

Planificador Uniprocador

Bobadilla, Manuel
Ghilino, Ramiro
Villarreal, Juan Manuel

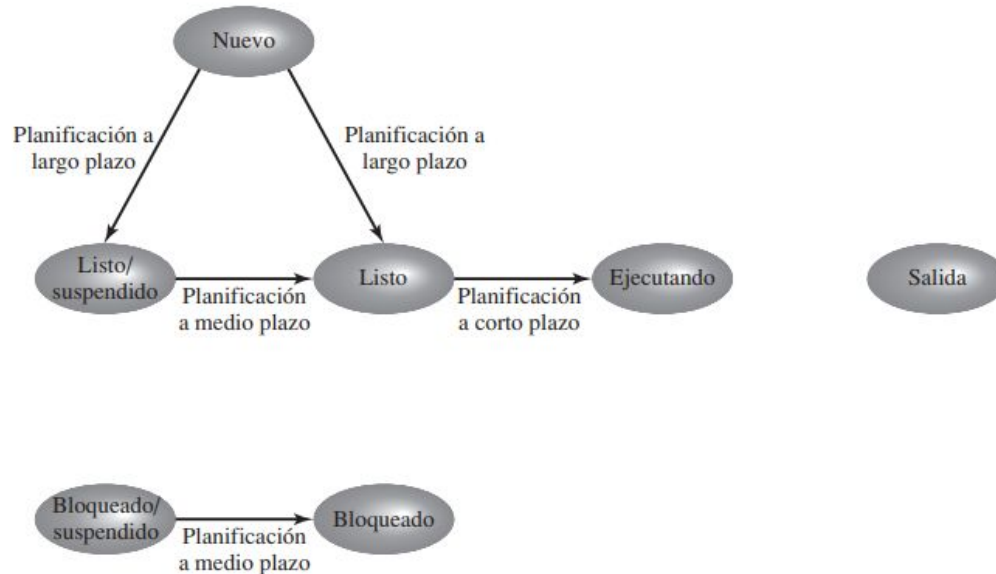
Tipos de planificación

El objetivo de la planificación de procesos es **asignar procesos a ser ejecutados por el procesador o procesadores a lo largo del tiempo**, de forma que se cumplan los objetivos del sistema tales como el tiempo de respuesta, el rendimiento y la eficiencia del procesador.

Planificación a largo plazo	La decisión de añadir un proceso al conjunto de procesos a ser ejecutados
Planificación a medio plazo	La decisión de añadir un proceso al número de procesos que están parcialmente o totalmente en la memoria principal
Planificación a corto plazo	La decisión por la que un proceso disponible será ejecutado por el procesador
Planificación de la E/S	La decisión por la que un proceso que está pendiente de una petición de E/S será atendido por un dispositivo de E/S disponible

Tipos de planificación

La planificación **afecta al rendimiento del sistema** porque determina **qué proceso esperará y qué proceso progresará**. Es un problema de manejo de colas para minimizar el retardo en la cola y para optimizar el rendimiento en un entorno de colas.



Largo plazo

Determina **qué programas se admiten** en el sistema para su procesamiento. Una vez admitido, un trabajo o programa de usuario se convierte en un **proceso** y se **añade a la cola del planificador a corto plazo**. El planificador a largo plazo creará procesos desde la cola siempre que pueda.

En este caso hay que tomar dos decisiones.

- El planificador debe decidir cuándo el sistema operativo puede tomar uno o más procesos adicionales.
- El planificador debe decidir qué trabajo o trabajos se aceptan y son convertidos en procesos.

La decisión de cuándo crear un nuevo proceso se toma **dependiendo del grado de multiprogramación deseado**.

La decisión de qué trabajo admitir el siguiente, puede basarse en un sencillo «primero en llegar primero en servirse» (FCFS), o puede ser una herramienta para gestionar el rendimiento del sistema.

Medio plazo

La planificación a medio plazo es **parte de la función de intercambio**.

Con frecuencia, la decisión de intercambio se basa en la **necesidad de gestionar el grado de multiprogramación**. En un sistema que no utiliza la memoria virtual, la gestión de la memoria es también otro aspecto a tener en cuenta.

De esta forma, la decisión de meter un proceso en la memoria, tendrá en cuenta las necesidades de memoria de los procesos que están fuera de la misma.

Corto plazo

El planificador a corto plazo, conocido también como activador, ejecuta mucho más frecuentemente y toma las decisiones de grano fino **sobre qué proceso ejecutar el siguiente**.

El planificador a corto plazo se invoca **siempre que ocurre un evento que puede conllevar el bloqueo del proceso actual** y que puede proporcionar la oportunidad de expulsar al proceso actualmente en ejecución en favor de otro. Algunos ejemplos de estos eventos son

- Interrupciones de reloj.
- Llamadas al sistema.
- Interrupciones de E/S.
- Señales (por ejemplo, semáforos).

Algoritmos de planificación

El **objetivo principal de la planificación a corto plazo** es asignar tiempo de procesador de tal forma que se optimicen uno o más aspectos del comportamiento del sistema.

Generalmente, se establece un conjunto de criterios con los que se pueden evaluar varias políticas de planificación. Se puede hacer una **distinción entre criterios orientados al usuario y criterios orientados al sistema**

Orientados al usuario

Orientados al usuario, relacionados con las prestaciones

Tiempo de estancia (*turnaround time*) Tiempo transcurrido desde que se lanza un proceso hasta que finaliza. Incluye el tiempo de ejecución sumado con el tiempo de espera por los recursos, incluyendo el procesador. Es una medida apropiada para trabajos por lotes.

Tiempo de respuesta (*response time*) Para un proceso interactivo, es el tiempo que transcurre desde que se lanza una petición hasta que se comienza a recibir la respuesta. A menudo un proceso puede producir alguna salida al usuario mientras continúa el proceso de la petición. De esta forma, desde el punto de vista del usuario, es una medida mejor que el tiempo de estancia. La planificación debe intentar lograr bajos tiempos de respuesta y maximizar el número de usuarios interactivos con tiempos de respuesta aceptables.

Fecha tope (*deadlines*) Cuando se puede especificar la fecha tope de un proceso, el planificador debe subordinar otros objetivos al de maximizar el porcentaje de fechas tope conseguidas.

Orientados al usuario, otros

Previsibilidad Un trabajo dado debería ejecutarse aproximadamente en el mismo tiempo y con el mismo coste a pesar de la carga del sistema. Una gran variación en el tiempo de respuesta o en el tiempo de estancia es malo desde el punto de vista de los usuarios. Puede significar una gran oscilación en la sobrecarga del sistema o la necesidad de poner a punto el sistema para eliminar las inestabilidades.

Orientados al sistema

Orientados al sistema, relacionados con las prestaciones

Rendimiento La política de planificación debería intentar maximizar el número de procesos completados por unidad de tiempo. Es una medida de cuánto trabajo está siendo realizado. Esta medida depende claramente de la longitud media de los procesos, pero está influenciada por la política de planificación, que puede afectar a la utilización.

Utilización del procesador Es el porcentaje de tiempo que el procesador está ocupado. Para un sistema compartido costoso, es un criterio significativo. En un sistema de un solo usuario y en otros sistemas, tales como los sistemas de tiempo real, este criterio es menos importante que algunos otros.

Orientados al sistema, otros

Equidad En ausencia de orientación de los usuarios o de orientación proporcionada por otro sistema, los procesos deben ser tratados de la misma manera, y ningún proceso debe sufrir inanición.

Imposición de prioridades Cuando se asignan prioridades a los procesos, la política del planificador debería favorecer a los procesos con prioridades más altas.

Equilibrado de recursos La política del planificador debería mantener ocupados los recursos del sistema. Los procesos que utilicen poco los recursos que en un determinado momento están sobreutilizados, deberían ser favorecidos. Este criterio también implica planificación a medio plazo y a largo plazo.

Prioridades

En muchos sistemas, a cada proceso se le asigna una prioridad y el planificador siempre elegirá un proceso de prioridad mayor sobre un proceso de prioridad menor.

Un problema de los esquemas de planificación con prioridades es que **los procesos con prioridad más baja pueden sufrir inanición**. Esto sucederá si hay siempre un conjunto de procesos de mayor prioridad listos para ejecutar. Si este comportamiento no es deseable, **la prioridad de un proceso puede cambiar con su antigüedad o historial de ejecución**.

Políticas de planificación alternativas

- Función de selección
- Tres medidas cuantitativas

w = tiempo usado en el sistema hasta este momento, esperando o ejecutando

e = tiempo usado en ejecución hasta este momento

s = tiempo total de servicio requerido por el proceso, incluyendo e ; generalmente, esta cantidad debe ser estimada o proporcionada por el usuario

Políticas de planificación alternativas

- Políticas expulsivas/no expulsivas
- Tiempo de estancia (tiempo de espera + tiempo de servicio)
- Tiempo de estancia normalizado (tiempo de estancia / tiempo de servicio)

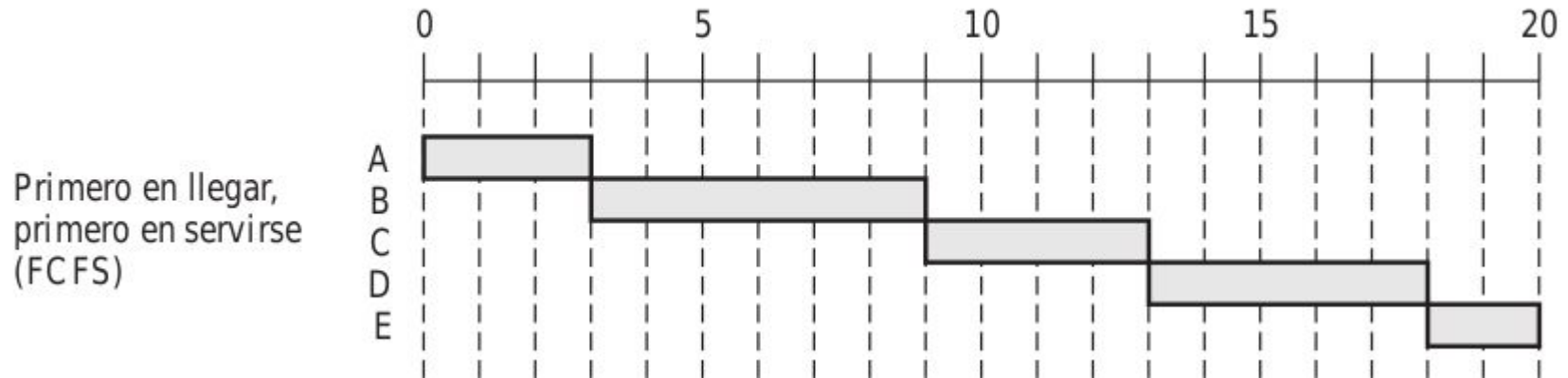
Tiempo de estancia normalizado = (tiempo de espera + tiempo de servicio) / tiempo de servicio
= tiempo de espera / tiempo de servicio + 1

Primero en Llegar Primero en Servirse - FCFS

- Ejecuta el proceso que ha estado más tiempo en la cola de listos
- Beneficia procesos largos
- Beneficia procesos limitados por procesador
- No expulsivo

Proceso	Tiempo de Llegada	Tiempo de Servicio (T_s)	Tiempo de Comienzo	Tiempo de Finalización	Tiempo de Estancia (T_r)	T_r/T_s
W	0	1	0	1	1	1
X	1	100	1	101	100	1
Y	2	1	101	102	100	100
Z	3	100	102	202	199	1,99
Media					100	26

Primero en Llegar Primero en Servirse - FCFS



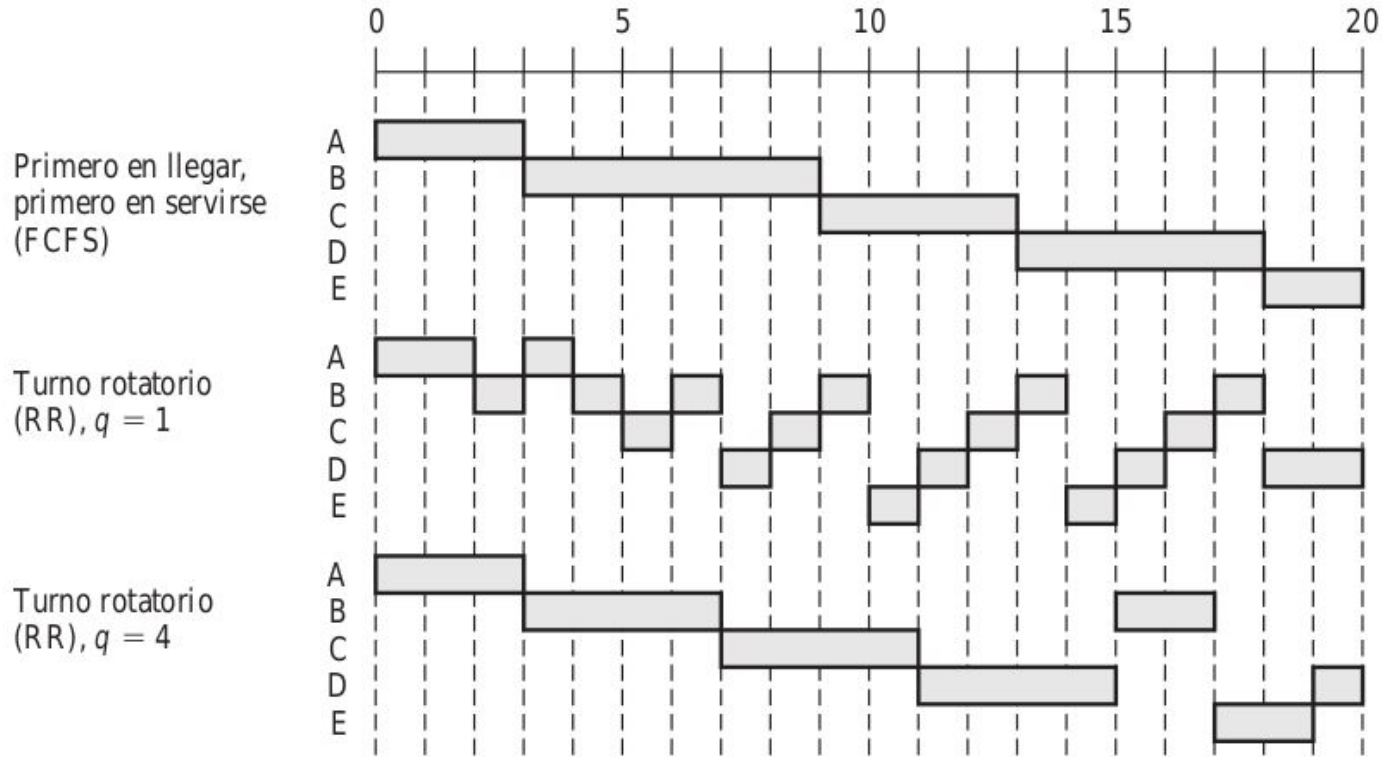
Primero en Llegar Primero en Servirse - FCFS

	Proceso	A	B	C	D	E	Media
	Tiempo de llegada Tiempo de servicio (T_s)	0 3	2 6	4 4	6 5	8 2	
FCFS	Tiempo de finalización	3	9	13	18	20	8.60 2.56
	Tiempo de estancia (T_e)	3	7	9	12	12	
	T/T_s	1.00	1.17	2.25	2.40	6.00	

Turno Rotatorio - RR

- Reduce el beneficio a los procesos largos
- Método expulsivo
- Se basa en el reloj
- Definición de quantum ligeramente mayor que el tiempo requerido para una función/interacción típica del proceso
- Trato desfavorable para procesos limitados por E/S

Turno Rotatorio - RR



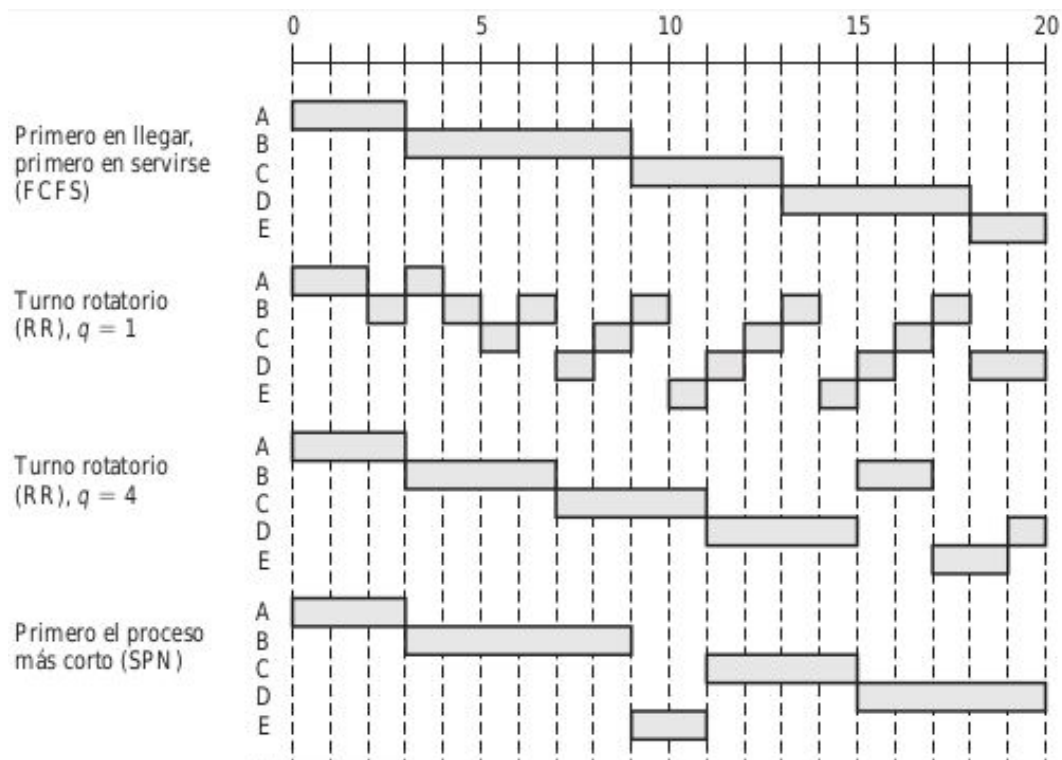
Turno Rotatorio - RR

	Proceso Tiempo de llegada Tiempo de servicio (T_s)	A 0 3	B 2 6	C 4 4	D 6 5	E 8 2	Media
FCFS	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	13 9 2.25	18 12 2.40	20 12 6.00	8.60 2.56
RR $q = 1$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	4 4 1.33	18 16 2.67	17 13 3.25	20 14 2.80	15 7 3.50	10.80 2.71
RR $q = 4$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	17 15 2.5	11 7 1.75	20 14 2.80	19 11 5.50	10.00 2.71

Primero el Proceso más Corto - SPN

- No expulsiva
- Selección del proceso más corto
- Mejora tiempos de respuesta
- Más variabilidad en los tiempos de respuesta
- Definir duración de cada proceso
- Posible inanición

SPN



SPN

	Proceso Tiempo de llegada Tiempo de servicio (T_s)	A 0 3	B 2 6	C 4 4	D 6 5	E 8 2	Media
FCFS	Tiempo de finalización Tiempo de estancia (T_j) T_j/T_s	3 3 1.00	9 7 1.17	13 9 2.25	18 12 2.40	20 12 6.00	8.60 2.56
RR $q = 1$	Tiempo de finalización Tiempo de estancia (T_j) T_j/T_s	4 4 1.33	18 16 2.67	17 13 3.25	20 14 2.80	15 7 3.50	10.80 2.71
RR $q = 4$	Tiempo de finalización Tiempo de estancia (T_j) T_j/T_s	3 3 1.00	17 15 2.5	11 7 1.75	20 14 2.80	19 11 5.50	10.00 2.71
SPN	Tiempo de finalización Tiempo de estancia (T_j) T_j/T_s	3 3 1.00	9 7 1.17	15 11 2.75	20 14 2.80	11 3 1.50	7.60 1.84

SPN - Definición Duración Proceso

- Debe ser proporcionado por el programador
- Generado por el SO

SPN - Definición Duración Proceso

$$S_{n+1} = \frac{1}{n} \sum_{i=1}^n T_i \quad (9.1)$$

donde,

T_i = tiempo de ejecución del procesador para la instancia i -ésima de este proceso (tiempo total de ejecución para los trabajos por lotes; tiempo de ráfaga de procesador para los trabajos interactivos)

S_i = valor predicho para la instancia i -ésima

S_1 = valor predicho para la primera instancia; no calculado

Para evitar volver a calcular la suma completa cada vez, podemos reescribir la ecuación como

$$S_{n+1} = \frac{1}{n} T_n + \frac{n-1}{n} S_n \quad (9.2)$$

SPN - Definición Duración Proceso

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

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

$$\alpha = 0,8$$

$$S_{n+1} = 0,8T_n + 0,16T_{n-1} + 0,032T_{n-2} + 0,0064T_{n-3} + \dots$$

SPN - Definición Duración Proceso

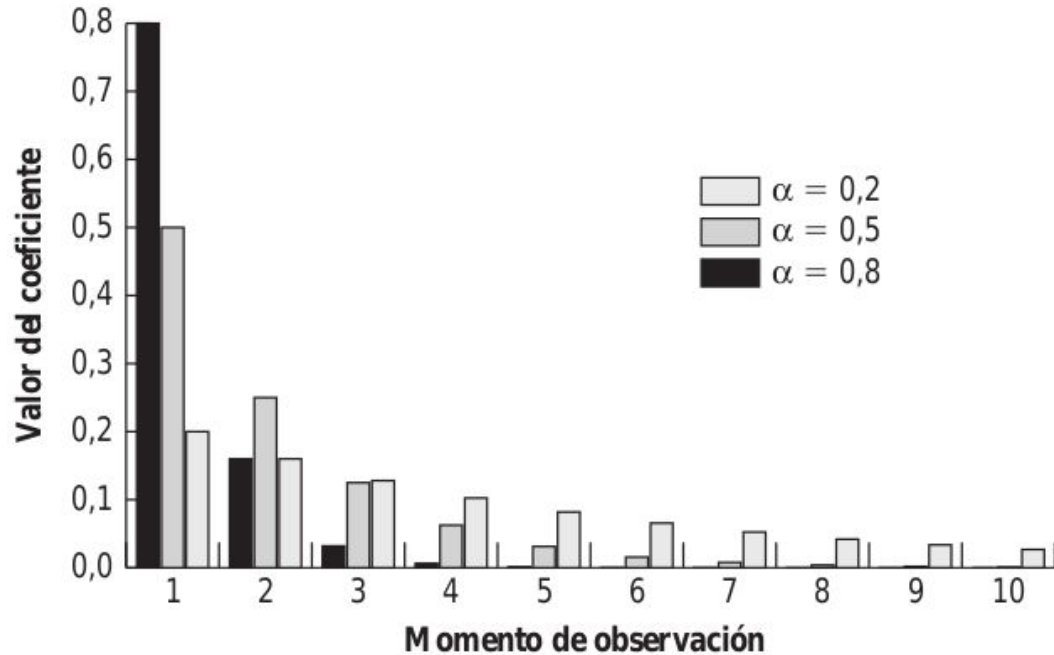
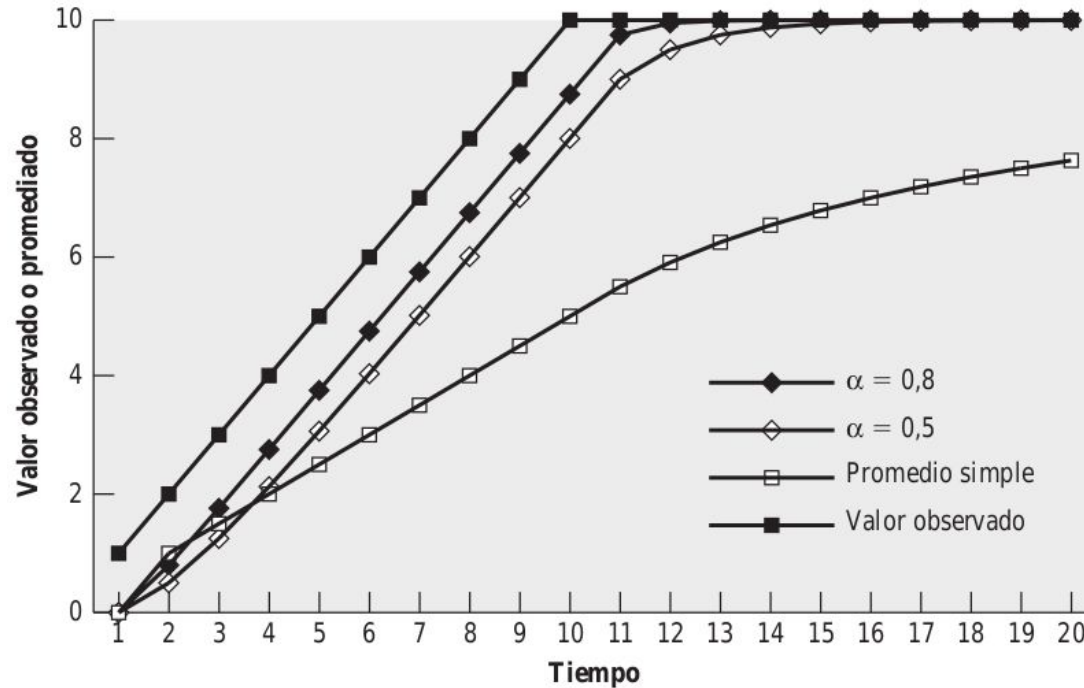


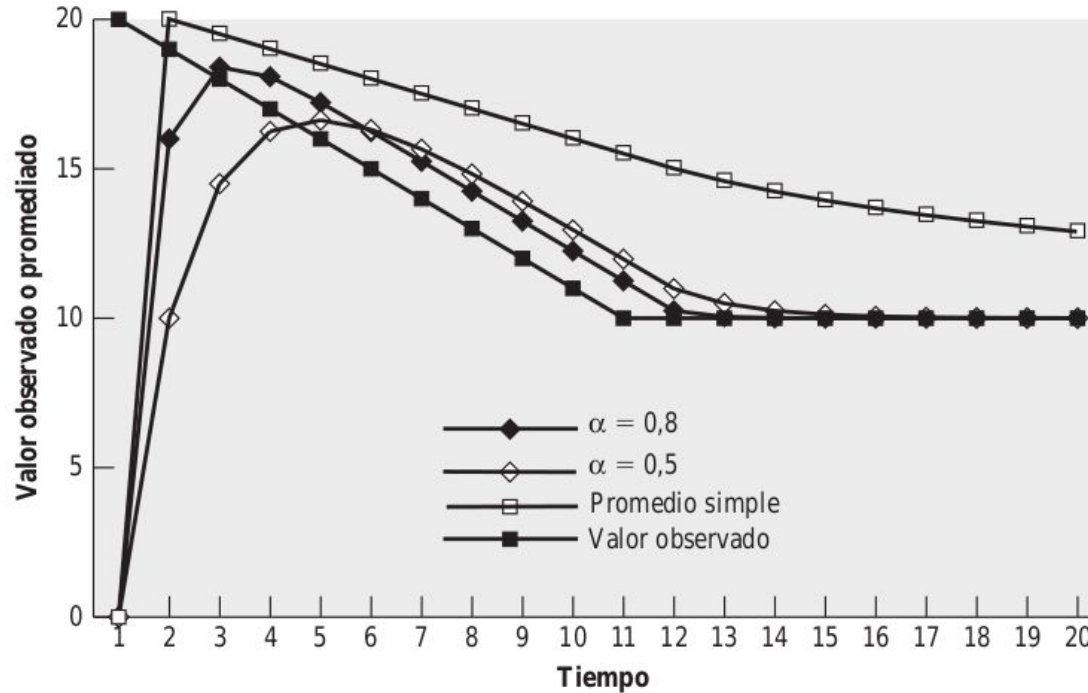
Figura 9.8. Coeficientes exponenciales suaves.

SPN - Definición Duración Proceso



(a) Función creciente

SPN - Definición Duración Proceso



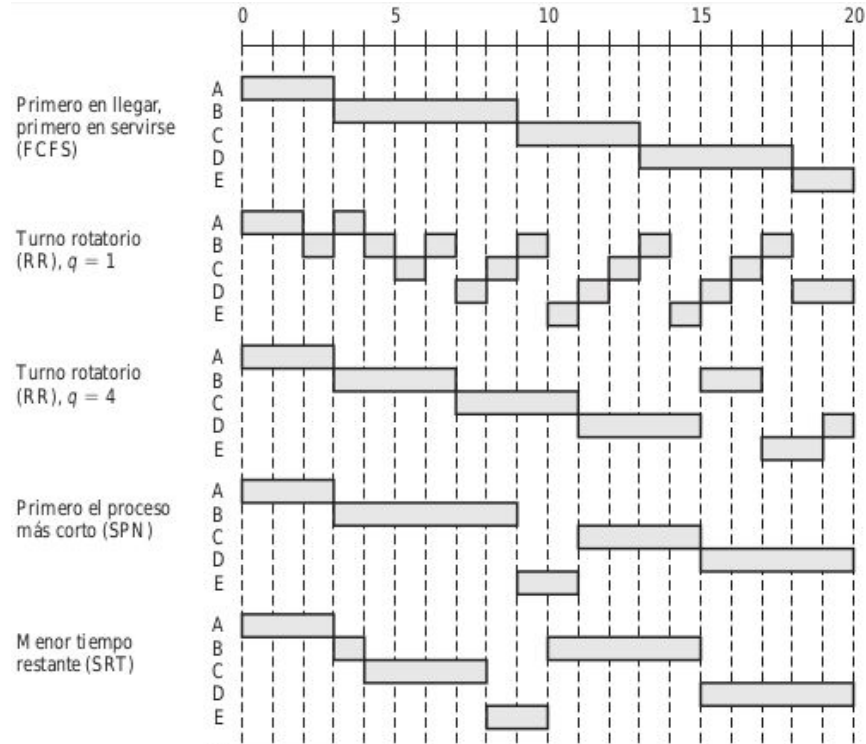
(b) Función decreciente

Figura 9.9. Uso del promedio exponencial.

Menor Tiempo Restante - SRT

- Versión expulsiva del SPN
- Se selecciona el proceso con el menor tiempo de proceso restante
- Se debe almacenar los tiempos de servicio transcurridos

Menor Tiempo Restante - SRT



Menor Tiempo Restante - SRT

	Proceso Tiempo de llegada Tiempo de servicio (T_s)	A 0 3	B 2 6	C 4 4	D 6 5	E 8 2	Media
FCFS	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	13 9 2.25	18 12 2.40	20 12 6.00	8.60 2.56
RR $q = 1$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	4 4 1.33	18 16 2.67	17 13 3.25	20 14 2.80	15 7 3.50	10.80 2.71
RR $q = 4$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	17 15 2.5	11 7 1.75	20 14 2.80	19 11 5.50	10.00 2.71
SPN	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	15 11 2.75	20 14 2.80	11 3 1.50	7.60 1.84
SRT	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	15 13 2.17	8 4 1.00	20 14 2.80	10 2 1.00	7.20 1.59

Primero el de Mayor Tasa de Respuesta - HRRN

- Elegir el proceso listo con el mayor valor de R
- No expulsivo
- Favorece a procesos más cortos y más viejos

$$R = \frac{(w + s)}{s}$$

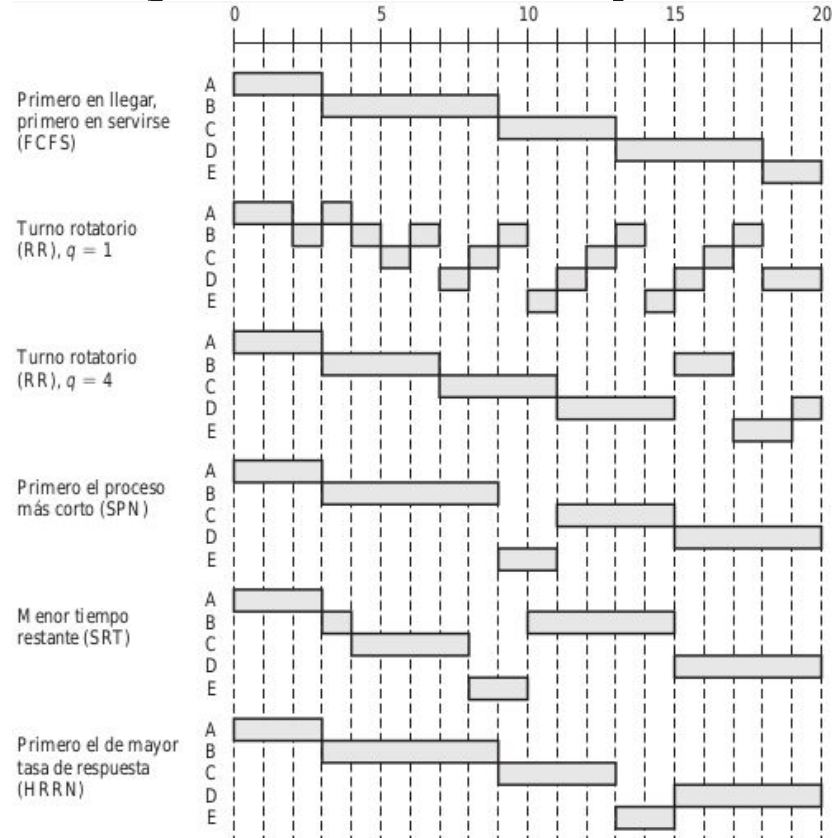
donde

R = tasa de respuesta

w = tiempo invertido esperando por el procesador

s = tiempo de servicio esperado

Primero el de Mayor Tasa de Respuesta - HRRN



Primero el de Mayor Tasa de Respuesta - HRRN

	Proceso Tiempo de llegada Tiempo de servicio (T_s)	A 0 3	B 2 6	C 4 4	D 6 5	E 8 2	Media
FCFS	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	13 9 2.25	18 12 2.40	20 12 6.00	8.60 2.56
RR $q = 1$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	4 4 1.33	18 16 2.67	17 13 3.25	20 14 2.80	15 7 3.50	10.80 2.71
RR $q = 4$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	17 15 2.5	11 7 1.75	20 14 2.80	19 11 5.50	10.00 2.71
SPN	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	15 11 2.75	20 14 2.80	11 3 1.50	7.60 1.84
SRT	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	15 13 2.17	8 4 1.00	20 14 2.80	10 2 1.00	7.20 1.59
HRRN	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	13 9 2.25	20 14 2.80	15 7 3.5	8.00 2.14

Retroalimentación - F

- A veces no se conoce el tiempo de los procesos
- En vez de favorecer a los procesos más cortos (desconocidos) penalizar a los más largos
- Expulsiva, política rotatoria
- Prioridades dinámicas (varias colas)
- En todas las colas, a excepción de la última se utiliza FCFS

Retroalimentación - F

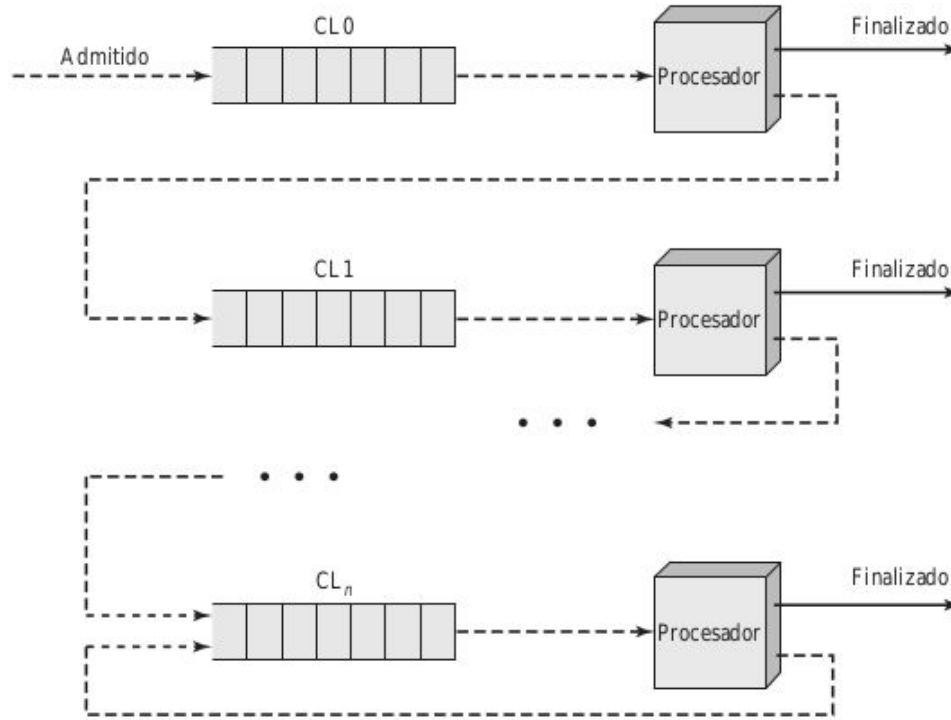
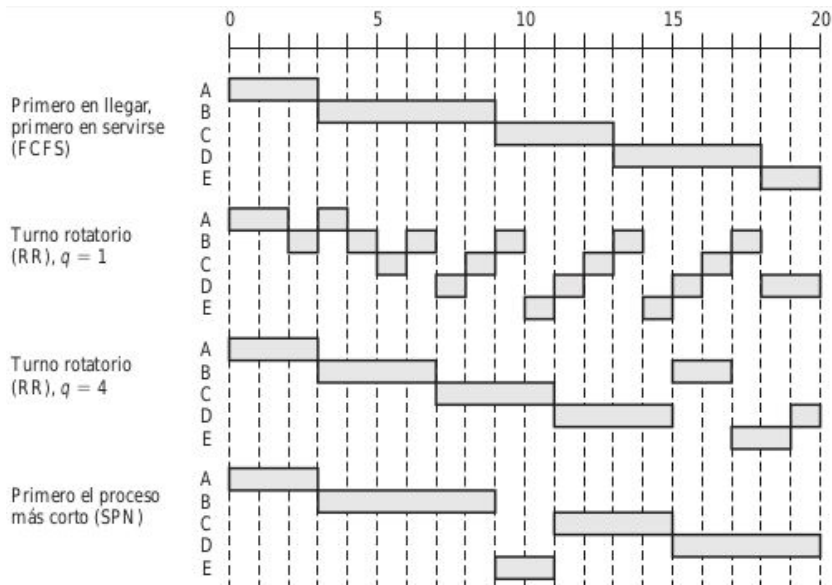


Figura 9.10. Planificación retroalimentada.

Retroalimentación - F

- Puede alargar mucho la estancia de procesos largos, puede haber inanición
- Se puede variar los tiempos de expulsión de cada cola
- Mover los procesos a una cola de mayor prioridad si esperan mucho tiempo

Retroalimentación - F

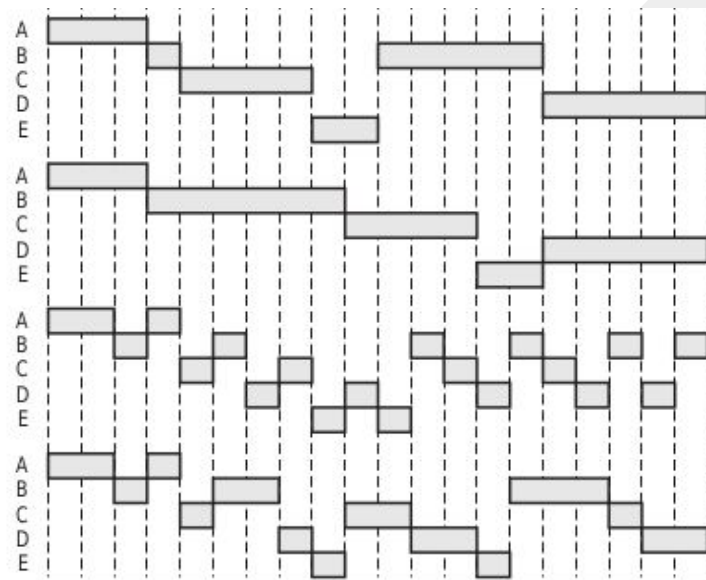


Menor tiempo restante (SRT)

Primero el de mayor tasa de respuesta (HRRN)

Retroalimentación $q = 1$

Retroalimentación $q = 2$



Retroalimentación - F

	Proceso Tiempo de llegada Tiempo de servicio (T_s)	A 0 3	B 2 6	C 4 4	D 6 5	E 8 2	Media
FCFS	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	13 9 2.25	18 12 2.40	20 12 6.00	8.60 2.56
RR $q = 1$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	4 4 1.33	18 16 2.67	17 13 3.25	20 14 2.80	15 7 3.50	10.80 2.71
RR $q = 4$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	17 15 2.5	11 7 1.75	20 14 2.80	19 11 5.50	10.00 2.71
SPN	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	15 11 2.75	20 14 2.80	11 3 1.50	7.60 1.84
SRT	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	15 13 2.17	8 4 1.00	20 14 2.80	10 2 1.00	7.20 1.59
HRRN	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	3 3 1.00	9 7 1.17	13 9 2.25	20 14 2.80	15 7 3.5	8.00 2.14
FB $q = 1$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	4 4 1.33	20 18 3.00	16 12 3.00	19 13 2.60	11 3 1.5	10.00 2.29
FB $q = 2'$	Tiempo de finalización Tiempo de estancia (T_e) T/T_s	4 4 1.33	17 15 2.50	18 14 3.50	20 14 2.80	14 6 3.00	10.60 2.63

Comparación de Rendimiento

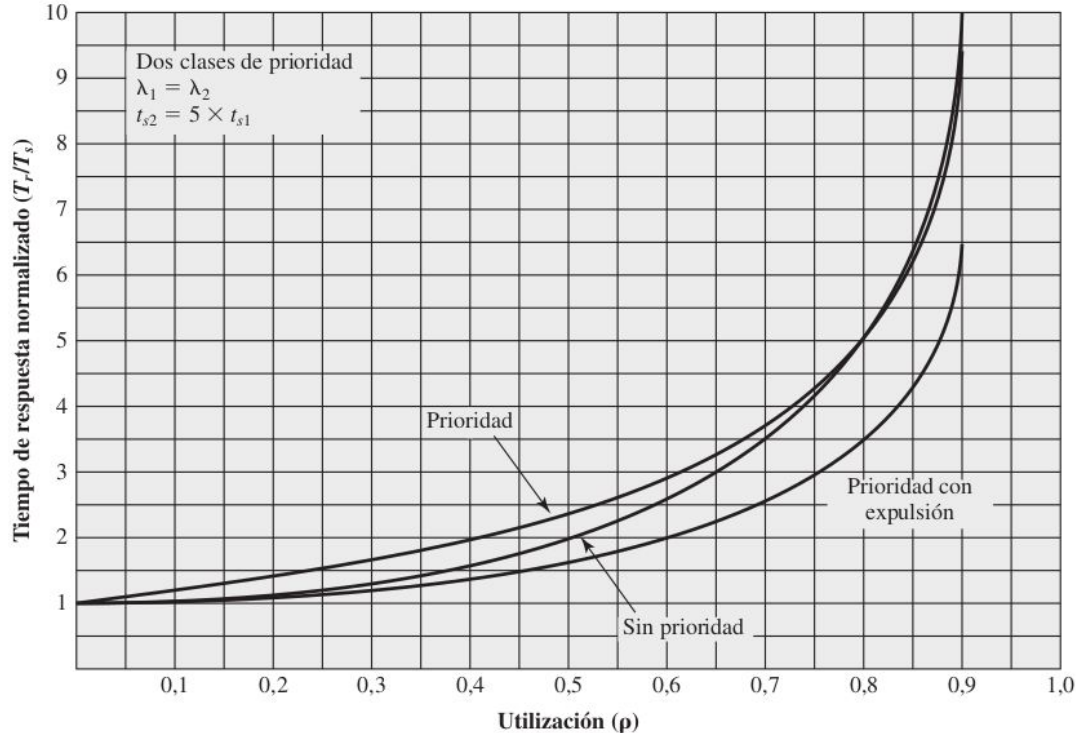


Figura 9.11. Tiempo total de respuesta normalizado.

Comparación de Rendimiento

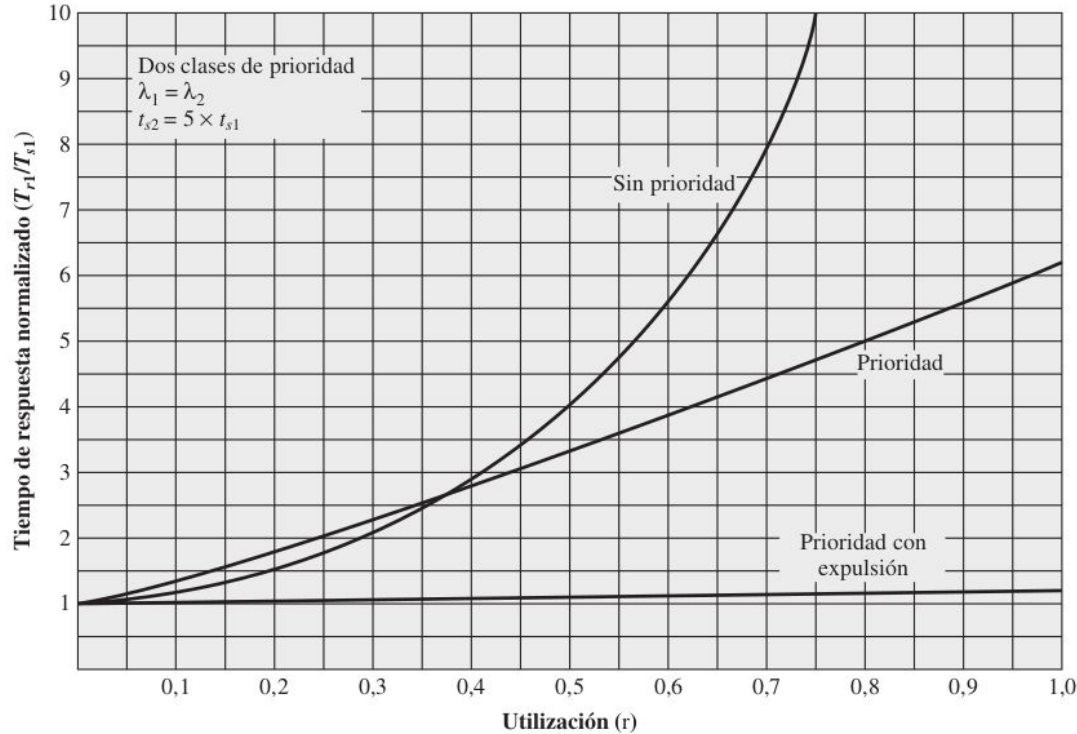


Figura 9.12. Tiempo de respuesta normalizado para los procesos más cortos.

Comparación de Rendimiento

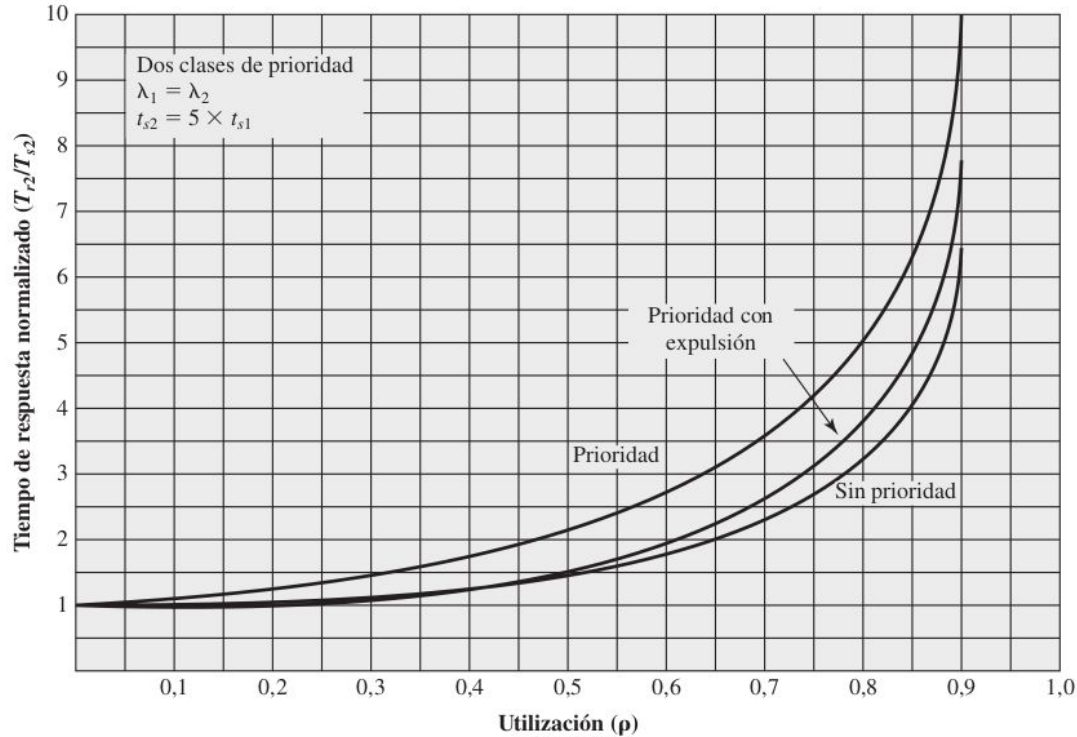


Figura 9.13. Tiempo de respuesta normalizado para los procesos más largos.

Comparación de Rendimiento

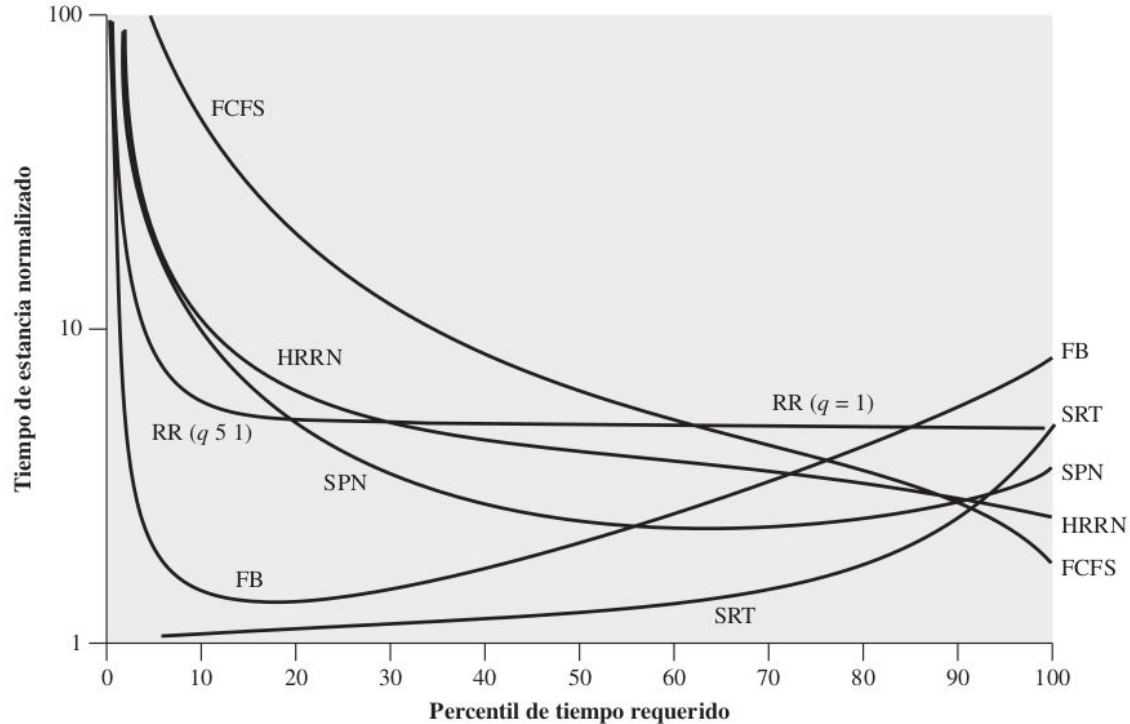


Figura 9.14. Resultados de la simulación para el tiempo de estancia normalizado.

Comparación de Rendimiento

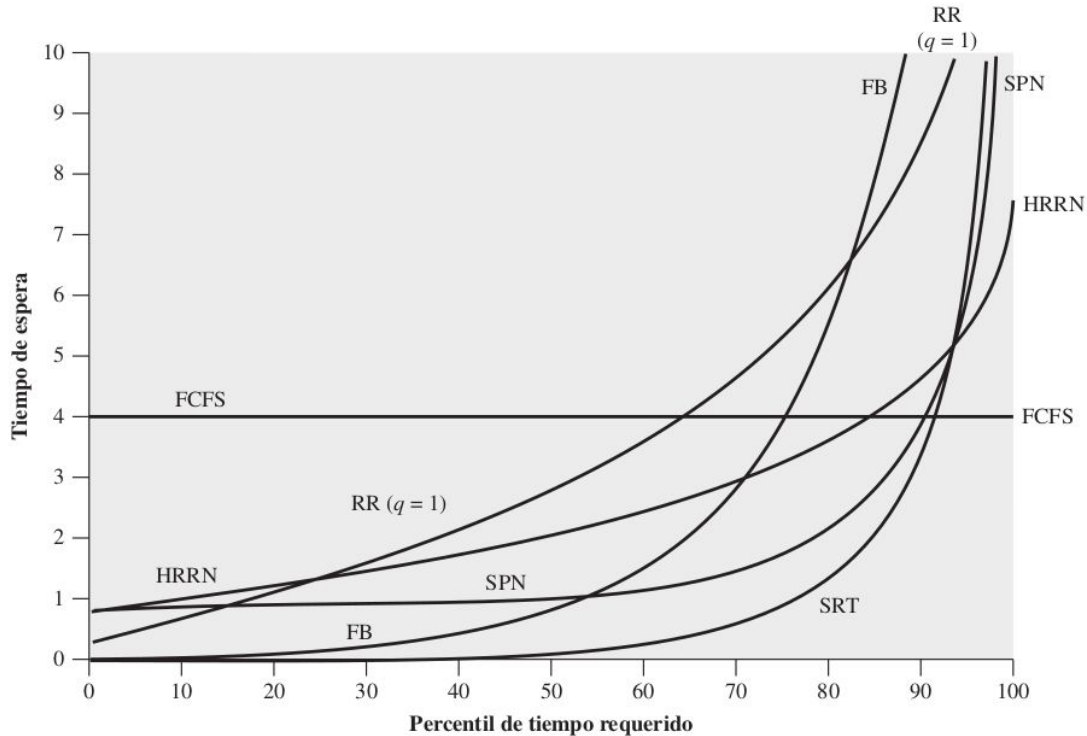


Figura 9.15. Resultados de la simulación para el tiempo de espera.

La Planificación de Contribución Justa (Fair-Share Scheduling)

- Algoritmo de planificación pensado para sistemas multiusuarios.
- No existe una única colección de procesos listos, sino múltiples conjuntos de procesos, uno para cada grupo de usuarios.
- Ideal para equipos de tiempo compartido, donde todos los grupos reciben la misma prima de servicio, por lo que solo se degrada el servicio para grupos con muchos usuarios.

¿Cómo se calculan prioridades y usos?

$$CPU_j(i) = \frac{CPU_j(i-1)}{2}$$

$$GCPU_k(i) = \frac{GCPU_k(i-1)}{2}$$

$$P_j(i) = Base_j + \frac{CPU_j(i)}{2} + \frac{GCPU_k(i)}{4 \times W_k}$$

Donde:

- $CPU_j(i)$ = Medida de utilización del procesador por el proceso j en el intervalo i .
- $GCPU_k(i)$ = Medida de utilización del procesador del grupo k en el intervalo i .
- $P_j(i)$ = Prioridad del proceso j al comienzo del intervalo i ; valores más pequeños equivalen a prioridades más altas.
- $Base_j$ = Prioridad base del proceso j .
- W_k = Prima asignada al grupo k , con la restricción que $0 < W_k \leq 1$ y $\sum W_k = 1$.

Tiempo	Proceso A			Proceso B			Proceso C		
	Prioridad	Contador CPU proceso	Contador CPU grupo	Prioridad	Contador CPU proceso	Contador CPU grupo	Prioridad	Contador CPU proceso	Contador CPU grupo
0	60	0 1 2 • • 60	0 1 2 • • 60	60	0 1 2 • • 60	0 1 2 • • 60	60	0 1 2 • • 60	0 1 2 • • 60
1	90	30	30	60	0 1 2 • • 60	0 1 2 • • 60	60	0 1 2 • • 60	0 1 2 • • 60
2	74	15 16 17 • • 75	15 16 17 • • 75	90	30	30	75	0	30
3	96	37	37	74	15	15 16 17 • • 75	67	0 1 2 • • 60	15 16 17 • • 75
4	78	18 19 20 • • 78	18 19 20 • • 78	81	7	37	93	30	37
5	98	39	39	70	3	18	76	15	18

Planificación UNIX Tradicional

Esta planificación está diseñada para proporcionar buenos tiempos de respuesta a usuarios interactivos a la vez que asegura que los trabajos de fondo de baja prioridad no sufran inanición.

Para lograr esto, se emplea una retroalimentación multinivel y utiliza planificación de turno rotatorio en cada una de las colas de prioridad.

El sistema hace uso de expulsión de 1 segundo, es decir, si un proceso no se bloquea o se completa en un segundo, es expulsado.

La prioridad está basada en el tipo de proceso y el histórico de ejecución.

Cálculos en Planificación UNIX

$$CPU_j(i) = \frac{CPU_j(i-1)}{2}$$

$$P_j(i) = Base_j + \frac{CPU_j(i)}{2} + ajuste_j$$

Donde:

- $CPU_j(i)$ = Medida de utilización del procesador por el proceso j en el intervalo i .
- $P_j(i)$ = Prioridad del proceso j al comienzo del intervalo i ; valores más pequeños equivalen a prioridades más altas.
- $Base_j$ = Prioridad base del proceso j .
- $ajuste_j$ = Factor de ajuste controlable por el usuario.

Planificación UNIX Tradicional

- La prioridad de cada proceso se recalcula una vez por segundo, momento en el que se realiza una nueva decisión de planificación.
- El propósito de la prioridad (*P*) es dividir todos los procesos en bandas fijas de niveles de prioridad.
- Para asegurar lo anterior, los componentes *CPU* y *ajuste* se limitan para impedir que los procesos migren de su banda asignada (por la *Base*).
- El objetivo de las bandas es optimizar el acceso a dispositivos de bloques (por ejemplo, disco) y para permitir al sistema operativo responder rápidamente a llamadas al sistema.

Bandas de Prioridad

Las bandas de prioridad, de mayor a menor prioridad, son:

- Intercambiador (*swapper*).
- Control de dispositivos de bloque E/S.
- Manipulación de ficheros.
- Control de dispositivos orientados a carácter de E/S.
- Procesos de usuario.



¿Preguntas?



**¡Muchas
Gracias!**

