Scheduling

Scheduling processes

- OS keeps PCBs, TCBs on different queues
  - Ready processes are on ready queue-OS chooses one to dispatch
  - Processes waiting for I/O are on appropriate device queue
- OS regulates PCB migration during life cycle of corresponding process

Why scheduling is interesting

- Processes are not created equal!

  CPU-bound process: long CPU bursts

  mp3 encoding, compilation, scientific applications

  I/O-bound process: short CPU bursts

  index a file system, browse small web pages

  Balanced

  playing video, moving windows around

Metrics

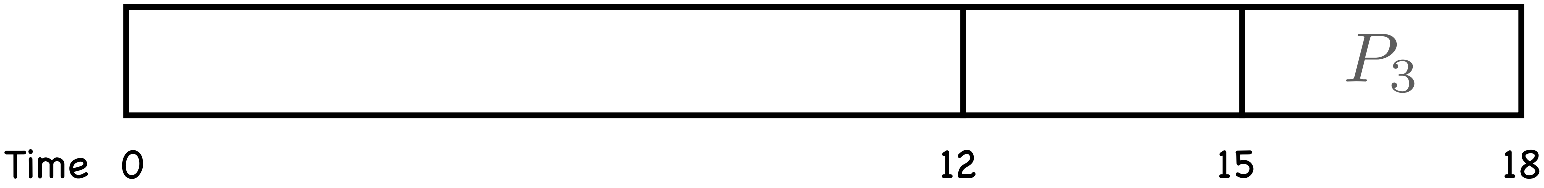
- CPU utilization: time CPU busy/time observed ↑
- Throughput: jobs completed/time observed ↑
- Turnaround time: time elapsed between submission and termination ↓
- Waiting time: time spent ready but not running ↓
- Response time: time elapsed between ready and first response produced ↓

FIFO

FIFO

- Processes with compute time 12, 3, and 3

Job arrival

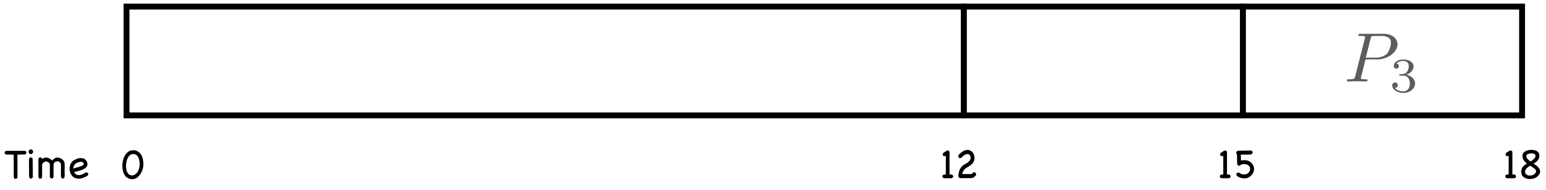


Average
Response Time:
(12+15+18)/3 = 15

FIFO

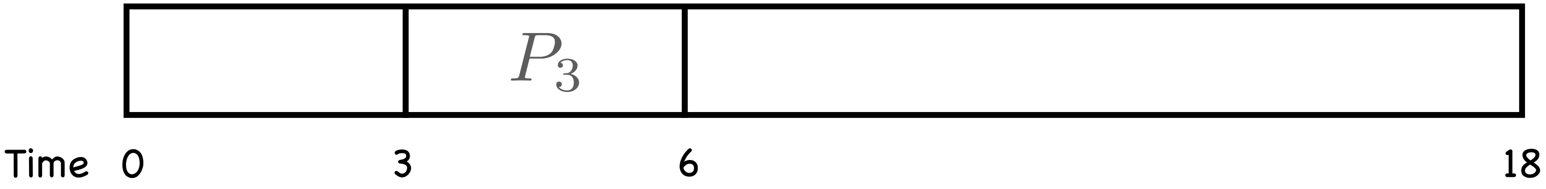
- Processes with compute time 12, 3, and 3

Job arrival



Average
Response Time:
(12+15+18)/3 = 15

Job arrival



Average
Response Time:
(3+6+18)/3 = 9

- Average response time very sensitive to arrival time

FIFO

Simple

Minimizes overhead

No starvation

Optimal average response time
(when all tasks are same size)

Poor average response time
when tasks have variable size

Not responsive to
interactive tasks

Round Robin

- Each process is allowed to run for a quantum
- Context is switched (at the latest) at the end of the quantum
- What is a good quantum size?
  - Too long, and it morphs into FIFO
  - Too short, and much time lost context switching
  - Typical quantum: about 100X cost of context switch

Round Robin

No starvation

Can reduce response time

Overhead of context switching

Mix of I/O and CPU bound

Simultaneous, equal length jobs

RR & FIFO Example

5 jobs, 100s each, quantum 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100				
2	100				
3	100				
4	100				
5	100				
Average					

RR & FIFO Example

5 jobs, 100s each, time sliced 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100	100		0	
2	100				
3	100				
4	100				
5	100				
Average					

RR & FIFO Example

5 jobs, 100s each, time sliced 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100	100		0	
2	100	199		99	
3	100				
4	100				
5	100				
Average					

RR & FIFO Example

5 jobs, 100s each, time sliced 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100	100		0	
2	100	199		99	
3	100	298		198	
4	100	397		297	
5	100	496		396	
Average					

RR & FIFO Example

5 jobs, 100s each, time sliced 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100	100		0	
2	100	199		99	
3	100	298		198	
4	100	397		297	
5	100	496		396	
Average		298		198	

RR & FIFO Example

5 jobs, 100s each, time sliced 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100	100	496	0	396
2	100	199		99	
3	100	298		198	
4	100	397		297	
5	100	496		396	
Average		298		198	

RR & FIFO Example

5 jobs, 100s each, time sliced 1s

Context switch time = 0; jobs arrive at 0,1,2,3,4

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	100	100	496	0	396
2	100	199	497	99	396
3	100	298	498	198	396
4	100	397	499	297	396
5	100	496	500	396	396
Average		298	498	198	396

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50				
2	40				
3	30				
4	20				
5	10				
Average					

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40				
3	30				
4	20				
5	10				
Average					

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40	90		40	
3	30				
4	20				
5	10				
Average					

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40	90		50	
3	30	120		90	
4	20	140		120	
5	10	150		140	
Average		110		80	

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40	90		50	
3	30	120		90	
4	20	140		120	
5	10	150	50	140	40
Average		110		80	

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40	90		50	
3	30	120		90	
4	20	140	90	120	70
5	10	150	50	140	40
Average		110		80	

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40	90		50	
3	30	120	120	90	90
4	20	140	90	120	70
5	10	150	50	140	40
Average		110		80	

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50		0	
2	40	90	140	50	100
3	30	120	120	90	90
4	20	140	90	120	70
5	10	150	50	140	40
Average		110		80	

RR & FIFO Example 2

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Turnaround Time		Wait Time	
Job	Length	FCFS	Round Robin	FCFS	Round Robin
1	50	50	150	0	100
2	40	90	140	50	100
3	30	120	120	90	90
4	20	140	90	120	70
5	10	150	50	140	40
Average		110	110	80	80

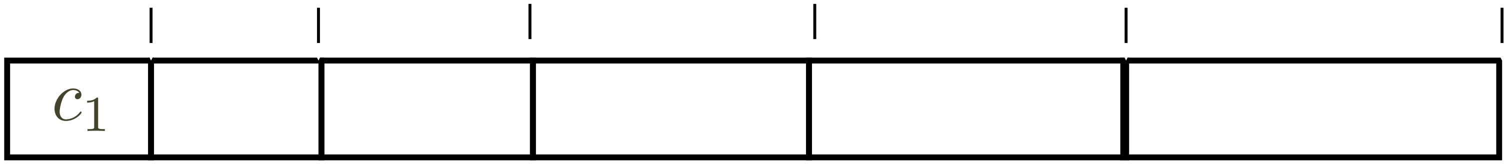
SJF: Shortest Job First

- Schedule jobs in order of estimated completion† time
- Optimal* average turnaround time (   )

†with preemption, remaining time *when jobs are available simultaneously

SJF: Shortest Job First

- Schedule jobs in order of estimated completion[†] time
- Optimal[*] average turnaround time (   )
- Intuition:



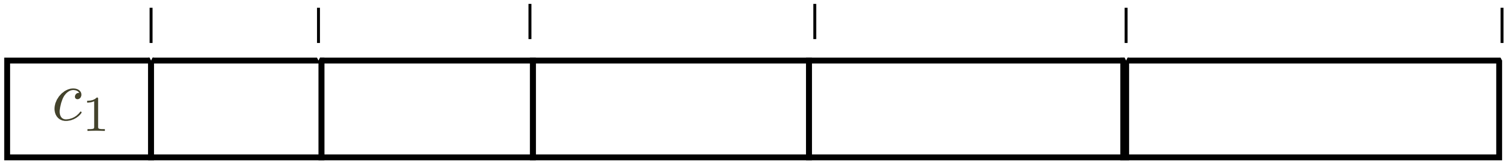
- Can switching execution order reduce response time?

[†] with preemption, remaining time

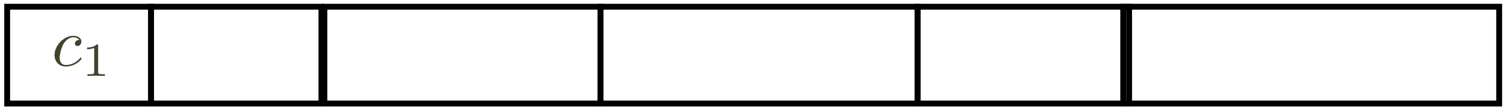
[*] when jobs are available simultaneously

SJF: Shortest Job First

- Schedule jobs in order of estimated completion[†] time
- Optimal* average turnaround time (   )
- Intuition:



- Can switching execution order reduce response time?



[†] with preemption, remaining time   *when jobs are available simultaneously

SJF

Optimal average response time

Pessimal variance in response time

Needs estimate of execution time

Can starve long jobs

Shortest Process Next (SJF for interactive jobs)

- Enqueue in order of estimated completion time

  Use recent history as indicator of near future

- Let duration of CPU burst

  estimated duration of CPU burst

- then $\tau_{n+1} = \alpha t_n + (1 - \alpha)\tau_n$

  which expands to

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50						
2	40						
3	30						
4	20						
5	10						
Average							

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50						
2	40						
3	30						
4	20						
5	10			10			0
Average							

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50						
2	40						
3	30						
4	20			30			10
5	10			10			0
Average							

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50						
2	40						
3	30			60			30
4	20			30			10
5	10			10			0
Average							

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50						
2	40			100			60
3	30			60			30
4	20			30			10
5	10			10			0
Average							

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50			150			100
2	40			100			60
3	30			60			30
4	20			30			10
5	10			10			0
Average				70			40

SJF Example

5 jobs of length 50s, 40s, 30s, 20s, 10s, time sliced 1s

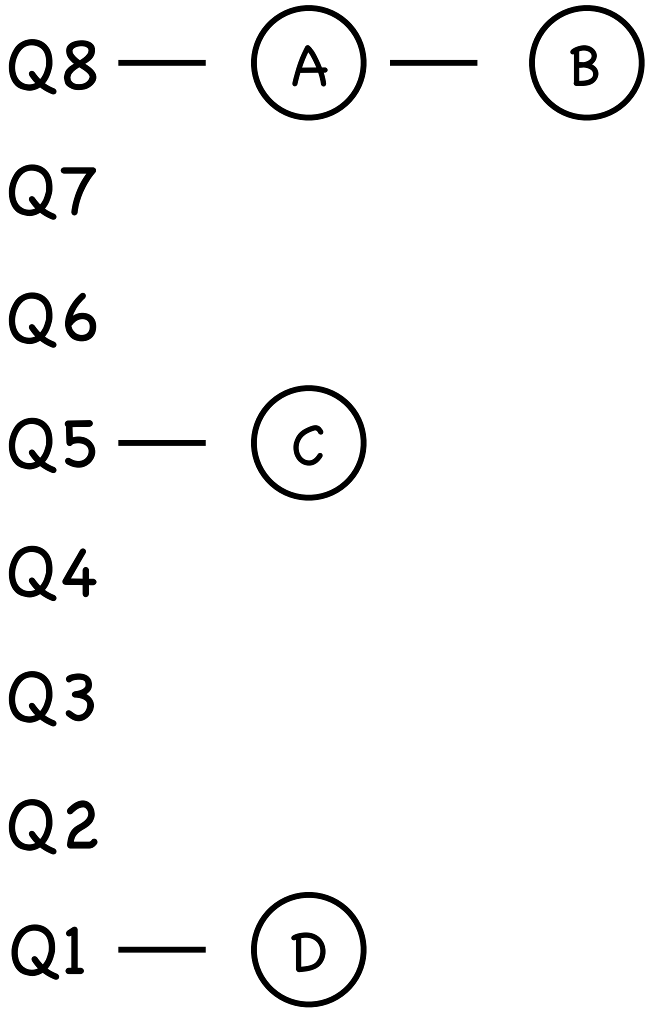
Context switch time = 0; Jobs arrive simultaneously

		Completion Time			Wait Time		
Jo	Leng	FCF	RR	SJF	FCF	RR	SJF
1	50	50	150	150	0	100	100
2	40	90	140	100	50	100	60
3	30	120	120	60	90	90	30
4	20	140	90	30	120	70	10
5	10	150	50	10	140	40	0
Average		110	110	70	80	396	40

Multi-level Feedback Queues

- Use the past to predict the future
  - If was I/O bound in the past, it is likely to be so in the future
  - Approximates SJF without calling 1-900
- Favors jobs that used little CPU...
  - ...but adaptive

The basic idea



Different queues at different priority levels

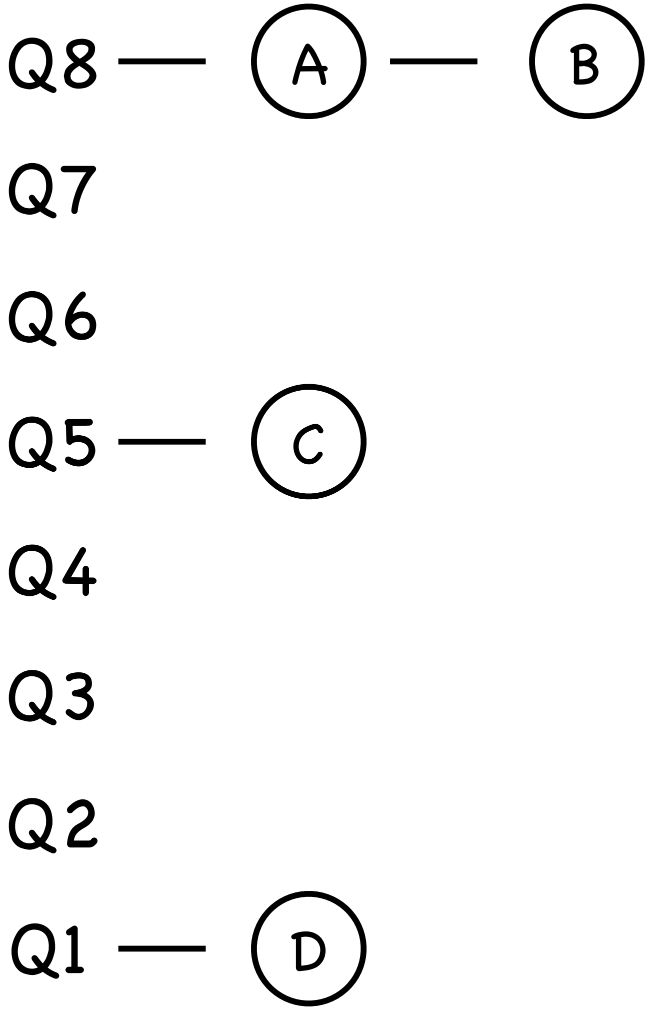
- run job a priority P if no job at higher priorities

RR at each level

Mobility

- Jobs start at top level
- Jobs that use their full quantum drop down a level
- Jobs that don't, stay at the same level

Mobility Rivisited



Mobility

- Jobs start at top level
- Jobs that use their full quantum drop down a level
- Jobs that don't, stay at the same level

To avoid starvation

- Priority boost!

To defeat cheating

- Drop even if quantum completed in stages