Sistemas Operativos II

Resumen Primera Parte

Emiliano Salvatori

Agosto 2019

1. Clase 2

1.1. Concepto de Programa y sus Estados

Un programa es una entidad pasiva, como el ejemplo en C escrito en clase. Solamente que es compilado y ejecutado. Esto está estático en un disco, cuando se compila se convierte en una entidad activa, ya que se convierte en un proceso.

El concepto de proceso es mucho más que código compilado, tiene muchos más atributos que una simple ejecución de un programa.

Pregunta de examen: realizar un grafo de los estados y explicarlos y tener en cuenta en poner los motivos de las flechas.

1.1.1. Estados

Los tipos de Estados son los siguientes:

- **Nuevo**: El proceso está siendo creado. Es cuando apenas se lanza un programa. El equivalente sería a darle doble click a un lanzador en el Sistema Operativo o cuando se ejecuta por terminal: *./ejecutable*. El proceso está siendo creado.
- **Preparado**: El proceso está a la espera de que le asignen tiempo para ejecutarse en el procesador. Lo que hace el SO es poner el programa en espera para los programas que están listos para correr en el procesador; se pueden visualizar por ejemplo cuando se corre *htop* se listan estos estados de procesos. Que están esperando que el kernel le asigne tiempo para que se corra en el micro.
- **En ejecución**: Es cuando el SO le asigna tiempo dentro del procesador para poder ejecutar el código. El estado significa que se están ejecutando las instrucciones pertenecientes al proceso.
- **Espera**: Es un estado que se le asigna el SO cuando se produce por ejemplo un fallo de página. [1].

  O también se produce cuando realiza operaciones de E/S que tardan mucho tiempo en realizar la petición al hardware. Como esta petición tarda mucho tiempo, el SO pone al proceso en Espera y cuando le es devuelta la petición, lo vuelve a ejecutar. Según la filmina, el proceso está esperando a que se produzca un suceso.
- **Terminado**: Es el estado que el SO asigna a un proceso cuando terminó de ejecutar lo que debía de procesar.

Significado de Interrupción: El sistema operativo le otorga a cada proceso según un quantum un determinado tiempo para ejecutarse dentro del procesador para luego quitarlo si es que no termina. Cuando se le acaba el tiempo para correr, el proceso pasa nuevamente a *preparado*. Importante: Las interrupciones se hacen por tiempo.

Los estados en los que pueda estar un proceso depeden también de lo que esté ejecutando el programa. Por ejemplo: Si se pone un *for* que hace muchas cuentas matemáticas, sólo estará en dos estados: Preparado y Ejecución porque la ALU por sí sola puede realizar estas cuentas matemáticas, sin la necesidad de recurrir a ninguna otra petición externa. En cambio si en ese emphfor se pone un imprimir por pantalla estará más tiempo en el estado de Espera.

[1] Se denomina de esta forma a una excepción arrojada cuando un programa informático requiere una dirección que no se encuentra en la memoria principal actualmente. Recordar que los programas cargan en memoria de a pedazos, cuando no se encuentra ese pedazo en la siguiente instrucción a ejecutar, se produce el fallo de página

1.2. Bloque de control de procesos

Se denomina Bloque de Control de Procesos al conjunto de atributos que tienen los procesos y que NO tiene un programa compilado. Estas características no las poseen los programas estáticos.

Pregunta de examen: liste las caracteristicas del Bloque de Control de Procesos (PCB).

- **Estado**: sería en qué lugar del grafo me encuentro. Qué estado le otorga el SO al proceso.
- **PID** : process ID es como el DNI del proceso.
- **Contador de programa**: puntero a la siguiente instrucción a ejecutar. Esto es indispensable porque el SO cada vez que quita un proceso del procesador porque se le acabó el tiempo de ejecución, requiere guardar las instrucciones en las que se quedó ejecutando, tiene que saber en qué parte de la ejecución quedó pendiente.
- **Registros del CPU**: Al igual que el punto anterior, también al quitar del procesador un proceso es necesario que se guarden las variables que estaba utilizando. Cuando el proceso se guarda porque se le acabó el quantum, debe tomar el registro de todas las variables que tiene el programa y guardarlas y saber en dónde terminó para que cuando se le vuelva a dar tiempo en el procesador saber desde dónde seguir.
- **Información de la planificación de la CPU**: Se tiene información acerca de la prioridad que el SO le da al proceso.
- **Información de gestión de memoria**: son los valores referidos a memoria utilizadad por el proceso, como ser: valor de registros básicos, tablas de paginación o segmentos, etc.
- **Información contable**: cuántos milisegundos corrió el proceso sobre el procesador, por ejemplo. Lleva la cuenta de cuánto tiempo estuvo corriendo el proceso.
- **Información de entrada salida**: qué dispositivo se está uttilizando, por ejemplo si hay un proceso que utiliza la impresora, el SO sabe que no puede asignar a otro proceso mientras el otro lo está utilizando.

1.3. Concurrencia

Dos procesos son concurrentes entre si, sólo si es que son independientes. Es decir, son independientes entre si cuando no importa el orden en que se ejecute, el resultado es siempre el mismo. Por ejemplo: si hay un proceso que imprime "HOLA.^{o}tro proceso imprimie "PERROτ se corren al mismo tiempo, son independientes porque no importa el orden el resultado siempre es el mismo. El resultado de un proceso NO depende del resultado del otro.

Para que exista independencia en la ejecución de procesos, estos deben tener comunicación entre si. Por ejmplo cuando se visita la página de Youtube y se reproduce un video. Un proceso está encargado de ir llenando el buffer del video mientras que otro proceso está pendiente de si se aprieta play para poder comenzar a poner en secuencia las imágenes. El primero de ellos siempre va adelante y siempre haciendo peticiones de E/S a la memoria asignada como buffer, manejándose de forma independiente.

Para el usuario final, el resultado es el mismo, se ve el video. Pero los procesos que atienden estos servicios NO son lo mismo, son varios procesos intercomunicados entre si brindando un mismo servicio. El programa en este caso es un sistema concurrente. [2]

Las cuestiones a tener en cuenta a la hora de diseñar procesos concurrentes:

- Comunicación entre procesos.
- Compartición y competencia por los recursos.
- Sincronización de la ejecución de varios procesos.
- Asignación del tiempo del procesador a los procesos.

[2] Dos o más procesos decimos que son concurrentes, paralelos, o que se ejecutan concurrentemente, cuando son procesados almismo tiempo, es decir, que para ejecutar uno de ellos, no hace falta que se haya ejecutado otro.

1.4. Paralelismo

Capacidad de ejecutar procesos de manera paralela. Está asociado al hardware. La única manera de ejecutar dos procesos A y B a la vez, es que lo permita el hardware. Se tiene cuando hay mas de un cause por los que se puede ejecutar un proceso, y esto lo dispone el hardware. Antes tenían procesadores con un solo cause, se corría un solo proceso dentro de un solo procesador. En los 80 se inventaron varios causes y al día de hoy los causes se duplicaron.

Recordar que **Paralelismo NO implica concurrencia**. Si se tiene paralelismo (con hardware adecuado) se va a necesitar la capacidad de correr procesos concurrentes. **Pero para que haya Concurrencia es necesario que haya Paralelismo**

1.5. Gestión de procesos

Multiprogramación: tener un solo cause y dividir en tajadas de tiempo varios programas y tener la sensación de que sucede todo a la vez. Se denomina multiprogramación a una técnica por la que dos o más procesos pueden alojarse en la memoria principal y ser ejecutados concurrentemente por el procesador o CPU.

Hay que tener en cuenta que los procesos van intercambiandose entre si un tiempo el proceso P1, otro tiemp el P2 y puede haber un P3 que lo utilice otro tiempo. Lo que NO puede pasar es que se superpongan los Procesos.

Multiprocesamiento: es tener varios causes. Se tiene varios procesadores para varios procesos. Multiprocesamiento o multiproceso es el uso de dos o más procesadores (CPU) en una computadora para la ejecución de uno o varios procesos (programas corriendo). Si se tiene multiprocesador se puede tener por ejemplo a P1 y P2 corriendo en simultaneo (dependiendo de cuántos cause tenga el procesador).

Procesamiento distribuido: Consiste en la gestión de varios procesos, ejecutándose en sistemas de computadoras múltiples y distribuidos. Es la idea de cluster.

La computación distribuida o informática en malla (grid) es un modelo para resolver problemas de computación masiva utilizando un gran número de ordenadores organizados en clústeres incrustados en una infraestructura de telecomunicaciones distribuida.

1.6. Función Fork

La función fork(), se utiliza para crear un nuevo proceso duplicando el proceso de llamada actual, siendo el proceso recién creado conocido como **proceso hijo** y el proceso de llamada actual conocido como **proceso padre**. Entonces podemos decir que fork () se usa para crear un proceso secundario al proceso de llamada.

Con esta instrucción, el cause de ejecución no es de arriba hacia abajo como en un programa simple realizado en C, sino que se tiene un cause distinto, bifurcado. Se puede correr de forma concurrente. Se tene dos procesos simultaneamente por ejemplo:

Como se puede ver *fork()* lo que hace es crear un proceso nuevo, es decir un proceso hijo pero dentro de la ejecución del proceso padre. Lo que hace el SO cuando encuentra esta instrucción es duplicar el espacio de memoria del padre, compartiendo todas las variables. Copia toda la memoria estática al proceso hijo.

Se debe recordar que Minix tiene una tabla de procesos que sólo puede albergar 100 procesos (de 0 a 99). Si se llena esa tabla, el fork() da error, simbolizando su valor de retorno con un -1. Solo puede crear 99 procesos nada más. El fork() da -1 y sale con error y este error lo toma el SO en el cause del proceso que lo invocó.

Con el fork no se puede saber cuándo un proceso se ejecuta primero. Por lo que se puede utilizar la función *sleep()* en el del hijo. Para dormir al hijo y que ejecute el hijo y no el padre.

Pregunta de examen: ¿existe otro estado que no sea el de *terminado*? Este estado posible se denomina *Estado Zombie*, que es cuando el hijo no tiene nada que ejecutar y está esperando que el padre muera o al revés. En sistemas operativos Unix, un proceso zombi o "defunct"(difunto) es un proceso que ha completado su ejecución pero aún tiene una entrada en la tabla de procesos, permitiendo al proceso que lo ha creado (padre) leer el estado de su salida. Metafóricamente, el proceso hijo ha muerto pero su .ªlma.ªún no ha sido recogida.

Preguna de examen: Escriba un código en C donde un proceso A cree un proceso hijo B. Esto sería similar a poner el *ejemplo5.c* citado anteriormente.

Se debe tener en cuenta que el PID puede tener 3 valores:

- Un número mayor que cero.
- Un número igual a cero. Puede ser cero por ejemplo para el PID del hijo de un proceso padre
- Un número igual a -1.

1.7. Hilos y procesos multihilo

Proceso padre tiene un espacio de memoria asociado y luego tiene un código. En el espacio tiene la pila, los datos, el contador de programa. Y tiene un solo cause. Cuando se hace invoca a la función fork(), se crea un hijo idéntico, lo que cambia solo es el PID. Por lo que es costoso en tiempo y en recurso invocar un fork().

Un proceso de Unix es cualquier programa en ejecución y es totalmente independiente de otros procesos. El comando de Unix *ps* nos lista los procesos en ejecución en nuestra máquina. Un proceso tiene su propia zona de memoria y se ejecuta "simultáneamente.ª otros procesos

La idea de hilo es agarrar el procso y utilizar el mismo tamaño de memoria de un solo proceso pero hacer correr varias hebras. Esto tiene muchos más beneficios que desdoblar un proceso en dos, ya que los hilos comparten espacio de memoria. Los hilos ya de por sí comparten espacio de salida. Cada hebra corre por distintos causes, si una se bloquea el proceso sigue consumiendo tiempo en el procesador ejecutando otros hilos y no saca el proceso de su ejecución como SI sucedería con un proceso hijo y padre.

Dentro de un proceso puede haber varios hilos de ejecución (varios threads). Eso quiere decir que un proceso podría estar haciendo varias cosas "a la vez". Los hilos dentro de un proceso comparten todos la misma memoria. Eso quiere decir que si un hilo toca una variable, todos los demás hilos del mismo proceso verán el nuevo valor de la variable. Esto hace imprescindible el uso de semáforos o mutex (EXclusión MUTua, que en inglés es al revés, funciones *pthread_mutex*) para evitar que dos threads accedan a la vez a la misma estructura de datos. También hace que si un hilo "se equivocaτ corrompe una zona de memoria, todos los demás hilos del mismo proceso vean la memoria corrompida. Un fallo en un hilo puede hacer fallar a todos los demás hilos del mismo proceso.

Conclusión: Un proceso es, por tanto, más costoso de lanzar, ya que se necesita crear una copia de toda la memoria de nuestro programa. Los hilos son más ligeros.

Ventajas de la programación multihilo:

- **Capacidad de respuesta**: permite que un programa continúe ejecutándose incluso aunque parte de él esté bloqueado.
- **Compartición de recursos**: por omisión, las hebras comparten la memoria y los recursos del proceso al que pertenecen.
- **Economía**: asignar memoria y recursos para crear procesos es costosa, es más económico realizar cambios de contexto entre hebras.
- **Multiprocesador**: se pueden ejecutar en paralelo en diferentes procesadores (un proceso monohebra solo se puede ejecutar en un sólo microprocesador).

1.8. Instancias de Kernel

Modelo Muchos a Uno: Otra forma es que muchas hebras llaman a una sola instanacia de kernel. Es un costoso cuando hay muchas hebras. Es eficiente pero el proceso completo se bloquea si una hebra realiza una llamada bloqueante al sistema.

Modelo Uno a uno: Asigna cada hebra de usuario a una de Kernel. Todo el tiempo se hacen solicitudes al kernel (como por ejemplo imprimir en pantalla). Cuano se tienen hebras se pueden una instancia de kernel que le responda a cada hebra. Cada hebra hace una solicitud a una instanaica del kernel. Esto se denomina Uno a Uno: un hilo solicita una instancia de Kernel. Este modelo proporciona mayor concurrencia que el anterior, se permiten herbas bloqueantes mientras se ejecutan otras. El incovenviente de ello es que crear una hebra de usuario requiere crear su correspondiente hebra de kernel por lo que repercute en el rendiemiento

Modelo muchos a muchos: La solución que se da es un híbrido entre ambas: muchos hilos solicitan al kernel una instancia en particular y el kernel evalúa cuándo crear una instancia para cada hebra. Multiplexa muchas hebras de usuario sobre un número menor o igual de hebras de Kernel. Se pueden crear tantas hebras de usuario como sea necesario y las correspondientes hebras del Kernel pueden ejecutarse en paralelo en un multiprocesador. Cuando una hebra realiza una llamada bloqueante, el kernel planifica otra hebra.

1.9. Hilos en modo Usuario

Los hilos en modo usuario tienen determinadas ventajas:

- El núcleo no sabe que existen.
- Tabla de subprocesos probada para cambios de contexto.
- Cambio de contexto mucho mas rápido entre hilos (no se pasa al kernel).

- Cada proceso puede tener su algoritmo de planificación.

Inconvenientes:

- Llamadas bloqueantes al sistema.
- Fallos de página.
- Tienen que ceder la CPU entre ellos: conmutación en el mismo proceso.
- Precisamente queremos hilos en procesos con muchas E/S para obtener paralelismo, es decir que se están bloqueando muy frecuentemente.

1.10. Hilos en Modo Kernel

Ventajas:

- El núcleo mantiene la tabla de hilos, que es un subconjunto de la de procesos.
- Las llamadas bloqueantes no necesitan funciones especiales.
- Los fallos de página no suponen un problema.
- Al bloquearse un hilo, el núcleo puede conmutar a otro hilo de otro proceso.

Inconvenientes

- Las llamadas bloqueantes son llamadas al sistema.
- La creación y destrucción de procesos es mas costoso, por lo que se trata de reutilizar hilos.

1.11. Diferencia entre procesos e hilos

Creación:

- **Procesos**: son costosos para crear
- **Hilos**: son bastante ligeros

Recursos(memoria):

- **Procesos**: Independientes
- **Hilos**: Compartidos

Comunicación:

- **Procesos**: compleja
- **Hilos**: Sencilla

Cambio por Sistema operativo:

- **Procesos**: Muy lento
- **Hilos**: Rápido

Proramación:

- **Procesos**: Reducida
- **Hilos**: Alta