

Curso: Sistemas Operativos
6. Sistema de Entrada/Salida
(Parte II)

6 Sistemas de Entrada/Salida

6.5 Planificación del disco

Planificación del disco.

Durante la últimas décadas, el incremento de la velocidad de los procesadores y de la memoria principal ha sobrepasado con creces el de la velocidad de acceso a disco. Los discos son actualmente al menos cuatro órdenes de magnitud más lentos que la memoria principal. Por tanto, el rendimiento del subsistema de almacenamiento de disco es de vital importancia, por lo que se han realizado muchas investigaciones en esquemas para mejorar ese rendimiento.

Los detalles reales de la operación de E/S del disco dependen de la computadora, del sistema operativo y de la naturaleza del hardware del canal de E/S y del controlador del disco.

Parámetros de rendimiento del disco.

Los parámetros a considerar para el rendimiento del disco son:

- *Tiempo medio de acceso*: tiempo medio que tarda la aguja en situarse en la pista y el sector deseado; es la suma del *Tiempo medio de búsqueda* (situarse en la pista), *Tiempo de lectura/escritura* y la *Latencia media* (situarse en el sector).

Parámetros de rendimiento del disco (continuación).

- *Tiempo medio de búsqueda*: tiempo medio que tarda la aguja en situarse en la pista deseada; es la mitad del tiempo empleado por la aguja en ir desde la pista más periférica hasta la más central del disco.
- *Tiempo de lectura/escritura*: tiempo medio que tarda el disco en leer o escribir nueva información. Depende de la cantidad de información que se quiere leer o escribir, el tamaño de bloque, el número de cabezales, el tiempo por vuelta y la cantidad de sectores por pista.
- *Latencia media*: tiempo medio que tarda la aguja en situarse en el sector deseado; es la mitad del tiempo empleado en una rotación completa del disco.
- *Velocidad de rotación*: Es la velocidad a la que gira el disco duro, más exactamente, la velocidad a la que giran el/los platos del disco, que es donde se almacenan magnéticamente los datos. La regla es: a mayor velocidad de rotación, más alta será la transferencia de datos, pero también mayor será el ruido y mayor será el calor generado por el disco duro. Se mide en número revoluciones por minuto (RPM). Un disco duro IDE en promedio gira a 5400 RPM (ya hay discos IDE de 7200 RPM), un disco SCSI generalmente gira a 7200 RPM (los hay de 10.000 RPM).

6 Sistemas de Entrada/Salida

6.5 Planificación del disco

Parámetros de rendimiento del disco (continuación).

Una velocidad de 5400 RPM permitirá una transferencia entre 80MB y 110MB por segundo con los datos que están en la parte exterior del cilindro o plato. A mayor velocidad de rotación, menor latencia media.

- *Tasa de transferencia*: velocidad a la que puede transferir la información a la computadora una vez que la aguja está situada en la pista y sector correctos. Puede ser *velocidad sostenida* o *de pico*.

Algoritmos de planificación de acceso a disco.

La transferencia de datos desde y hacia los discos son uno de los procesos más lentos de los que gestiona el sistema operativo. El tiempo de búsqueda es una de las mediciones que ocupan más tiempo en la realización de una operación de E/S, lo que influye mucho en el rendimiento del sistema.

Si tenemos un gran número de operaciones que realizar en disco, procedentes de procesos que hacen solicitudes de lectura o escritura de bloques, resulta que el orden en que se manipulen las solicitudes tiene un impacto significativo en el rendimiento global del sistema. Es por esto que se tienen varios algoritmos; revisaremos algunos de ellos.

Algoritmos de planificación de acceso a disco.

Para cada algoritmo de planificación de acceso a disco que trataremos a continuación, vamos a revisar cuántas pistas tiene que recorrer la cabeza lectora/escritora (tiempo de búsqueda), así como el número de cambios de dirección de movimiento de dicha cabeza, para atender solicitudes de lectura o escritura en las siguientes pistas, llegadas en el orden mostrado: 17, 30, 24, 37, 15, 27, 11, 75, 20, 5

El disco cuenta con 80 pistas e inicialmente la cabeza está en la pista 28, dirección al centro del disco. La pista cero es la más externa. A la cabeza le toma 0.8 ms ir de una pista a la otra pista contigua y el cambio de dirección de 5 ms.

Algoritmo FCFS (first come, first served).

Las solicitudes se forman en una cola, conforme van llegando y así se van atendiendo. Este algoritmo es fácil de implementar y tiene la ventaja de que es *justo* en el sentido de que el proceso que solicita primero es el primero en ser servido. Calculemos ahora las pistas que va recorriendo la cabeza para atender cada solicitud, y el total; así como el número de cambios de dirección de la cabeza.

Pistas solicitadas	17	30	24	37	15	27	11	75	20	5
# de pistas recorridas de solicitud a solicitud	11	13	6	13	22	12	16	64	55	15
Cambio dirección cabeza	sí	sí	sí	sí	sí	sí	sí	sí	sí	no

Algoritmo FCFS (continuación).

Por lo que el total de pistas recorridas es de 227; y hubo 9 cambios de dirección.

Entonces el tiempo de búsqueda total para atender las solicitudes fue:

$$227 * 0.8 + 9 * 5 = \underline{226.6 \text{ ms.}}$$

Algoritmo Pickup

Es una variante del algoritmo FCFS. En este algoritmo, las solicitudes suelen tomarse en orden, como en el FCFS, pero a medida que el sistema mueve la cabeza se detiene para cualquier pista que no haya pasado aún y que tiene una solicitud en la cola. En Linux realiza este algoritmo y lo denomina *Noop*.

Aplicando el algoritmo en las solicitudes de ejemplo, partiría de la pista 28 (donde está localizada la cabeza) y comenzaría a moverse a la pista 17, la primera solicitud de la cola; pero en su camino debe atender las pistas 27, 24 y 20. La secuencia total sería: 27, 24, 20, 17, 30, 37, 15, 11, 75, 5

Pistas solicitadas	27	24	20	17	30	37	15	11	75	5
Num pistas recorridas de solicitud a solicitud	1	3	4	3	13	7	22	4	64	70
Cambio dirección cabeza	sí	no	no	no	sí	no	sí	no	sí	sí

Algoritmo Pickup (continuación).

Por lo que el total de pistas recorridas es de 191; y hubo 5 cambios de dirección.

Entonces el tiempo de búsqueda total para atender las solicitudes fue:

$$191 * 0.8 + 5 * 5 = \underline{177.8 \text{ ms.}}$$

Vemos que el tiempo de búsqueda se reduce en un 21.5% al algoritmo FCFS.

Algoritmo SSTF (shortest seek time first).

Este algoritmo se enfoca en atender la pista que lleve menor tiempo de búsqueda, o sea atender la pista más cercana físicamente.

Aplicando el algoritmo en las solicitudes de ejemplo, partiría de la pista 28 (donde está localizada la cabeza); la entrada más próxima en la cola es 27, por lo que se movería a ella, luego la más cercana es la 30 de modo que se movería ahí. La secuencia total de atención sería: 27, 30, 24, 20, 17, 15, 11, 5, 37, 75.

Pistas solicitadas	27	30	24	20	17	15	11	5	37	75
Num pistas recorridas de solicitud a solicitud	1	3	6	4	3	2	4	6	32	38
Cambio dirección cabeza	sí	sí	sí	no	no	no	no	no	sí	sí

Algoritmo SSTF(continuación).

Por lo que el total de pistas recorridas es de 99; y hubo 5 cambios de dirección.

Entonces el tiempo de búsqueda total para atender las solicitudes fue:

$$99 * 0.8 + 6 * 5 = \underline{109.2 \text{ ms.}}$$

Observamos que baja aún más el tiempo de búsqueda.

Este algoritmo, sin embargo tiene algunas desventajas: la primera solicitud en la cola (17) se atiende cuando ya se atendieron a la mitad de solicitudes, también observamos que la cabeza se mantiene en la zona donde se encontraba inicialmente, lo que da como resultado que un proceso que estaba teniendo acceso a bloques que estaban ubicados en la parte donde se mueve primordialmente la cabeza, tiende a obtener un mejor servicio que uno que estaba teniendo acceso a bloques fuera de dicha área.

Es así que se considera un mecanismo de priorización donde se asigna una prioridad de forma gradual, a las solicitudes que se encuentren por más tiempo en la cola, y sean atendidas más pronto. Esta variante se le denomina algoritmo *ASSTF(aged shortest seek time first)*, o sea antigüedad y menor tiempo de búsqueda primero.

Algoritmo Look.

Este algoritmo también se denomina “algoritmo de ascensor”. Consiste en que una vez que la cabeza tiene una dirección, no invierte su dirección de búsqueda hasta que ha atendido las solicitudes que hay en esa dirección.

Aplicando el algoritmo en las solicitudes de ejemplo, partiría de la pista 28 (donde está localizada la cabeza) en dirección al centro del disco, o sea a pistas más altas. Entonces atendería a la pista 30 y continuaría en atender las siguientes más altas. La secuencia total de atención sería: 30, 37, 75, 27, 24, 20, 17, 15, 11, 5.

Pistas solicitadas	30	37	75	27	24	20	17	15	11	5
Num pistas recorridas de solicitud a solicitud	2	7	38	48	3	4	3	2	4	6
Cambio dirección cabeza	no	no	no	sí	no	no	no	no	no	no

Por lo que el total de pistas recorridas es de 117; y hubo solo 1 cambio de dirección. Entonces el tiempo de búsqueda total para atender las solicitudes fue:

$$117 * 0.8 + 1 * 5 = \underline{98.6 \text{ ms.}}$$

Observamos que es el tiempo de búsqueda más bajo de todos los algoritmos antes vistos.

Algoritmo C-Look (Circular-Look).

Es una variante del Look más justo. Su funcionamiento consiste en que cuando la cabeza se ha movido hacia la última pista en una dirección, en lugar de invertir la dirección y buscar la pista más próxima, se sigue buscando la más alejada en la cola en la dirección opuesta. Luego empieza a realizar las búsquedas moviéndose de vuelta en la dirección en la que estaba desplazándose antes.

Aplicando el algoritmo en las solicitudes de ejemplo, partiría de la pista 28 (donde está localizada la cabeza) en dirección al centro del disco, o sea a pistas más altas. Entonces atendería a la pista 30 y continuaría en atender las siguientes más altas. Luego se va a atender la pista más baja, o sea, la 5 y nuevamente hacia las más altas. La secuencia total de atención sería: 30, 37, 75, 5, 11, 15, 17, 20, 24, 27.

Pistas solicitadas	30	37	75	5	11	15	17	20	24	27
Num pistas recorridas de solicitud a solicitud	2	7	38	70	6	4	2	3	4	3
Cambio dirección cabeza	no	no	no	sí	no	no	no	no	no	no

Por lo que el total de pistas recorridas es de 139; y hubo sólo 1 cambio de dirección. Entonces el tiempo de búsqueda total para atender las solicitudes fue:

$139 * 0.8 + 1 * 5 = \underline{116.2 \text{ ms.}}$ Es más alto que el algoritmo Look pero es más justo.

Algoritmo F-Scan.

Es una variante del algoritmo Look. Ocupa dos colas, por ejemplo X y Y. Primero atiende las solicitudes de la cola X, usando el algoritmo Look. Si durante la atención de la cola X llegan más solicitudes, éstas las coloca en la cola Y. Una vez que atiende todas las solicitudes de la cola X, inicia con la atención de solicitudes de la cola Y, la cual no recibe ninguna solicitud más ya que son enviadas a la cola X; y así sucesivamente.

Este mecanismo es un compromiso entre la imparcialidad de FCFS y la eficiencia de Look sin la larga búsqueda costosa de C-Look. Por tanto evita largos periodos de inanición.

Algoritmo N-step Scan.

Es una última variante del algoritmo Look; organiza por lotes las solicitudes en grupos de N solicitudes. Primero atiende las solicitudes de un lote con el algoritmo Look, antes de pasar al siguiente. Así evita la postergación indefinida de una solicitud a ser atendida.

Veamos el efecto de N-step Scan de acuerdo al tamaño de N. Si N es igual a 1, N-step Scan es efectivamente FCFS, y si N es suficientemente grande para que casi todas las solicitudes sean atendidas en la primera atención, N-step Scan es equivalente a Look.

Planificadores de Linux.

Linux maneja varios planificadores de disco. Veamos tres que actualmente utiliza.

Planificador anticipado.

Fue un tiempo el planificador predeterminado en Linux. Fusiona solicitudes como el algoritmo Pickup y usa la secuencia de ascensor unidireccional del algoritmo Look.

Una característica única es que intenta anticipar las operaciones de lectura al mantener apagado un bit después de un comando de lectura asíncrona si considera que un proceso puede solicitar más datos cercanos. Si llega una solicitud del último proceso, puede invertir la dirección de búsqueda si la distancia no es muy grande.

Planificador de tiempo límite.

También fusiona los algoritmos Pickup y Look, además impone un tiempo límite sobre todas las operaciones para evitar inanición de recursos. El planificador de tiempo límite otorga prioridad a las solicitudes de lectura en la medida en que no haya pasado el tiempo límite para una solicitud de escritura. Este planificador se usa en sistemas de bases de datos, especialmente si los discos son unidades de alto rendimiento.

Planificador de cola completamente justo (CFQ, Completely Fair Queuing).

Este planificador , de igual manera fusiona los algoritmos Pickup y Look; también trata de dar a todos los procesos que usan un dispositivo particular el mismo número de solicitudes síncronas de E/S sobre un intervalo medido.

Otros dispositivos de almacenamiento masivo.

Hasta el momento, hemos estudiado dos medios de almacenamiento masivo: Las cintas y los discos magnéticos rotativos. Ahora veamos algunos otros que han sido o son tendencia.

Discos ópticos.

Los discos ópticos utilizan diversas técnicas de grabación. En una de ellas, se utiliza calor intenso generado por láser para quemar agujeros microscópicos en un recubrimiento metálico. En otra, el calor del láser crea ampollas en la superficie. En una tercera, se altera la reflectividad de la superficie.

Los primeros discos ópticos fueron dispositivos de *grabación única y lectura múltiple* (WORM, write-once-read-many). En la actualidad, ejemplos de discos ópticos y sus tecnologías de grabación (escritura) son :

- Por moldeado durante la fabricación, mediante un molde de níquel (CD-ROM y DVD ROM),
- Por la acción de un haz láser (CD-R y CD-RW, también llamado CD-E),
- Por la acción de un haz láser en conjunción con un campo magnético (discos magneto-ópticos - MO).

Otros dispositivos de almacenamiento masivo.

Discos ópticos (continuación).

Por su gran capacidad de almacenamiento, las aplicaciones más comunes de los discos ópticos son:

- Bases de datos en CD ROM para bibliotecas de datos invariables (enciclopedias, distribución de software, manuales de software, demos, etc.),
- Servidores de archivos en una red local,
- Uso de CD-R (grabables por el usuario) para backups, y bibliotecas de Imágenes.

Algunas características importantes de los discos ópticos:

- Son medios removibles con capacidad para almacenar masivamente datos en pequeños espacios.
- Son portables (removibles de la unidad) y seguros en la conservación de datos.
- Tienen bajo costo por byte almacenado.
- Los CD-ROM se producen masivamente
- Los CD son más seguros en la conservación de datos que los discos magnéticos debido a que la capa que los almacena es inmune los campos magnéticos caseros, y están protegidos de la corrosión ambiental, etc.

Otros dispositivos de almacenamiento masivo.

Almacenamiento en estado sólido.

Desde la última década ha crecido el uso de medios de almacenamiento de estado sólido, los cuales no contienen partes móviles. Las características de estos medios de almacenamiento son muy distintas de las de los discos; a pesar de tener esta característica, se presentan ante el sistema operativo como si fuera lo mismo, pudiera ser que para ser utilizados pronto y sin tener que esperar a que los desarrolladores de sistemas operativos adecuaran los controladores, se conectan mediante la misma interfaz y empleando la misma semántica que un disco rotativo. Esto es porque las unidades de estado sólido cuentan con una capa de traducción que emula el comportamiento de un disco duro.

Tipos de medios de almacenamiento de estado sólido.

- *NVRAM Unidades RAM No Volátil.* Almacenan la información en Chips de RAM estándar, con un respaldo de batería para mantener la información cuando se desconecta la corriente externa.

Su principal ventaja es la velocidad y durabilidad; el tiempo de acceso o escritura de datos es la misma de la memoria principal del sistema y este tiempo es el mismo independientemente de la dirección que se solicite.

Otros dispositivos de almacenamiento masivo.

Su principal desventaja es el precio; por byte almacenado en una RAM, es ciento de veces más caro que el medio magnético. el tiempo de acceso o escritura de datos es la misma de la memoria principal del sistema y este tiempo es el mismo independientemente de la dirección que se solicite. Otra desventaja es que si bien el medio no se degrada con el uso, la batería sí, lo que podría perderse la información.

- *Memoria flash.* Derivada de los EEPROM (Electrically Erasable Programmable Read-Only Memory), del cual toma las operaciones de lectura, escritura y borrado. En las memorias flash, el espacio de almacenamiento está dividido en muchas celdas, y el controlador puede leer, borrar o escribir a cada uno de ellos por separado.

En la actualidad, hay dos tipos principales de dispositivos flash.

Unidades de estado sólido (SSD). Son unidades flash de alta velocidad y capacidad; presentan una interfaz similar a la que tienen los discos duros; hoy en día, la más común es SATA. Su velocidad de lectura es muy superior a la de discos duros, no así su velocidad de escritura y borrada que es igual. El precio por el mismo volumen de almacenamiento es 5 y 10 veces el de los discos duros.

Estas unidades se emplean tanto como unidades independientes o como un componente de la tarjeta madre en dispositivos móviles como teléfonos y tabletas.

Otros dispositivos de almacenamiento masivo.

Transporte de archivos. Esta tecnología también está presente en las diversas unidades extraíbles o móviles, como las unidades USB, SD, Memory Stick, Compact flash, entre otros. La principal diferencia entre éstas son los diferentes conectores que emplean. Todas estas tecnologías presentan dispositivos que varían fuertemente en capacidad, velocidad y durabilidad.

RAID

RAID es un acrónimo del inglés que significa *Redundant Array of Independent Disks*, literalmente *matriz de discos independientes redundantes*, aunque no todos los sistemas RAID proporcionan redundancia.

La finalidad de un sistema RAID de discos es la de proteger los datos en caso de que un disco duro falle, o en algunos casos tiene como función principal mejorar la velocidad de lectura de varios discos que conforman un único volumen.

En otras palabras, consiste en crear un único volumen con varios discos duros funcionando en conjunto, y con este conjunto se puede conseguir redundancia (tolerancia a fallos en el caso de que uno falle, conocido como *disk mirroring*) o mayor velocidad (conocido como *disk striping*), haciendo que ese conjunto sea en realidad un tándem.

Otros dispositivos de almacenamiento masivo.

Transporte de archivos. Esta tecnología también está presente en las diversas unidades extraíbles o móviles, como las unidades USB, SD, Memory Stick, Compact flash, entre otros. La principal diferencia entre éstas son los diferentes conectores que emplean. Todas estas tecnologías presentan dispositivos que varían fuertemente en capacidad, velocidad y durabilidad.

RAID

RAID es un acrónimo del inglés que significa *Redundant Array of Independent Disks*, literalmente *matriz de discos independientes redundantes*, aunque no todos los sistemas RAID proporcionan redundancia.

La finalidad de un sistema RAID de discos es la de proteger los datos en caso de que un disco duro falle, o en algunos casos tiene como función principal mejorar la velocidad de lectura de varios discos que conforman un único volumen.

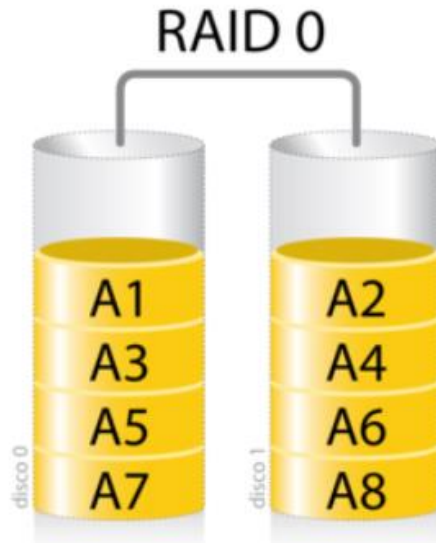
En otras palabras, consiste en crear un único volumen con varios discos duros funcionando en conjunto, y con este conjunto se puede conseguir redundancia (tolerancia a fallos en el caso de que uno falle, conocido como *disk mirroring*) o mayor velocidad (conocido como *disk striping*), haciendo que ese conjunto sea en realidad un tándem.

RAID (continuación)

Hay muchos tipos de RAID; veamos algunos de ellos.

RAID 0

Este tipo de RAID supone el concepto principal que proporciona mayor velocidad al sistema. La información se va escribiendo en dos discos de manera alterna, es decir, un bit en uno, y otro bit en otro, de manera que el ancho de banda es literalmente el doble y por eso se mejora notablemente el rendimiento en este modo. Además, se duplica la capacidad del volumen, es decir, si usamos dos discos duros de 1 TB cada uno, tendríamos un volumen de 2 TB.



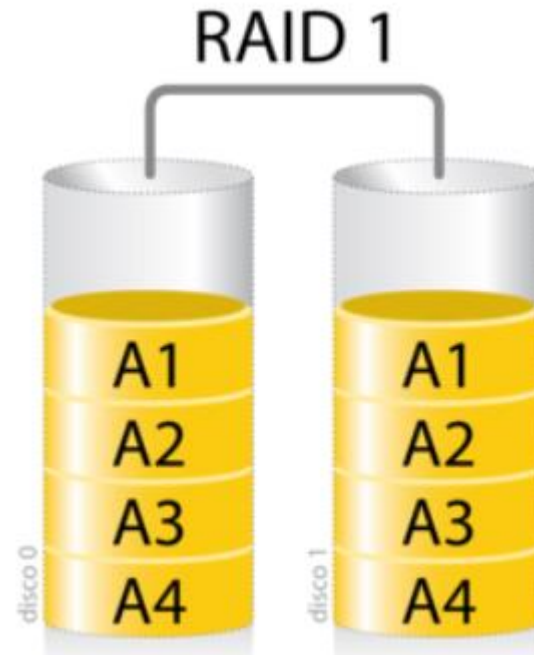
La desventaja de este tipo de RAID es que si fallara alguno de los dos discos duros, la información de los dos se echaría a perder puesto que se encontraría repartida entre los dos.

RAID (continuación)

RAID 1

Este es el otro tipo básico de RAID, y supone el concepto principal de redundancia. En este modo, los datos se escriben en los dos discos de manera simultánea, siendo el uno una copia exacta del otro, motivo por el que se conoce a este modo como *de espejo (mirroring)*. En este caso, si se estropeará uno de los dos discos no pasaría nada porque los datos estarían todavía en el otro, y bastaría con reemplazar el disco estropeado por uno nuevo para volver a restablecer el RAID 1.

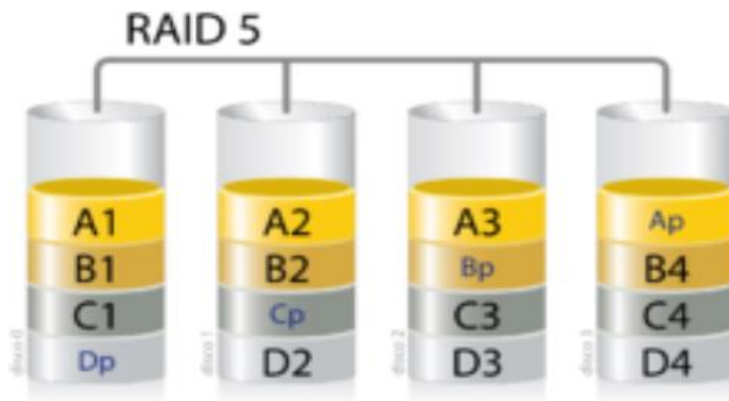
La desventaja de este tipo de RAID es que no se gana ningún rendimiento, al contrario, ya que todos los datos deben escribirse dos veces. Además, el tamaño del volumen será el del disco de menor capacidad. Es decir, si usáramos un disco de 1 TB y otro de 500 GB, tendríamos un volumen de 500 GB en RAID 1.



RAID 5

Un RAID 5 (también llamado distribuido con paridad) es una división de datos a nivel de bloques que distribuye la información de paridad entre todos los discos miembros del conjunto. El RAID 5 ha logrado popularidad gracias a su bajo coste de redundancia. Generalmente, el RAID 5 se implementa con soporte hardware para el cálculo de la paridad. RAID 5 necesitará un mínimo de 3 discos para ser implementado.

Cada vez que un bloque de datos se escribe en un RAID 5, se genera un bloque de paridad dentro de la misma división (*stripe*). Un bloque se compone a menudo de muchos sectores consecutivos de disco. Una serie de bloques (un bloque de cada uno de los discos del conjunto) recibe el nombre colectivo de división (*stripe*). Si otro bloque, o alguna porción de un bloque, es escrita en esa misma división, el bloque de paridad (o una parte del mismo) es recalculada y vuelta a escribir



El disco utilizado por el bloque de paridad está escalonado de una división a la siguiente, de ahí el término *bloques de paridad distribuidos*. Las escrituras en un RAID 5 son costosas en términos de operaciones de disco y tráfico entre los discos y la controladora.

6 Sistemas de Entrada/Salida

Preguntas base para siguiente examen. Contestarlas en sus notas.

- De los algoritmos de planeación de acceso al disco:
 - a) Identificar cuáles son justos
 - b) ¿Cuáles propician que una solicitud de E/S se postergue por mucho tiempo?
 - c) ¿Cuáles son más veloces?.
- Distinguir las ventajas y desventajas de los discos ópticos y de los dispositivos de estado sólido.
- Objetivos de un sistema RAID de discos.

Ejercicio D

Para los algoritmos de planificación de acceso al disco indicados, revisar cuántas pistas tiene que recorrer la cabeza lectora/escritora (tiempo de búsqueda), así como el número de cambios de dirección de movimiento de dicha cabeza, para atender solicitudes de lectura o escritura en las siguientes pistas, llegadas en el orden mostrado: 25, 18, 11, 60, 72, 55, 30, 28, 45, 5

Si el disco cuenta con 80 pistas e inicialmente la cabeza está en la pista 10 , dirección al centro del disco, la pista cero es la más externa, a la cabeza le toma 0.7 ms ir de una pista a la otra pista contigua y el cambio de dirección de 3 ms, calcula el tiempo de búsqueda total para atender todas las solicitudes.

Preguntas base para siguiente examen (continuación). Contestarlas en sus notas.

Ejercicio D (continuación).

- a) Algoritmo FCFS (first come, first served).
- b) Algoritmo Pickup.
- c) Algoritmo SSTF (shortest seek time first).
- d) Algoritmo Look.
- e) Algoritmo C-Look (Circular-Look).