

Planificación de CPU

Agenda

- 1 Objetivos
- 2 Conceptos básicos
- 3 Criterios de planificación
- 4 Algoritmos de planificación

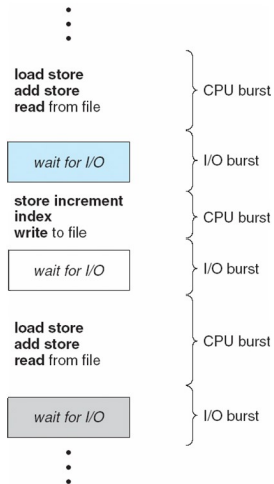
Objetivos

- Presentar los mecanismos de planificación de la CPU, que constituyen los cimientos de los SSOO multiprogramados.
- Describir los distintos algoritmos para la planificación de la CPU.

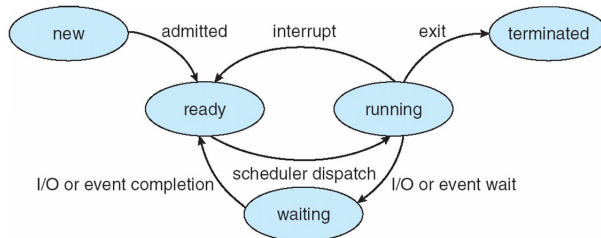
Conceptos básicos

- En un sistema de único procesador, solo puede ejecutarse un proceso por vez (los demás procesos deben esperar).
- Objetivo de la multiprogramación es tener continuamente varios procesos en ejecución (maximizar uso de la CPU).
- En sistemas simples, CPU permanece inactiva cuando hay que esperar por E/S.
- Ejecución de proceso consta de:
 - **Ráfaga de CPU**, ciclo de ejecución en la CPU.
 - **Ráfaga de E/S**, espera de E/S.
- Ejecución de un proceso comienza con una ráfaga de CPU, E/S, CPU, ...etc. CPU concluye con solicitud a sistema para terminar ejecución.

Ciclo de ráfagas de CPU y E/S



Estados de un proceso



Planificador de la CPU

- Cuando CPU queda inactiva, SO selecciona proceso de cola de preparados (ready) que están en memoria y le asigna la CPU.
- Cola de procesos preparados no necesariamente FIFO (first-in,first-out), pueden implementarse otros tipos (prioridad, árbol, etc.).
- Registros de las colas son PCBs.

Planificación, ¿cuándo tomar la decisión?

- Puede ser necesario tomar decisiones sobre planificación en las siguientes circunstancias, cuando un proceso:
 - 1.- cambia su estado de ejecución (running) a espera (waiting) (solicitud E/S, invocación a wait).
 - 2.- cambia su estado de ejecución (running) a preparado (ready) (se produce una interrupción).
 - 3.- cambia su estado de espera (waiting) a preparado (ready) (al completarse E/S).
 - 4.- termina (terminated).
- En planificación **sin expropiación**, el proceso mantiene la CPU hasta que ocurre 1 ó 4. (entrega CPU por voluntad propia).
- En los demás casos es planificación apropiativa (**con expropiación**).

Despachador (Scheduler Dispatch)

- Componente que participa en la planificación de la CPU.
- Proporciona el control de la CPU sobre los procesos seleccionados.
- Su función implica lo siguiente:
 - cambio de contexto.
 - cambio al modo usuario.
 - salto a la posición correcta dentro del programa de usuario para reiniciar dicho programa.
- Debe ser lo mas rápido posible, se invoca en cada conmutación de proceso.
- **Latencia de despacho**, tiempo en detener un proceso e iniciar otro.

Criterios de planificación

- Diferentes algoritmos de planificación tienen diferentes propiedades que pueden favorecer una clase de procesos sobre otros. Considerar tales propiedades a la hora de decidir qué algoritmo utilizar.
- Criterios para comparar los distintos algoritmos:
 - **Utilización de la CPU**, mantener la CPU lo más ocupada posible (40% ligeramente cargado, 90% intensamente utilizado).
 - **Tasa de procesamiento - Throughput**, número de procesos que se completan por unidad de tiempo.
 - **Tiempo de ejecución - Turnaround time**, cuanto tarda en ejecutarse un proceso (cargarse en memoria + esperar en cola de preparados + ejecutarse en la CPU + operaciones E/S).
 - **Tiempo de espera - Waiting time**, suma de los períodos invertidos en esperar en la cola de preparados.
 - **Tiempo de respuesta - Response time**, tiempo que proceso tarda en empezar a responder desde que se le envía una solicitud.

Criterios de optimización

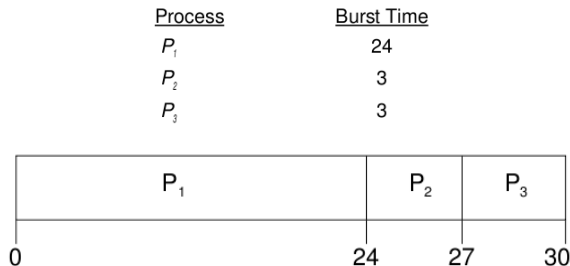
- Maximizar uso de la CPU. (utilizar la mayor parte del tiempo)
- Maximizar tasa de procesamiento. (procesar mayor cantidad de procesos por unidad de tiempo)
- Minimizar tiempo de ejecución. (procesos se ejecuten lo más rápido posible)
- Minimizar tiempo de espera. (procesos no deban esperar mucho tiempo)
- Minimizar tiempo de respuesta. (crítico en sistemas de tiempo compartido)

Algoritmos de planificación

- La planificación de la CPU aborda el problema de decidir a qué proceso de la cola de preparados (ready) asignar la CPU.
- Existen muchos algoritmos de planificación, aplicando cada uno de ellos diferentes criterios para la selección del proceso al cual asignar la CPU.

FCFS

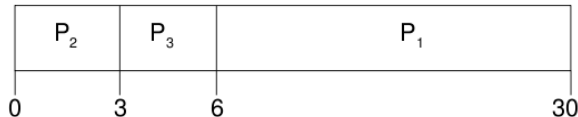
- Es el más simple. first-come, first-served (primero en llegar, primero en ser servido), se gestiona con una cola FIFO.
- En $t=0$, Procesos llegan en el orden: P1, P2, P3



- Tiempo de espera para $P1=0$, $P2=24$, $P3=27$
- Tiempo medio de espera es: $(0 + 24 + 27)/3 = 17$

FCFS

- Supongamos que llegan en $t=0$ en el siguiente orden: P2, P3, P1

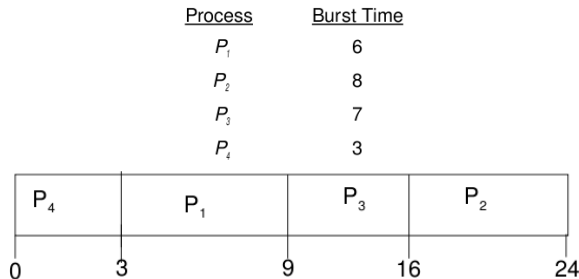


- Tiempo de espera para P1=6, P2=0, P3=3
- Tiempo medio de espera es: $(6 + 0 + 3)/3 = 3$
- Mucho mejor que el anterior.

SJF

- Selección del trabajo más corto (shortest-job-first)
- Se asigna la CPU al proceso que tiene ráfaga de CPU más corta.
- Si son iguales, se usa FCFS para romper el empate.
- Dificultad al tener que conocer duración de siguiente ráfaga de CPU.

SJF



- Tiempo de espera para $P_1 = 3$, $P_2 = 16$, $P_3 = 9$, $P_4 = 0$
- Promedio tiempo espera = $(3 + 16 + 9 + 0)/4 = 7$

Prioridades

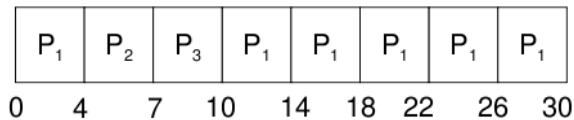
- A cada proceso se le asocia una prioridad. CPU se asigna a prioridad más alta.
- Procesos con misma prioridad se planifican en orden FCFS.
- SJF es un caso particular de prioridad.
- Prioridades pueden ser:
 - **Apropiativa.** Al llegar nuevo proceso, si prioridad es mayor, quita la CPU.
 - **Cooperativa.** Al llegar nuevo proceso, si prioridad es mayor, se coloca al principio de la cola.
- Problema - **muerte por inanición.** procesos de baja prioridad bloqueados indefinidamente.
- Solución - **envejecimiento.** cambiar prioridad hasta que se ejecute.

Round Robin

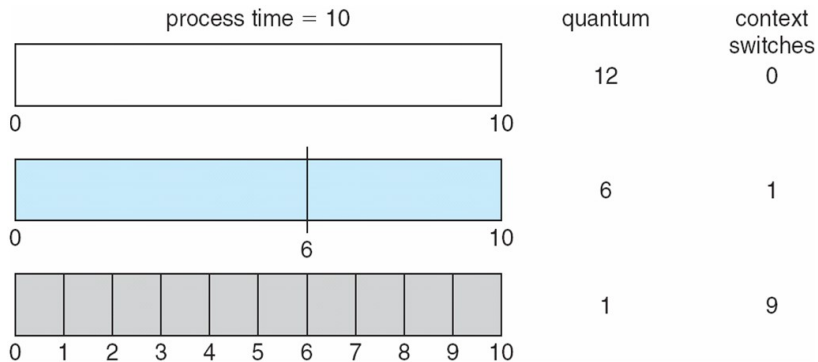
- Planificador por turnos. Especialmente para sistemas de tiempo compartido.
- Similar a FCFS pero con desalojo para conmutar entre procesos.
- **q - Cuanto de tiempo**, unidad de tiempo (10-100 milisegundos).
- Se asigna la CPU a cada proceso durante intervalo de tiempo.
- Cola se trata circularmente.
- Performace:
 - q grande, FIFO
 - q pequeño, debe ser grande comparado con el tiempo de cambio de contexto, de lo contrario overhead muy alto.

Round Robin $q=4$

<u>Process</u>	<u>Burst Time</u>
P_1	24
P_2	3
P_3	3



Round Robin, tiempos de q y cambio contexto

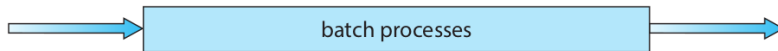
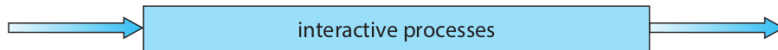
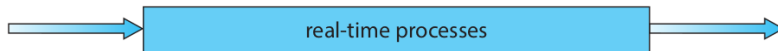


Colas Multinivel

- Procesos se clasifican en grupos diferentes. Requisitos diferentes de tiempos de respuesta:
 - de primer plano (foreground), interactivos. (RR)
 - de segundo plano (background), lotes. (FCFS)
- Cola de preparados (ready) es particionada en colas separadas.
- Cada cola tiene su propio algoritmo de planificación.

Colas Multinivel

highest priority



lowest priority

Colas Multinivel Retroalimentadas

- En colas multinivel, procesos se asignan a una cola (según su naturaleza) y no se mueven a otra.
- Retroalimentadas, permite mover un proceso de una cola a otra.
- Evita el bloqueo indefinido.

