

Memoria virtual (1/7)

- Es un esquema de gestión de memoria en donde puede que sólo una parte del espacio de direcciones virtuales de un proceso residente sea cargada realmente en memoria física
- En otras palabras la memoria virtual permite la ejecución de procesos parcialmente cargados
- Esto se consigue manteniendo una imagen del espacio de direcciones virtuales completo de un proceso en memoria secundaria, y trayendo a memoria principal parte de esa imagen cuando sea necesaria. La elección de que sección traer, cuando y donde es efectuada por el s.o.



Memoria virtual (2/7)

- La memoria virtual puede implementarse como extensión de la gestión de memoria paginada, segmentada o como una combinación de ambas
- La tarea adicional del hardware de traducción de direcciones en sistemas virtuales es detectar si el elemento referenciado esta en memoria real o no. Para ello se añade **un bit de presencia**, a cada entrada de la TP (en el caso de gestión paginada)



Memoria virtual (3/7)

- En el caso de que en el momento de la traducción el bit este a cero, el hardware genera una excepción por elemento ausente para anunciar el hecho al s.o. A esta excepción se denomina **fallo de página**
- Por lo tanto el fallo de página se produce cuando en la tabla de páginas del proceso no hay asociado un marco de página a la página que se quiere acceder



Memoria virtual (4/7)

- Cuando el proceso en ejecución experimenta un fallo de página **queda suspendido** hasta que la página que falta sea incorporada en memoria principal, así el procedimiento a seguir es el siguiente:
 - ◆ Se verifica si la dirección es válida. Si la dirección fuera inválida se enviará una señal al proceso o se abortaría
 - ◆ Si es una referencia válida, pero aún no se ha traído esta página, el s.o. detecta un fallo de página y determina la página virtual requerida para traerla. En este caso la instrucción queda interrumpida y se guarda el estado del proceso, para poder continuarlo en el mismo lugar y estado
 - ◆ Se selecciona un marco libre. Si no existe un marco libre, se tendría que ejecutar



Memoria virtual (5/7)

- ◆ Cuando el marco queda limpio, el s.o. examina la dirección en el disco donde se encuentra la página necesaria y planifica una operación de lectura de la misma. Mientras se carga la página, el proceso sigue suspendido y se permite ejecutar otro proceso
- ◆ Cuando se completa la lectura del disco, la tabla de páginas se actualiza para indicar que ya se dispone de la página en memoria
- ◆ La instrucción que produjo el fallo regresa al estado de comienzo y se planifica su ejecución, pudiéndose acceder a la página como si siempre hubiese estado en memoria



Memoria virtual (6/7)

- La memoria virtual puede producir la interrupción de ciertas instrucciones **durante** su ejecución. Para abordar este problema:
 - ◆ Deshacer los efectos parciales.
 - ◆ Reanudar la instrucción desde el punto exacto de interrupción
 - ◆ Preexaminar las referencias a memoria, antes de comenzar la ejecución de la instrucción



Memoria Virtual (7/7)

- Políticas de la gestión de la memoria virtual:
 - ◆ **Política de asignación:** que cantidad de memoria real se asigna a un proceso
 - ◆ **Política de acceso:** que elementos se incorporan y cuando
 - ◆ **Política de sustitución:** si no existe memoria libre cuando se debe añadir un nuevo elemento, que elemento se desaloja
 - ◆ **Política de ubicación:** donde se coloca el elemento nuevo



Política de sustitución (1/4)

- Gobierna la elección de la página víctima cuando se va a producir el desalojo
- Dependiendo de si ha sido modificada o no el elemento retirado de memoria puede tener que ser escrito en disco o simplemente descartado. Para ello se utiliza el **bit de modificación o de suciedad** en la TP
- Es posible elegir aleatoriamente una página para sustituirla por la solicitada, pero el rendimiento es mayor si se elige una página que no se usa



Política de sustitución (2/4)

- Se pretende que la razón de fallos de página sea lo más baja posible, por lo que se dispone de los siguientes algoritmos
 - ◆ **FIFO**: se sustituye la página que lleva más tiempo en memoria. En realidad, no es necesario guardar el tiempo de entrada, ya que se puede crear una cola, según el orden de entrada, con todas las páginas de la memoria, cuando hay que sustituir una página, se elige la primera de la cola y la que se trae se inserta al final de la cola.
Sufre la **anomalía de Belady** que consiste en que aumentan los fallos de página al aumentar el número de marcos de página para asignación
 - ◆ **OPT o OPTIMO**: se sustituye la página que tardará más en volverse a utilizar. Este algoritmo es irrealizable, ya que el s.o. no sabe a priori que páginas se referenciarán en el futuro



Política de sustitución (3/4)

- ◆ **LRU** se sustituye la página que no ha sido utilizada durante un periodo mayor de tiempo. Es equivalente a aplicar el algoritmo óptimo hacia atrás en el tiempo. Dos posibles realizaciones:
 - ◆ Mediante una pila que mantiene los números de las páginas, cada vez que una página se referencia, su número se elimina de la pila y se coloca en la cumbre de la misma. De esta forma, en la parte superior de la pila se tiene siempre el número de la última página usada y en el fondo el de la página que hace más tiempo que se usó
 - ◆ Mediante registros contadores
- ◆ **Segunda oportunidad o Reloj** es una cola circular, con un puntero que indica cual es la página que se sustituirá a continuación, cuando se necesita un marco de página el puntero avanza hasta que encuentra una página con un bit de referencia a cero, según avanza el puntero se ponen a cero los bits de referencia



Política de sustitución (4/4)

- La política de sustitución puede ser:
 - ◆ **Local:** la página desalojada es del proceso en cuestión
 - ◆ **Global:** la página a desalojar puede ser de cualquier proceso en memoria



Ejercicio 5

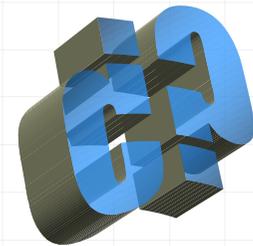
Práctica de política de sustitución

Dada la siguiente cadena de referencia:

0 1 2 3 0 1 4 0 1 2 3 4

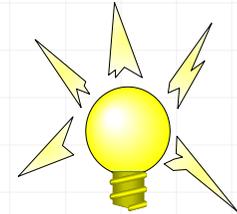
¿Cuántos fallos de página tendrán lugar si se dispone de 4 marcos de página inicialmente vacíos para el algoritmo FIFO?

¿Y si el algoritmo de sustitución fuese el LRU?



Solución 5 (1/4)

Práctica de política de sustitución



Para el algoritmo de sustitución FIFO:

0	1	2	3	3	3	4	0	1	2	3	4
x	0	1	2	2	2	3	4	0	1	2	3
x	x	0	1	1	1	2	3	4	0	1	2
x	x	x	0	0	0	1	2	3	4	0	1

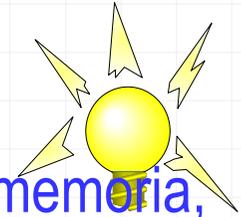
Como se deduce de la figura, inicialmente la memoria está vacía luego las cuatro primeras referencias causan fallos de página que provocan que se traigan las cuatro primeras páginas



Solución 5 (2/4)

Práctica de política de sustitución

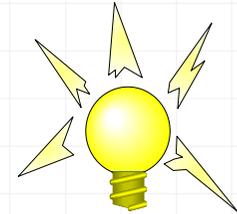
Las dos siguientes páginas que se solicitan, 0 y 1, si se hallan en memoria, por lo que no hay que hacer nada. A continuación, se solicita la página 4, que no está en memoria, produciéndose, de nuevo, un fallo de página, para traer dicha página a memoria hay que buscar, inicialmente, la página a sustituir, dicha página es la de la cabeza de la cola, por ser la más antigua, la página 0. Del mismo modo, las siguientes referencias a páginas también producen fallos de página, porque no se hallan en memoria, las páginas a sustituir serán las de la cabeza de la cola, por ser las más antiguas en memoria. Finalmente, tiene lugar 10 fallos de página



Solución 5 (3/4)

Práctica de política de sustitución

Para el algoritmo de sustitución LRU:



<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>0</u>	<u>1</u>	<u>4</u>	<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
x	0	1	2	3	0	1	4	0	1	2	3
x	x	0	1	2	3	0	1	4	0	1	2
x	x	x	0	1	2	3	3	3	4	0	1

Igual que en el caso anterior, inicialmente la memoria está vacía, luego las cuatro primeras referencias producen fallos de página, por lo que hay que traer esas páginas de memoria



Solución 5 (4/4)

Práctica de política de sustitución



A continuación, se referencian las páginas, 0 y 1, que se hallan en memoria, luego no se producen fallos de páginas, pero se debe mantener el orden en la pila, por lo que cada vez que se referencia una página su número se elimina de la pila y se coloca en la cumbre. De esta forma, en la parte superior de la pila, se tiene siempre el número de la última página referenciada y en la parte inferior la página que hace más tiempo que se usó. Así cuando se referencia la página 4 que no se halla en memoria y por lo tanto se tiene un fallo de página, la página que se sustituye es la de la parte inferior de la pila, en este caso, la página 2. Continuando con este procedimiento finalmente tiene lugar 8 fallos de página



Política de asignación (1/3)

- La relación entre la frecuencia de fallos de página y la cantidad de memoria real asignada a un programa no es lineal corresponde a una **curva de Parocoro**
- Un problema a tener en cuenta si se asignan pocas páginas a los procesos es el de la **hiperpaginación, catástrofe o (thrashing)**
 - ◆ La explicación de esto es que con el alto grado de multiprogramación es imposible que todos los procesos mantengan suficientes páginas en memoria para que no se generen un gran número de fallos de páginas. Esto implica que se satura el canal de transferencia con el disco, que se bloquean muchos procesos esperando un intercambio de páginas y que el procesador queda infrutilizado



Política de asignación (2/3)

- Como política de asignación tenemos **la teoría del conjunto operativo**. Dicho conjunto corresponde al conjunto de páginas referenciadas por el programa durante un intervalo reciente de tiempo, dicho intervalo es una ventana cuyo tamaño depende del diseño. Además, también se utiliza como política de sustitución:

" Ejecutar un proceso solamente si el conjunto de trabajo está por completo en memoria principal, y no eliminar una página que es parte del conjunto de trabajo de un proceso"



Política de asignación (3/3)

- El conjunto operativo esta basado en **el principio de localidad**

Del estudio del comportamiento de los programas, se observa que existe una fuerte tendencia de los programas a favorecer ciertos subconjuntos de sus espacios de direcciones durante la ejecución. A este fenómeno se le conoce como **localidad de referencia**, y puede ser:

- ◆ **Localidad de referencia espacial**: tendencia a referenciar posiciones agrupadas (arrays)
 - ◆ **Localidad de referencia temporal** tendencia a referenciar la misma posición o grupo de posiciones varias veces durante breves intervalos de tiempo (bucles)
- Una **localidad** es un pequeño grupo de páginas no necesariamente adyacentes a las cuales aluden la mayoría de las referencias a memoria durante un periodo de tiempo

