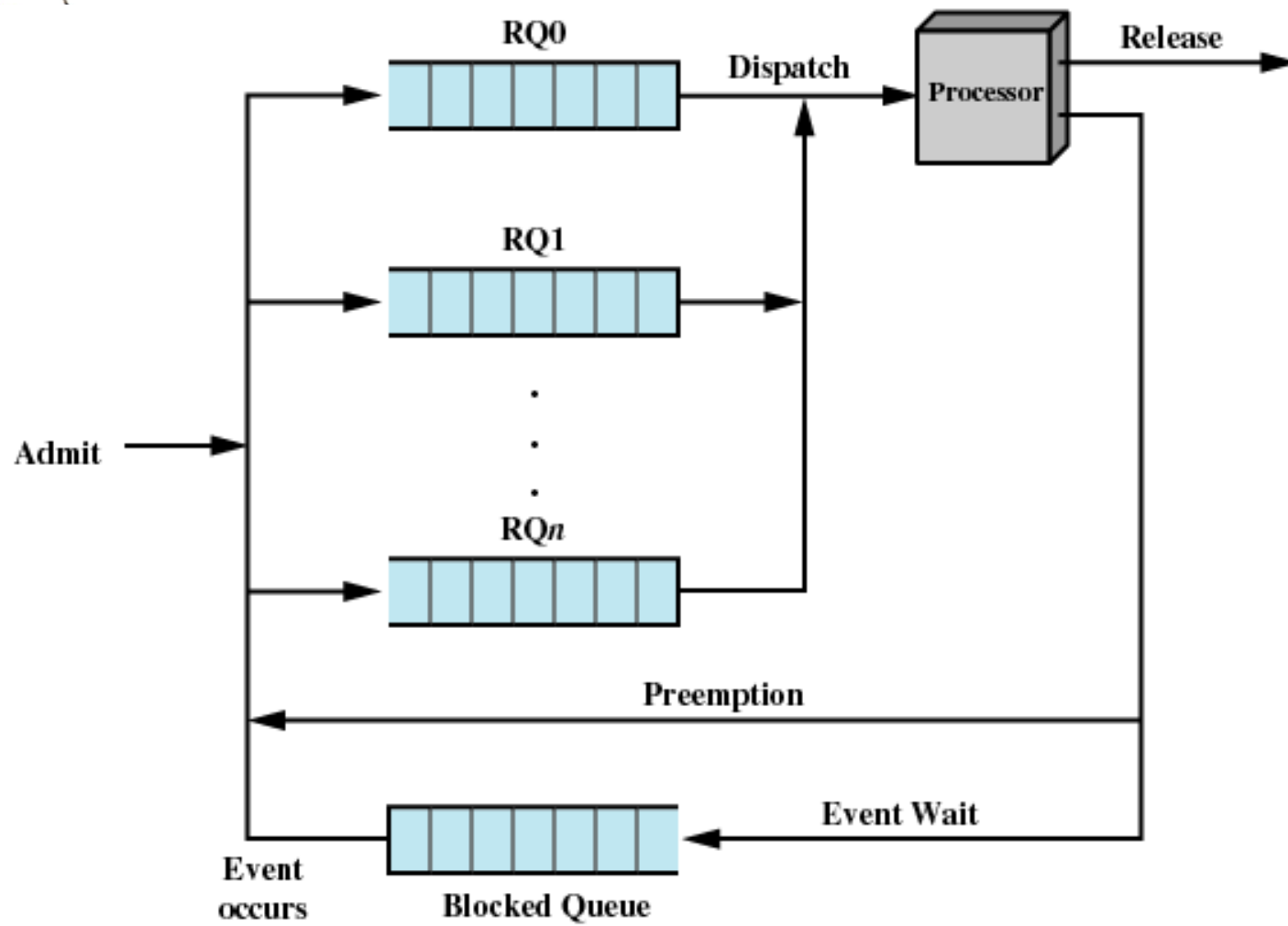
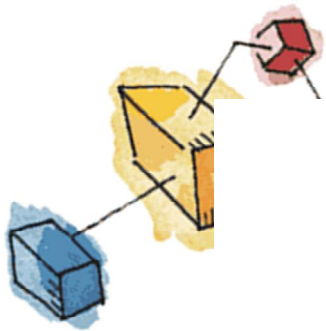
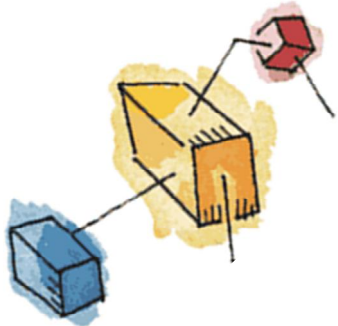


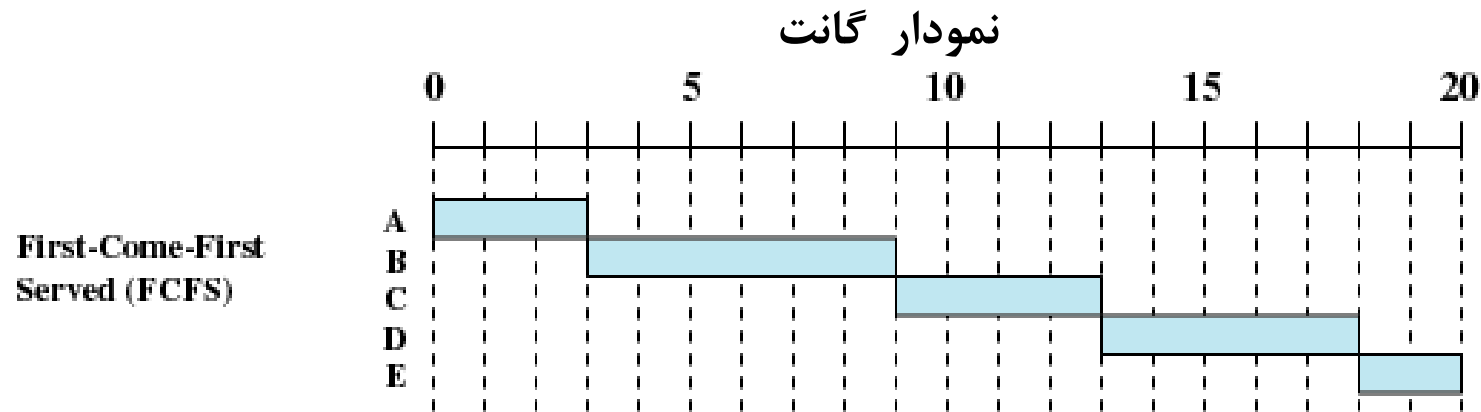
Figure 9.4 Priority Queuing





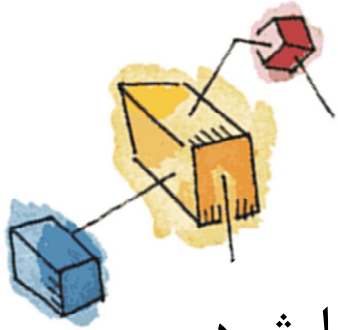
سرویس دهی به ترتیب ورود (FCFS/FIFO)

- هر فرآیند جدید به صف آماده اضافه می شود.
- هر زمان که فرآیند جاری از حالت اجرا خارج شود، قدیمی ترین فرآیند از صف آماده انتخاب می شود تا پس از آن انتخاب شود.
- زمانبندی انحصاری



Process	Arrival Time	Service Time
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

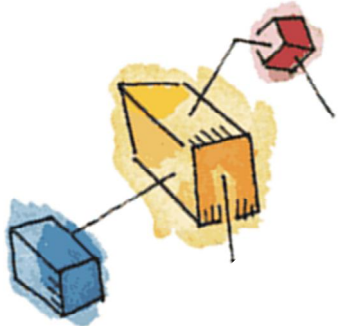




سرویس دهی به ترتیب ورود (FCFS/FIFO)

- یک فرآیند کوتاه ممکن است زمان زیادی منتظر بماند تا بتواند اجرا شود.
- به نفع فرآیندهای با تنگنای پردازنده (یا مقید به پردازنده یا CPU-bound) بوده و به ضرر فرآیندهای با تنگنای I/O می باشد.
- در سرویس دهی به ترتیب ورود، فرآیندهای مقید به I/O ناچار هستند منتظر بمانند تا اجرای فرآیندهای مقید به CPU تکمیل شود.
- فرآیندهای با تنگنای پردازنده، فرآیندهایی هستند که مدت زمان نیازشان به پردازنده خیلی بیشتر از مدت زمان نیاز آنها به I/O است.
- متوسط زمان انتظار برای این الگوریتم معمولاً طولانی است.

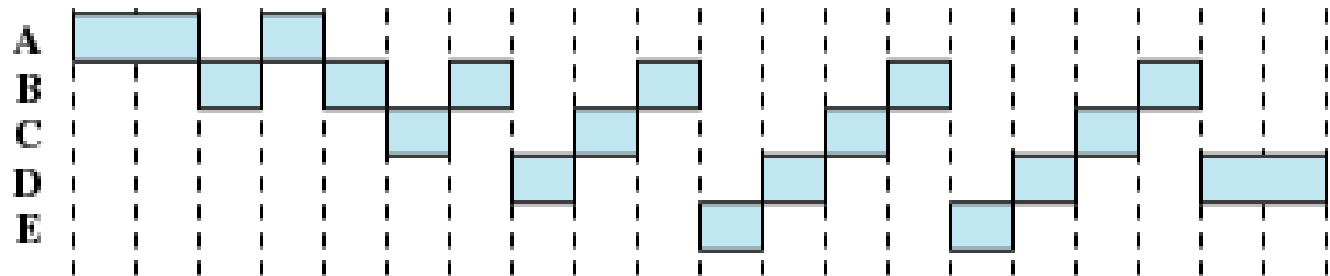




نوبت گردشی (Round-Robin)

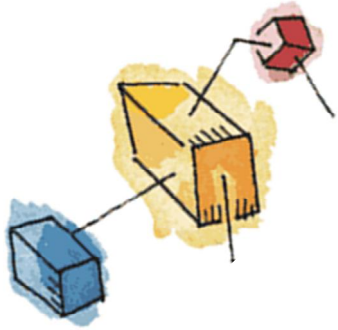
- صف فرآیندهای آماده بصورت یک صف حلقوی در نظر گرفته می شود.
- وقت پردازنده به برش های زمانی با یک طول مشخص تقسیم شده و هر فرآیند در صف آماده که نوبتش رسیده باشد می تواند به میزان یک برش زمانی (یا کوانتوم زمانی) اجرا شود.

Round-Robin
(RR), $q = 1$



Process	Arrival Time	Service Time
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2



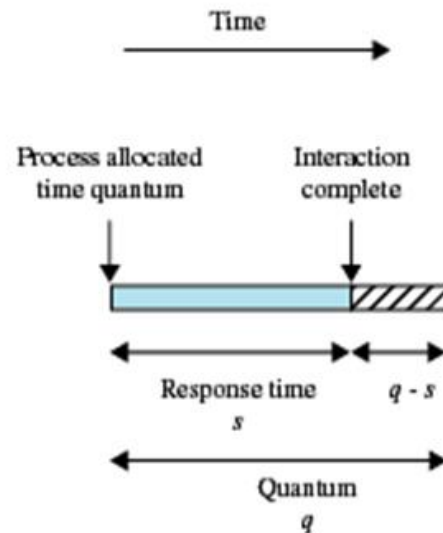
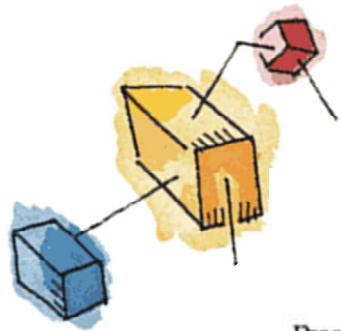


نوبت گردشی (Round-Robin)

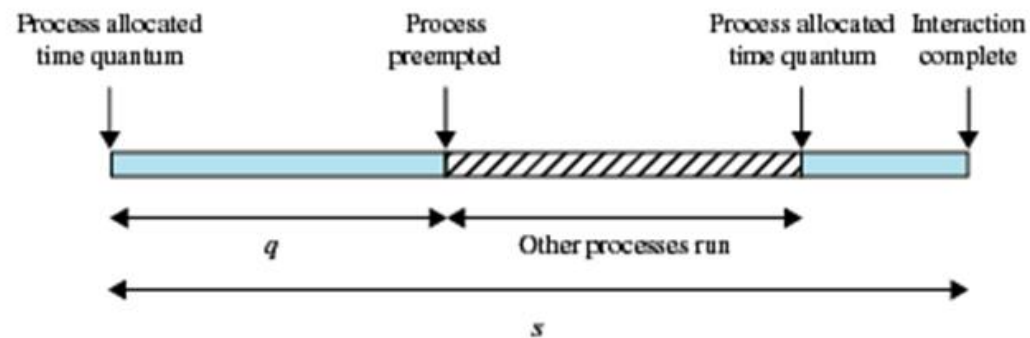
- وقفه های ساعت در بازه های زمانی متناوبی صادر می شوند.
- با اتمام برش زمانی، وقفه ای صادر شده و کنترل پردازنده از فرآیند جاری گرفته می شود و به فرآیند بعدی در صف آماده انتقال می یابد.
- روش نوبتی-گردشی، یک سیاست زمانبندی غیرانحصاری است که در آن، بر اساس زمان ساعت، فرآیندها قبضه می شوند.
- مناسب سیستم های اشتراک زمانی



در یک سیستم اشتراک زمانی:

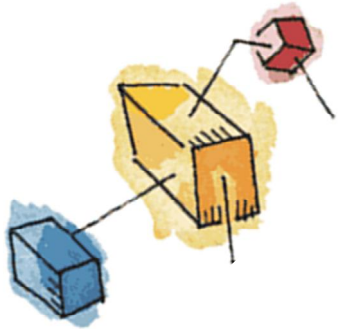


(a) Time quantum greater than typical interaction



(b) Time quantum less than typical interaction

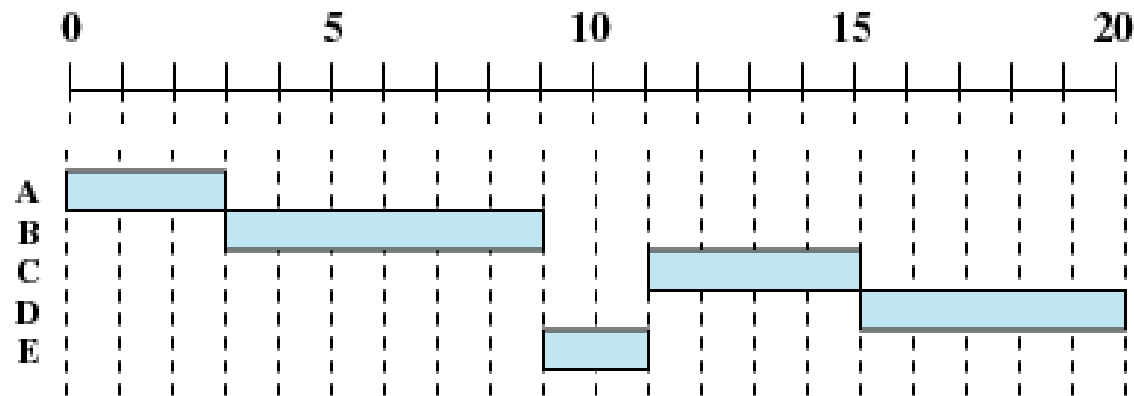




کوتاه ترین زمان اجرا (SJF یا SPN)

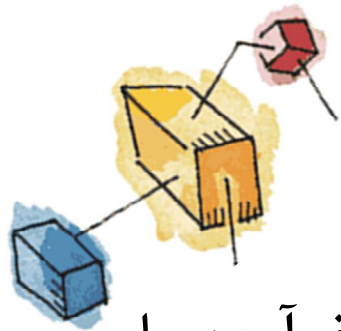
- یک سیاست زمانبندی انحصاری.
- هر بار فرآیندی که کوتاه ترین زمان پردازش را دارد انتخاب می شود.
- نیازمند تخمین زمان پردازش
- فرآیند کوتاه از فرآیندهای بلندتر پیشی می گیرد.
- احتمال قحطی زدگی برای فرآیندهای طولانی تر وجود دارد.

Shortest Process
Next (SPN)



Process	Arrival Time	Service Time
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2





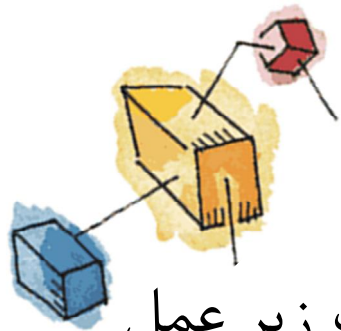
کوتاه ترین زمان اجرا (SJF یا SPN)

- در این روش نیاز است الگوریتم از زمان های پردازش فرآیندها اطلاع داشته باشد یا آنها را تخمین بزند.
- ساده ترین راه برای محاسبه: میانگین گیری

$$S_{n+1} = \frac{1}{n} \sum_{i=1}^n T_i$$

- T_i : زمان اجرای فرآیند برای i امین نوبت
- S_i : زمان پیش بینی شده برای i امین نوبت





کوتاه ترین زمان اجرا (SPN یا SJF)

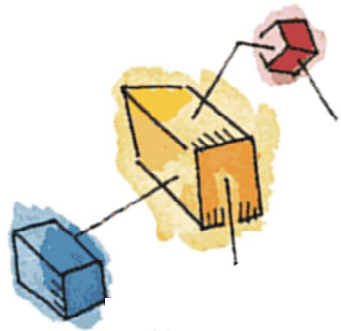
- برای اجتناب از انجام کل محاسبه در هر بار، می توان به صورت زیر عمل کرد:

$$S_{n+1} = \frac{1}{n}T_n + \frac{n-1}{n}S_n$$

- این رابطه به همه نوبت ها وزن یکسان می دهد. در حالیکه نوعا مایل هستیم به نوبت های اخیر، وزن بیشتری بدهیم زیرا احتمال اینکه رفتار آینده را منعکس کنند بیشتر است. پس به جای میانگین گیری ساده، از متوسط نمایی استفاده می کنیم.

$$S_{n+1} = \alpha T_n + (1 - \alpha)S_n \quad (0 < \alpha < 1)$$





کوتاه ترین زمان اجرا (SPN یا SJF)

$$S_{n+1} = \alpha T_n + (1 - \alpha) S_n \quad (0 < \alpha < 1)$$

- با باز کردن رابطه، داریم:

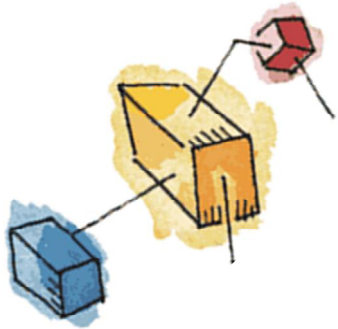
$$S_{n+1} = \alpha T_n + (1 - \alpha) \alpha T_{n-1} + \dots + (1 - \alpha)^i \alpha T_{n-i} + \dots + (1 - \alpha)^n S_1$$

- با توجه به اینکه α و $1-\alpha$ هر دو کوچکتر از ۱ هستند، ضریب جمله به تدریج کوچکتر می شود. مثلاً اگر $\alpha=0.8$ داریم:

$$S_{n+1} = 0.8 T_n + 0.16 T_{n-1} + 0.032 T_{n-2} + 0.0064 T_{n-3} + \dots$$

پس مشاهدات قدیمی تر، اثر کمتری در محاسبه متوسط دارند.





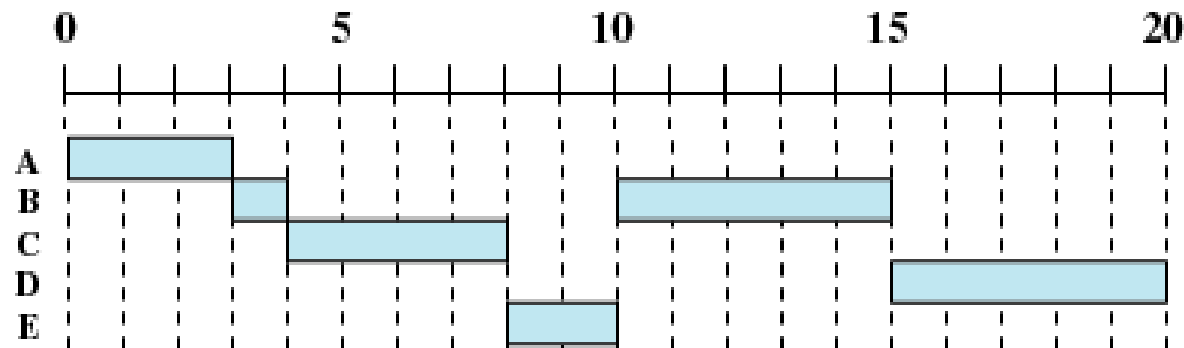
کمترین زمان اجرای باقیمانده (SRT)

• شبیه سیاست SPN (یا SJF) اما غیرانحصاری

– اگر فرآیندی وجود دارد که زمان اجرای آن از زمان باقیمانده فرآیند جاری کمتر است، فرآیند جاری متوقف می شود و اجرای آن فرآیند شروع می شود

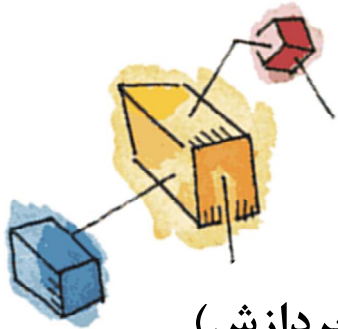
• نیازمند تخمین زمان پردازش

Shortest Remaining Time (SRT)



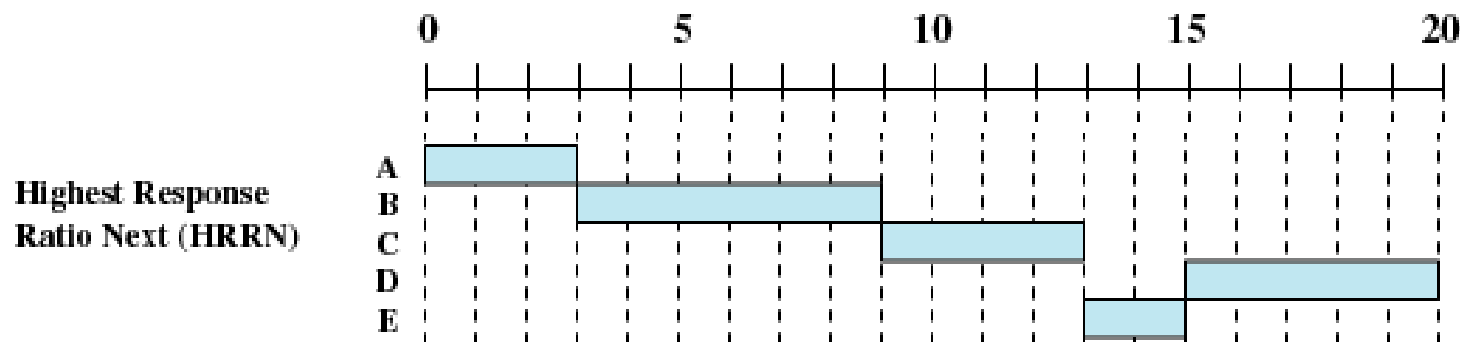
Process	Arrival Time	Service Time
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2





بالاترین نسبت زمان پاسخ دهی (HRRN)

- مدت زمان مورد نیاز برای پردازش / (مدت زمان انتظار + مدت زمان مورد نیاز برای پردازش)
- هر بار فرآیندی که بزرگترین نسبت را دارد برای پردازش انتخاب می شود.
- این روش، یک سیاست زمانبندی انحصاری (non-preemptive) می باشد.

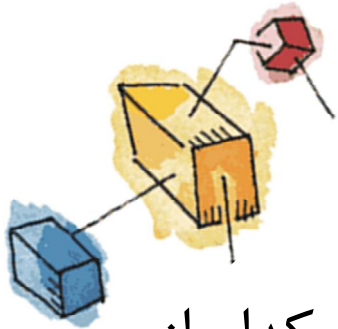


Process	Arrival Time	Service Time
A	0	3
B	2	6
C	4	4
D	6	5
E	8	2

$$\frac{\text{time spent waiting} + \text{expected service time}}{\text{expected service time}}$$



بازخورد (Feedback)



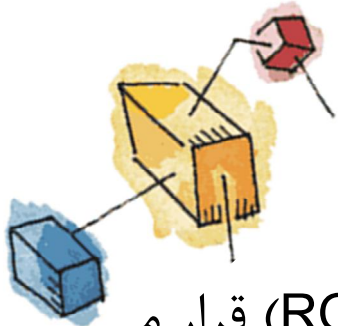
- اگر هیچ تخمینی از زمان پردازش فرآیندها نداشته باشیم، هیچکدام از سه روش قبل (SPN و SRT و HRRN) قابل استفاده نیستند.
– این سه روش، هر کدام به نوعی به کارهای کوچکتر اولویت می دادند.

- راه دیگر برای اولویت دادن به کارهای کوچکتر، جریمه کردن کارهایی است که زمان طولانی تری به اجرا درآمده اند.

- در واقع روی زمان سپری شده تمرکز می کنیم.

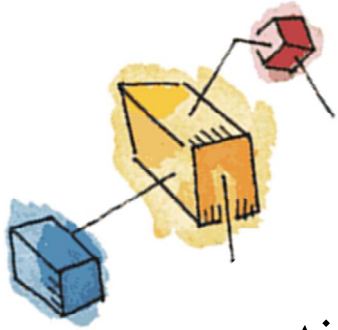


بازخورد (Feedback)



- هر فرآیند در ابتدا که وارد سیستم می شود، در صف با اولویت بالاتر (RQ0) قرار می گیرد.
- هر فرآیند پس از اینکه یک دور در صف RQ0 به اجرا درآمد، به ته صف RQ1 منتقل می شود.
- هر فرآیند پس از اینکه یک دور در صف RQ1 به اجرا درآمد، به ته صف RQ2 منتقل می شود.
- و
- اجرای فرآیندهای کوتاه به سرعت به انتها می رسد بدون اینکه به صف های پایین منتقل شوند.
- فرآیندهای بلندتر به تدریج به صف های پایین تر منتقل می شوند و در صف آخر می مانند.
- بنابراین فرآیندهای جدیدتر و کوتاه تر، به فرآیندهای طولانی تر اولویت می یابند و فرآیندهایی که زمان طولانی تری در حال اجرا بوده اند، جریمه می شوند.
- نیازی به اطلاع و تخمین باقیمانده زمان اجرای مورد نیاز فرآیند ندارد.





بازخورد (Feedback)

- فرآیندهای صف های پایین تر ممکن است دچار گرسنگی شوند.
- می توان پس از مدتی آنها را به صف های بالاتر منتقل کرد.
- هر صف زمان بندی خود را دارد. مثلاً نوبتی گردش یا FCFS
- در زمان بندی نوبتی گردش، می توان کوانتوم زمانی صف های با اولویت بالاتر را کوچکتر در نظر گرفت و برای صف های بعدی، آن را افزایش داد.



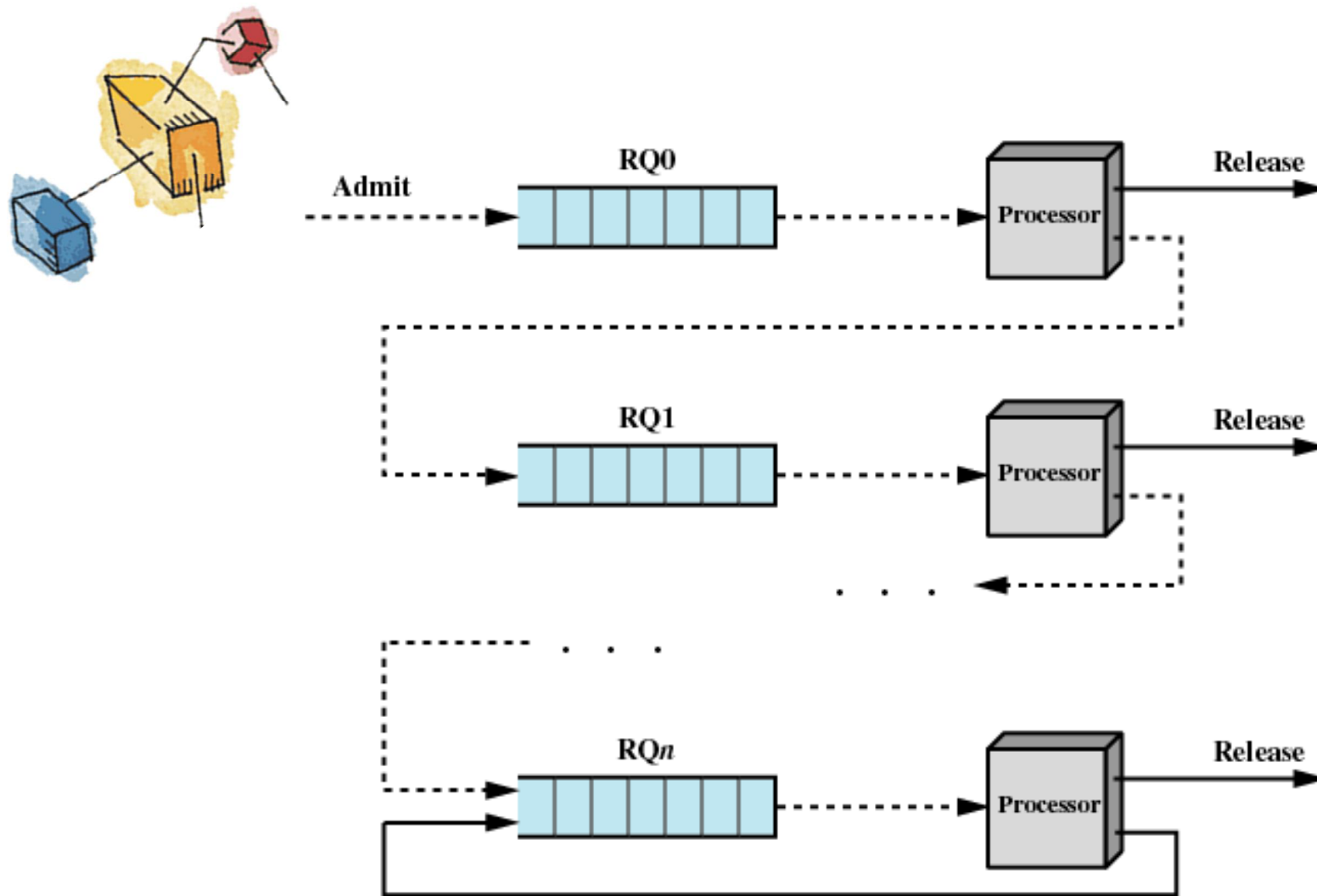
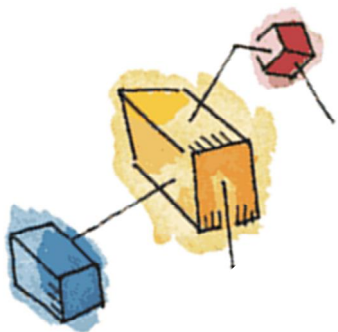


Figure 9.10 Feedback Scheduling





پایان فصل نهم

