



CPU 스케줄링



CPU 스케줄링

- CPU를 잘 사용하기 위해 프로세스를 잘 할당하는 작업
- 목표
 - CPU 이용률은 높게
 - 주어진 시간에 많은 일을 (Throughput)
 - 준비 큐에 있는 프로세스는 적게
 - 응답시간은 짧게



선점 / 비선점 스케줄링

- 비선점(nonpreemptive): 프로세스 종료 or I/O 등의 이벤트가 있을 때까지 보장 (처리시간 예측 용이함)
- 선점(preemptive): OS가 CPU의 사용권을 선점할 수 있는 경우, 강제 회수하는 경우 (처리시간 예측 어려움)



비선점형 스케줄링의 종류

- 어떤 프로세스가 CPU를 점유하고 있다면 이를 뺏을 수 없다. 강제로 프로세스를 중지하지 않는다.
- FCFS(First Come First Served)
 - 큐에 도착한 순서대로 CPU 할당
 - 실행시간이 짧은게 뒤로 가면 평균 대기 시간이 길어짐\
 - 호위 효과(Convoy Effect) 발생 가능



호위 효과(Convoy Effect): 몇 개의 시간이 오래 걸리는 프로세스로 인해 전체 OS가 느려지는 현상

- SJF (Shortest Job First)
 - 실행시간이 가장 짧다고 판단되는 작업을 먼저 수행
 - FCFS보다 평균 대기 시간 감소, 짧은 작업에 유리
 - 기아(Starvation) 현상 발생 가능



기아(Starvation) 현상: 실행 시간이 긴 프로세스가 우선순위가 계속 밀려 실행되지 못하고 무기한으로 대기하게 되는 현상

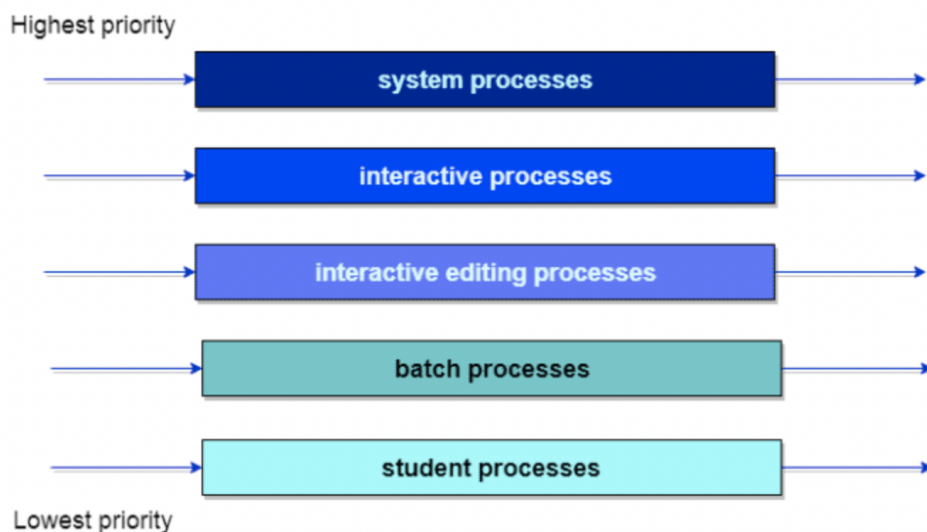
- **HRN (Highest Response-ratio Next)**

- 각각의 프로세스의 우선순위를 계산하여 점유 불평등을 보완한 방법 (SJF의 단점 보완)
- 우선순위 = (대기시간 + 실행시간) / 실행시간

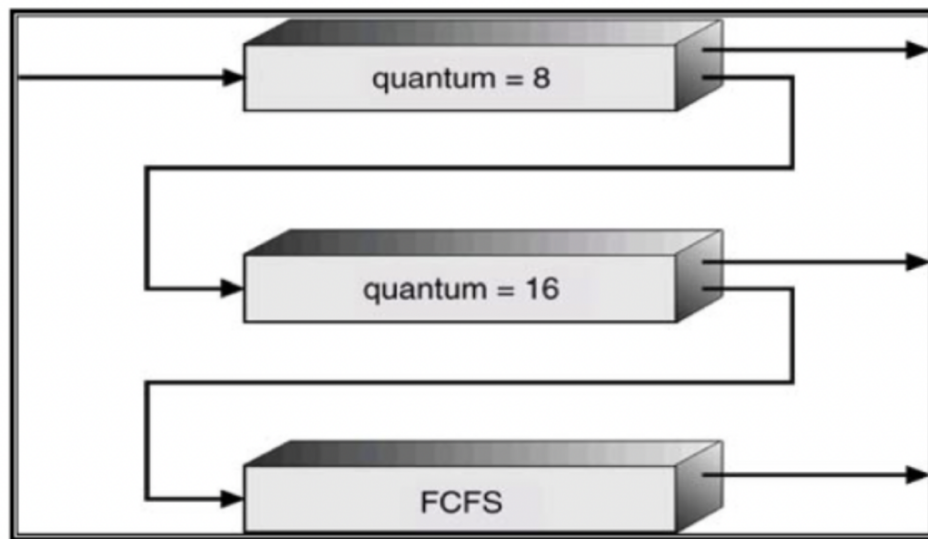


선점형 스케줄링의 종류

- OS가 강제로 CPU를 선점할 수 있다. CPU 사용 독점을 방지할 수 있어 효율적이다. 잦은 문맥 교환으로 오버헤드가 발생할 수 있다.
- **RR (Round Robin)**
 - 각각의 프로세스에 동일한 할당 시간을 부여해서 CPU 이용
 - 할당 시간 내에 처리하지 못하면 강제 중단 후 다음 작업으로 넘어간다
 - 응답 시간이 빨라지지만 할당 시간이 길면 FCFS처럼 작동하고, 할당 시간이 너무 짧으면 문맥 교환이 잦아져서 오버헤드가 증가한다
- **SRTF (Shortest Remaining Time First)**
 - 현재 실행되고 있는 프로세스의 남은 시간보다 다음 프로세스의 실행 시간이 더 짧으면 다음 프로세스를 실행하도록 바꾼다.
 - 평균 대기 시간을 줄일 수 있지만 다음 프로세스의 CPU burst time을 예측하는 것이 어렵다는 문제가 존재한다
- **다단계 큐 (Multilevel Queue)**



- 작업들을 여러 개의 그룹으로 나누어 여러 개의 큐를 이용하는 방법
 - 우선순위가 낮은 큐들이 실행되지 못하는 문제(기아 현상)을 방지하고자 각 큐마다 다른 **Time Quantum** (할당 시간)을 설정한다
 - 우선순위가 높은 큐는 작은 **Time Quantum** 할당. 우선순위가 낮은 큐는 큰 **Time Quantum** 할당.
- 다단계 피드백 큐 (Multilevel Feedback Queue)



- MQ 알고리즘과 다르게 프로세스가 큐 간의 이동이 가능하다.
- MQ 알고리즘은 정해진 시간동안 작업을 처리하지 못하면 동일한 우선순위의 준비 큐에 들어가지만, MFQ 알고리즘의 경우 우선순위가 낮은 큐에 들어감으로써 프로세스의 기아 현상을 해결할 수 있다. 이 기법을 **에이징(aging)**이라고 한다.

💡 CPU 성능 척도

- 시스템 입장
 - Utilization(이용률): 전체시간 중 CPU가 일한 시간
 - Throughput(처리량): 주어진 시간 내 CPU가 처리한 작업 양
- 프로그램(사용자) 입장
 - Turnaroud-Time(소요시간, 반환시간): 프로세스가 CPU를 쓰기 위해 들어와서 다 쓰고 나가기까지의 걸린 시간 == **CPU Burst Time**
 - Waiting-Time(대기시간): 준비 큐에서 기다리는 시간
 - Response-Time(응답시간): 준비 큐에 들어와서 처음 CPU를 할당 받기까지 걸리는 시간