

Sistemas Operativos - Apuntes para final

Gianfranco Zamboni

27 de septiembre de 2020

Índice

1. Introducción	5
1.1. El sistema operativo	5
2. Procesos y API	6
2.1. Creación y control de un proceso	6
2.1.1. Creación de un proceso	6
2.1.2. Terminación	7
2.1.3. Un poco más sobre las syscalls mencionadas	7
2.2. Estados de un proceso	8
2.3. Process Control Block (PCB)	8
2.4. Scheduler	9
2.4.1. Context Switching	9
2.4.2. Colas de procesos	10
2.5. Inter Process Communication (IPC)	10
2.5.1. Pasaje de mensaje	11
2.5.2. Sockets	12
2.5.3. Pipes	12
2.6. E/S bloqueante / no bloqueante	12
2.7. Manejo básico de un shell Unix	13
2.7.1. File descriptors:	13
2.7.2. Comandos de consola	13
3. Scheduling	14
3.1. Objetivos de la política de scheduling	14
3.2. Scheduling con y sin desalojo	16
3.3. Políticas de scheduling	16
3.4. Scheduling para tiempo real y para SMP	16

4. Sincronización entre procesos	16
4.1. Contención	16
4.2. Condiciones de carrera	16
4.3. Secciones críticas	16
4.4. TestAndSet	16
4.5. Busy waiting / sleep	16
4.6. Productor - Consumidor	16
4.7. Semáforos	16
4.8. Introducción a deadlock	16
4.9. Monitores	16
4.10. Variables de condición	16
5. Programación concurrente	16
5.1. Algoritmos wait-free	16
5.2. Algoritmos lock-free	16
5.3. CAS	16
5.4. ABA	16
5.5. Programación de multicores	16
5.6. Invalidación de caché	16
5.7. Reorden de instrucciones	16
6. Administración de memoria	16
6.1. Segmentación	16
6.2. Paginación	16
6.3. Swapping	16
6.4. MMU	16
6.5. Memoria virtual	16
6.6. Copy-on-write	16
6.7. Algoritmos de reemplazo de páginas	16
7. Administración de entrada/salida	16
7.1. Polling, interrupciones, DMA	16
7.2. Almacenamiento secundario	16
7.3. Drivers	16
7.4. Políticas de scheduling de E/S a disco	16
7.5. Gestión del disco (formateo, booteo, bloques dañados)	16
7.6. RAID	16
7.7. Copias de seguridad	16
7.8. Spooling	16
7.9. Clocks	16

8. Sistemas de archivos	16
8.1. Responsabilidades del FS	16
8.2. Punto de montaje	16
8.3. Representación de archivos	16
8.4. Manejo del espacio libre	16
8.5. FAT, inodos	16
8.6. Atributos	16
8.7. Directorios	16
8.8. Caché	16
8.9. Consistencia, journaling	16
8.10. Características avanzadas	16
8.11. NFS, VFS	16
9. Protección y seguridad	16
9.1. Conceptos de protección y seguridad	16
9.2. Matrices de permisos	16
9.3. MAC vs. DAC	16
9.4. Autenticación, autorización y auditoría	16
9.5. Funciones de hash de una vía	16
9.6. Encriptación simétrica	16
9.7. RSA	16
9.8. Privilegios de procesos	16
9.9. Buffer overflows	16
9.10. Inyección de parámetros	16
9.11. Condiciones de carrera	16
9.12. Sandboxes	16
9.13. Principios generale de seguridad	16
10. Sistemas distribuidos	16
10.1. Taxonomía de Flynn	16
10.2. Arquitecturas de HW y SW para sistemas distribuidos	16
10.3. RPC	16
10.4. Threads	16
10.5. Pasaje de mensajes	16
10.6. Orden parcial entre eventos	16
10.7. Livelock	16
10.8. Acuerdo bizantino	16
10.9. Intuición de safety, liveness, fairness	16
10.10 Algoritmo del banquero	16
10.11 Panadería de Lamport	16
10.12 Modelos de fallas y métricas de complejidad	16

10.13	Exclusión mutua y locks distribuidos	16
10.14	Elección de líder	16
10.15	Instantánea global consistente	16
10.16	2PC	16
11.	Conceptos avanzados	16
11.1.	Virtualización	16
11.2.	Contenedores	16
11.3.	Cloud computing	16

1. Introducción

Un sistema informático tiene cuatro componentes:

- El hardware (CPU, memoria y dispositivos de entrada salida) proveen los recursos básicos del sistema.
- Las aplicaciones definen la forma en que estos recursos va a ser usados para resolver los problemas del usuario.
- El sistema operativo controla el hardware y coordina su uso entre las distintas aplicaciones de los usuarios.

1.1. El sistema operativo

Un sistema operativo provee un entorno de ejecución de programas. Para esto tiene ciertos componentes que difieren de sistema en sistema pero que entre lo más comunes se encuentran:

- **Drivers:** Programas manejan los detalles de bajo nivel relacionados con la operación de los distintos dispositivos.
- **Núcleo:** (o Kernel) Es el sistema operativo, propiamente dicho. Se encarga de las tareas fundamentales y contiene los diversos sub-sistemas que iremos viendo a lo largo de la materia.
- **Sistema de archivos:** Forma de organizar los datos en el disco para gestionar su acceso, permisos, etc.
 - **Archivo:** Secuencia de bits con un nombre y una serie de atributos que indican permisos.
 - **Binario del sistema:** Son archivos que no forman parte del kernel pero suelen llevar a cabo tareas muy importantes o proveer las utilidades básicas del sistema.
 - **Archivos de configuración:** Son archivos especiales con información que el sistema operativo necesita para funcionar.
 - **Directorio:** Colección de archivos y directorios que contiene un nombre y se organiza jerárquicamente.
 - **Directorios del sistema:** Son directorios donde el propio SO guarda archivos que necesita para su funcionamiento.
 - **Dispositivo virtual:** Abstracción de un dispositivo físico bajo la forma, en general, de un archivo de manera tal que se pueda abrir, leer, escribir, etc.
 - **Usuario:** La representación, dentro del propio Sistema Operativo, de las personas o entidades que pueden usarlo. Sirve principalmente como una forma de aislar información entre distintos usuarios reales y de establecer limitaciones.
 - **Grupo:** Una colección de usuarios

Systems calls (Syscalls): Son rutinas que proveen una interfaz a los servicios disponibles en el sistema. Generalmente, son rutinas escritas en C/C++ y llamarlas implican un cambio de contexto y privilegios del proceso a ser ejecutado. Algunos de estos servicios son:

- Creación y control de procesos.
- Pipes.
- Señales
- Operaciones de archivos y directorios
- Excepciones
- Errores del bus
- Biblioteca C

2. Procesos y API

Proceso: Es un programa en ejecución y todas las estructuras que debe mantener el sistema para su correcto funcionamiento. Entre ellas, los valores de los registros, el program counter, el stack del proceso, el área de memoria en la que se cargan las instrucciones a ejecutar (el programa propiamente dicho) y el área de datos (contiene variables globales) y un heap de memoria (que es donde reserva memoria dinámica).

2.1. Creación y control de un proceso

2.1.1. Creación de un proceso

Durante su ejecución, un proceso puede crear nuevos procesos. En este caso llamamos **proceso padre** al proceso que crea nuevos procesos y **procesos hijos** a los procesos creados. Además, cada proceso hijo podrá crear otros procesos, formando así un **árbol** de procesos.

En la mayoría de los sistemas operativos, se asigna a cada proceso un identificador numérico único llamado **process id** o **pid** que sirve como índice para poder acceder a varios de sus atributos.

El nuevo proceso hijo creado necesitara de ciertos recursos para poder llevar a cabo su objetivo. Estos recursos podrán ser obtenidos directamente del sistema operativo o estar restringidos a los recursos del proceso padre.

Cuando un proceso crea un nuevo proceso (**fork()**) hay dos posibilidades:

- El padre y su hijo continúan ejecutando concurrentemente.
- El padre espera a que alguno o todos sus hijos terminen para seguir ejecutando (**wait()**).

Además, el nuevo proceso puede ser un duplicado de su padre o cargar un nuevo programa (**exec()**)

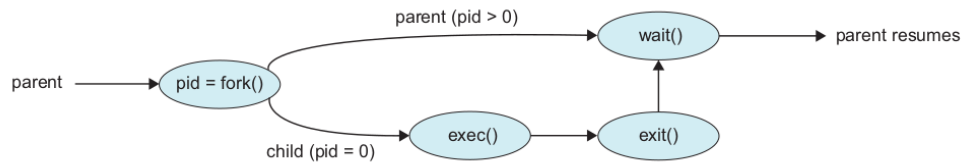


Figure 3.10 Process creation using the `fork()` system call.

2.1.2. Terminación

Un proceso termina cuando ejecuta su última instrucción y pide al sistema operativo que lo borre usando la syscall `exit()`. En este momento, el proceso envía el código de estado con el que terminó a su padre y el sistema operativo libera todos los recursos que le había asignado.

La terminación puede ocurrir por varios motivos. Un proceso puede causar la terminación de otro usando las syscalls adecuadas (usualmente, esto puede hacerlo solo el proceso padre).

2.1.3. Un poco más sobre las syscalls mencionadas

- `pid_t fork()`: Crea un nuevo proceso. En el caso del creador se retorna el PID del hijo. En el caso del hijo, retorna 0.
- `int execve(const char* filename, char* const argv[], char* const envp[])`: Sustituye la imagen de memoria del programa por la del programa ubicado en filename.
- `pid_t forkv()`: Crea un hijo sin copiar la memoria del padre, el hijo tiene que hacer `exec`. Un proceso creado con esta syscall comienza con sus páginas de memoria apuntando a las mismas que las de su padre. Recién cuando alguno escriba en memoria, se hace la copia. Esto se llama **Copy-On-Write**.
- `pid_t wait(int* status)`: Bloquea al padre hasta que algún hijo termine o hasta que alcance el status indicado.
- `pid_t waitpid(pid_t pid, int* status)`: Igual al anterior pero espera a que el hijo con PID `pid` llegue a ese status.
- `void exit(int status)`: Finaliza el proceso actual.
- `clone(...)`: Crea un nuevo proceso. El hijo comparte parte del contexto con el padre. Es usado en la implementación de threads.

Fork Bomb: Un proceso crea infinitos hijos.

2.2. Estados de un proceso

Durante su ejecución, un proceso va modificando su estado acorde al siguiente diagrama:

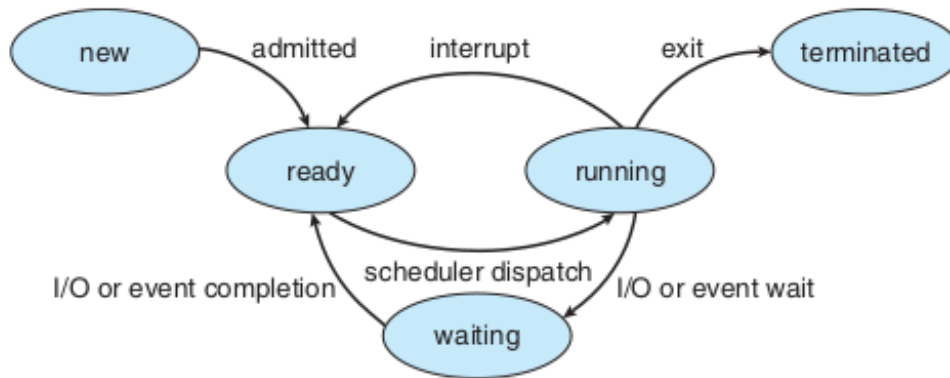


Figure 3.2 Diagram of process state.

- **Nuevo (New):** El proceso se está creando.
- **Listo (Ready):** El proceso está listo para ser ejecutado.
- **Bloqueado (Waiting):** El proceso está esperando a que ocurra algo (por ejemplo a que se complete una operación de entrada salida)
- **Ejecutando (Running):** El proceso se está ejecutando.
- **Terminado (Terminated):** El proceso terminó de ejecutarse.

Es importante notar que solo un proceso puede estar **ejecutando** en un instante de tiempo de dado. Sin embargo, muchos procesos pueden estar bloqueados o listos simultáneamente.

2.3. Process Control Block (PCB)

Cada proceso está representado, en el sistema operativo, por un Process Control Block o Task Control Block. El mismo contiene información asociada a cada proceso, entre ellas:

- El estado del proceso
- El program counter (la dirección de memoria de la próxima instrucción a ser ejecutada)
- Los registros del CPU,
- Información de Scheduling, como puede ser prioridad de ejecución
- Información de manejo de memoria, como las direcciones de los directorios de paginas o las tablas de segmentación.

- Estadísticas de uso del sistema
- Información de E/S (que dispositivos está usando, lista de archivos abiertos, etc)

2.4. Scheduler

Multiprocesador: Un equipo con más de un procesador.

Multiprogramación: La capacidad de un SO de tener varios procesos en ejecución.

Multiprocesamiento: Se refiere al tipo de procesamiento que sucede en los multiprocesadores.

Multitarea: Es una forma especial de multiprogramación, donde la conmutación entre procesos se hace de manera tan rápida que da la sensación de que varios programas están corriendo en simultáneo.

Multithreaded: Son procesos en los cuales hay varios *mini procesos* corriendo en paralelo (de manera real o ficticia).

2.4.1. Context Switching

Una interrupción hace que el sistema operativo cambie la tarea de una CPU y corra un rutina del kernel. Cuando esto ocurre, el sistema, debe guardar el **contexto** del proceso que se está ejecutando para poder retornar a la ejecución del mismo cuando termine de atenderla. Este contexto es el PCB del proceso.

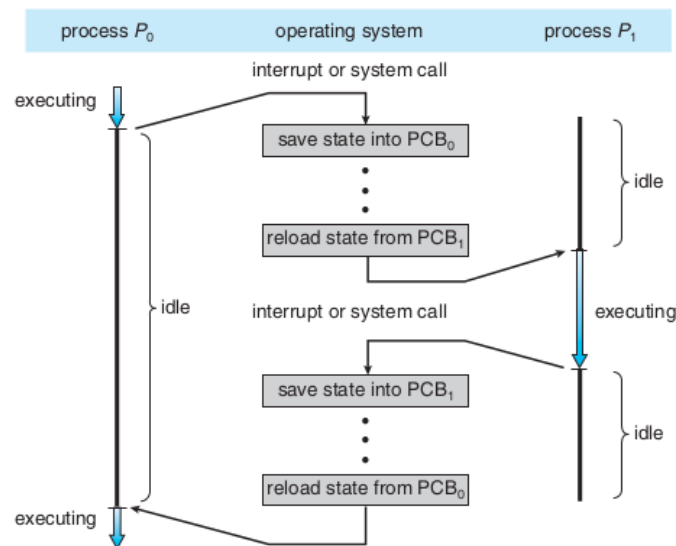


Figure 3.4 Diagram showing CPU switch from process to process.

2.4.2. Colas de procesos

Una de las características de los sistemas operativos actuales es que nos permiten correr varias aplicaciones a la vez, sin embargo solo un proceso puede estar ejecutando en la CPU. Para darnos la sensación de simultaneidad, el SO hace lo que se llama **preemption**.

Preemption: A cada proceso le asigna un **quantum** o una cantidad de tiempo durante la cual es ejecutado. Una vez que transcurrido ese tiempo, se guarda el contexto del proceso en memoria y se carga el contexto del próximo proceso a ejecutar.

El objetivo de la multiprogramación es intercambiar las tareas que se ejecutan en la CPU lo suficientemente rápido como para que el usuario pueda interactuar con cada programa como si estuviesen ejecutándose simultáneamente. Para esto existe un proceso especial llamado **Scheduler** que selecciona el próximo proceso que debe ser ejecutado.

El mismo dispone de varias colas de procesos en los que se encuentran los PCB de los procesos en ejecución. Los PCB se van encolando y desencolando de cada cola dependiendo del estado de cada proceso.

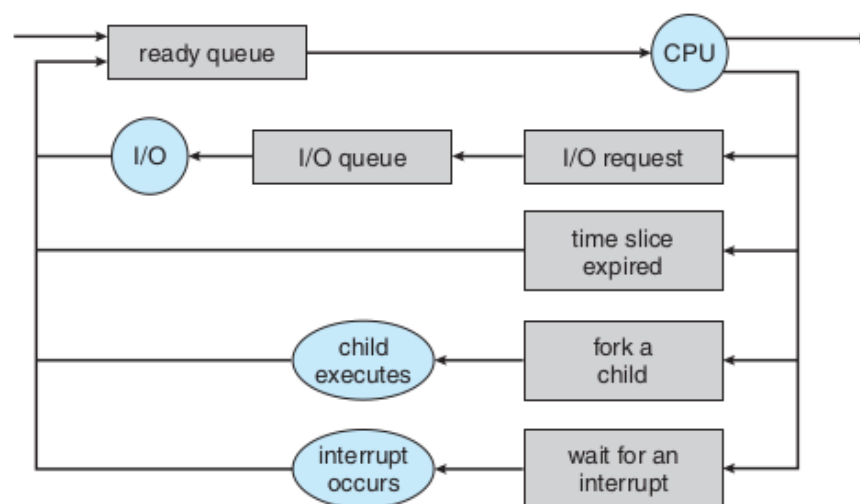


Figure 3.6 Queueing-diagram representation of process scheduling.

2.5. Inter Process Communication (IPC)

Los procesos que se están ejecutando concurrentemente en el sistema operativa pueden ser **procesos independientes** o **cooperativos**.

Procesos independiente: Es un proceso que no puede afectar o ser afectado por otros procesos ejecutándose en el sistema operativo.

Procesos cooperativos: Son procesos que no son independientes (pueden afectar o ser afectados por otros procesos).

Los procesos cooperativos requieren de un mecanismo de comunicación que les permita intercambiar datos e información. Hay dos modelos fundamentales para hacer esto: **memoria compartida** y **Pasaje de mensajes**.

En el modelo de memoria compartida, se establece un área de memoria que es compartida entre procesos. Los procesos pueden intercambiar información leyendo y escribiendo en este área.

En el modelo de pasaje de mensajes, la comunicación se realiza a través de mensajes entre los procesos cooperativos.

2.5.1. Pasaje de mensaje

Los mensajes proveen a los procesos un mecanismo que permite que los procesos se comuniquen entre sí y sincronicen sus acciones. Un sistema que permite el pasaje de mensajes provee al menos dos operaciones:

`send(message)` `receive(message)`

Los mensajes pueden tener una longitud fija o variable. La primera opción es más fácil de implementar pero hace que la programación de tareas sea más tediosa. Los mensajes de longitud variable necesitan un sistema más complejo pero permite más agilidad a la hora de programar tareas.

Si un proceso P y Q se quieren comunicar, entonces debe existir un **link de comunicación** entre ellos. Para crear este link, se deben tener en cuenta el tipo de comunicación que se desea ofrecer:

■ Direcccionamiento

- **Conexión directa:** Cada proceso debe explicitar el nombre del proceso destinatario/receptor:

- `send(P, message)` - Envía un mensaje al proceso P
- `receive(Q, message)` - Recibe un mensaje del proceso Q

En este caso, el link de comunicación se establece automáticamente entre cada par de procesos que quiere comunicarse. Y cada proceso sabe la identidad del otro.

- **Comunicación indirecta:** Los mensajes son enviados a buzones (**mailboxes**) o puertos (**puertos**). Cada buzón tiene un identificador único. Un proceso puede comunicarse con otro a través de varios buzones.

- `send(A, message)` - Envía un mensaje al buzón A
- `receive(A, message)` - Recibe un mensaje del buzón A

En este esquema, un link se establece entre dos procesos solo si comparten un buzón. Además un link puede estar asociado a varios pares de procesos e incluso puede haber varios links entre dos mismos procesos.

- **Sincronización:** El pasaje de mensajes puede ser bloqueante (síncrono) o no bloqueante (asíncrono):
 - **Envío bloqueante:** El proceso que envía un mensaje espera a que el mismo sea recibido por su destinatario.
 - **Envío no bloqueante:** El proceso envía el mensaje y sigue su ejecución.
 - **Recepción bloqueante:** El proceso se bloquea hasta que recibe un mensaje.
 - **Recepción no bloqueante:** El receptor recibe un mensaje válido o null.
- **Buffering:** Los mensajes intercambiados entre procesos deben almacenarse en una cola temporal. Estas colas pueden ser de tres tipo:
 - **Capacidad Cero:** El link no puede tener mensajes en espera. El remitente debe bloquearse hasta que el mensaje sea recibido.
 - **Capacidad acotada:** La cola tiene una capacidad finita n por lo que puede haber a lo sumo n en espera. Si la cola no está llena se puede enviar un mensaje, sino el remitente debe bloquearse hasta que se que haya espacio disponible.
 - **Capacidad infinita:** El remitente nunca se bloquea.

2.5.2. Sockets

Los sockets son los extremos de una comunicación que usan dos procesos para comunicarse a través de una red. Cada socket está identificado por una dirección IP y un número de puerto.

En general, se usa una arquitectura cliente-servidor: El servidor espera a que un cliente haga un pedido y, una vez que lo recibe, acepta la conexión del socket del cliente para completar la conexión. El número del socket del servidor, en general, va a ser menor o igual a 1024. Estos puertos son los puertos conocidos (*well known ports*) e implementan

Cuando el proceso cliente inicia la conexión, la computadora que lo está ejecutando le asigna un puerto arbitrario cuyo número es mayor a 1024.

2.5.3. Pipes

Un **pipe** es un canal que provee una de las formas más simples de comunicación entre dos procesos aunque tienen sus limitaciones. Al implementar un pipe, se debe tener ciertas consideraciones:

1. ¿El pipe permite comunicación bidireccional o unidireccional?
2. Si es bidireccional, es **half-duplex** (para mandar información se debe esperar a que el otro extremo del pipe termina de hacerlo) o **full-duplex** (la información puede viajar de un lado a otro y viceversa simultáneamente).
3. ¿Debe existir alguna relación entre los procesos que se están comunicando (por ejemplo, padre-hijo)?

4. ¿Los procesos se van a poder comunicar dentro de una red o tienen que estar en la misma maquina?

Ordinary pipes: Permiten que dos procesos se comuniquen en modo servidor-consumidor: El servidor escribe en un extremo del pipe (extremo de escritura) y el consumidor lee desde el otro extremo (extremo de lectura). Estos pipes son unidireccionales. Si se necesita una comunicación bidireccional se debe crear otro pipe que permita mandar datos en la otra dirección.

Este tipo de pipes requieren que los procesos que se comunican tengan una relación padre-hijo y dejan de existir una vez que la comunicación termina.

Named pipes: Proveen comunicación bidireccional y no necesitan que los procesos estén relacionados. Una vez que es establecido, varios procesos pueden usarlo para comunicarse y continúan existiendo incluso después de que las comunicaciones hayan finalizado.

2.6. E/S bloqueante / no bloqueante

Cundo un proceso necesita escribir o leer información de algún dispositivo necesita realizar operaciones de **entrada/salida**

Estas operaciones son muy lentas por lo que quedarse bloqueado es un desperdicio de tiempo.

Busy Waiting: El proceso no hace nada pero no libera el CPU. Se gastan ciclos de procesamiento en hacer nada.

Para evitar esto se utilizan algunas técnicas que permiten al SO operativo seguir ejecutando mientras espera la respuesta de los dispositivos:

- **Polling:** El proceso libera la CPU pero todavía recibe un quantum que desperdicia hasta que la E/S esté terminada.
- **Interrupciones:** Esto permite la multiprogramación. El SO no le otorga más quantum al proceso hasta que su E/S esté lista. El hardware comunica esto mediante una interrupción que hace que el proceso se despierte.

2.7. Manejo básico de un shell Unix

2.7.1. File descriptors:

En Unix, cada proceso se crea con una tabla que le permite identificar cuales son los archivos que tiene abiertos. Cada índice es un **file descriptors** que usa el Kernel para saber como leer/escribir datos en los distintos archivos (en Unix, el teclado y la pantalla se modelan como archivos).

Además, cada proceso hereda de su proceso padre tres archivos abiertos que ocupan los file descriptors 0, 1 y 3 y representan la entrada estándar (**stdin**), la salida estandar (**stdout**) y el error estandar (**stderr**), respectivamente.

Linux provee de dos llamadas al sistema que nos permiten leer/escribir a un archivo usando su file descriptor **fd**:

```
ssize_t read(int fd, void *buf, size_t count)
ssize_t write(int fd, const void *buf, size_t count);
```

Aquí **buf** es un puntero a donde se almacenan los datos a leer o escribir y **count** la cantidad de bytes que hay escribir/leer.

2.7.2. Comandos de consola

- **echo** escribe lo que le pasemos como parámetro en su **stdout**.

```
echo .Esto es un mensaje"
```

- > Se le indica a la consola que el **stdout** se redirija a un archivo:

```
echo .Esto es un mensaje» mensaje.txt
```

- | Redirige el **stdout** de un proceso hacia el **stdin** de otro:

```
echo .Esto es un mensaje" | wc -c
```

En este caso, el primer proceso ejecuta el comando **echo** que imprime en **stdout** el mensaje **.^Esto es un mensaje"**. El segundo proceso recibe por **stdin** lo que se escribió en el **stdout** del proceso que ejecutó **echo**

3. Scheduling

3.1. Objetivos de la política de scheduling

- La política de scheduling es una de las principales huellas del sistema operativo.
-

- 3.2. Scheduling con y sin desalojo
- 3.3. Políticas de scheduling
- 3.4. Scheduling para tiempo real y para SMP

4. Sincronización entre procesos

- 4.1. Contención
- 4.2. Condiciones de carrera
- 4.3. Secciones críticas
- 4.4. TestAndSet
- 4.5. Busy waiting / sleep
- 4.6. Productor - Consumidor
- 4.7. Semáforos
- 4.8. Introducción a deadlock
- 4.9. Monitores
- 4.10. Variables de condición

5. Programación concurrente

- 5.1. Algoritmos wait-free
- 5.2. Algoritmos lock-free
- 5.3. CAS
- 5.4. ABA
- 5.5. Programación de multicores
- 5.6. Invalidación de caché
- 5.7. Reorden de instrucciones

6. Administración de memoria

- 6.1. Segmentación
- 6.2. Paginación
- 6.3. Swapping
- 6.4. MMU
- 6.5. Memoria virtual
- 6.6. Copy-on-write
- 6.7. Algoritmos de reemplazo de páginas

7. Administración de entrada/salida