

TEMA 18

SISTEMAS OPERATIVOS: GESTIÓN DE ENTRADAS/SALIDAS

ÍNDICE

1. INTRODUCCIÓN
2. PRINCIPIOS HARDWARE EN ENTRADAS/SALIDAS
 - 2.1. Dispositivos de entrada/salida
 - 2.2. Controladores de dispositivo
 - 2.3. Puertos de entrada/salida
 - 2.4. Acceso directo a la memoria (DMA)
 - 2.5. Canales de entrada/salida
3. PRINCIPIOS DEL SOFTWARE DE ENTRADA/SALIDA
 - 3.1. Manejadores de interrupciones
 - 3.2. Manejadores de dispositivos
 - 3.3. Software de entrada/salida independiente del dispositivo
 - 3.4. Software del usuario
4. ENTRADAS/SALIDAS CONTROLADAS POR PROGRAMAS
5. INTERRUPCIONES
6. SPOOL
7. BIBLIOGRAFÍA

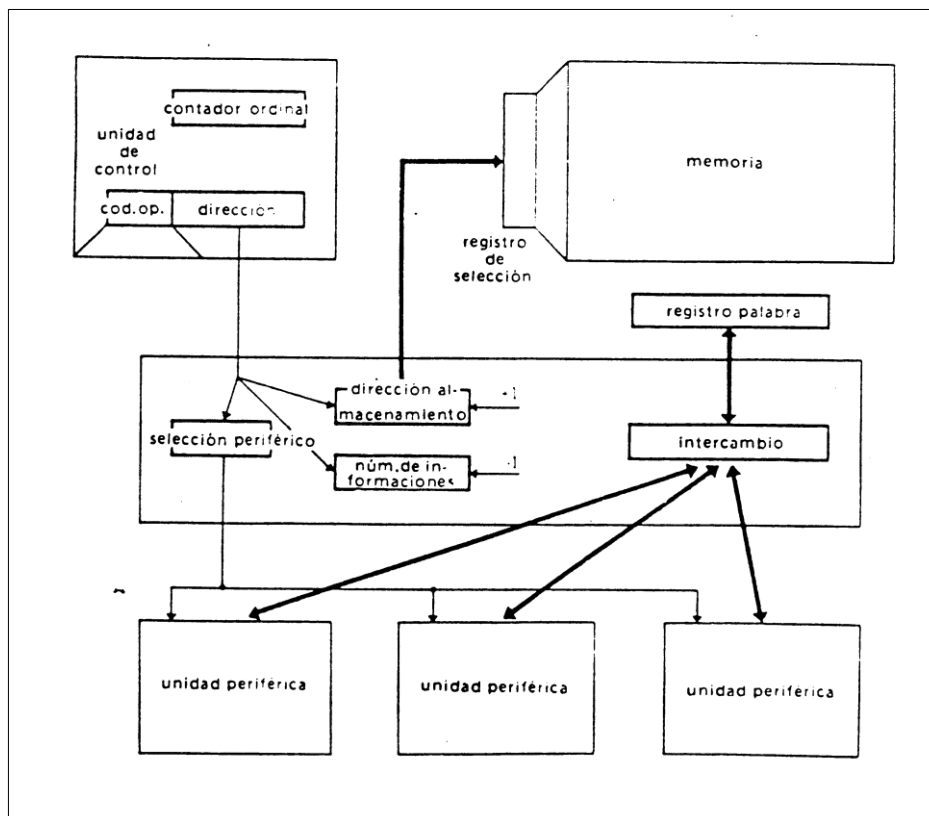
1. INTRODUCCIÓN

Con el término **entrada-salida** (E/S) se designa cualquier transferencia de información desde o hacia la memoria o el procesador.

Comprende tanto la transferencia entre los diversos niveles de la memoria como la comunicación con el exterior (periféricos).

Un dispositivo de entrada-salida es aquel capaz de transferir información entre la memoria o procesador y cualquier soporte de información externo.

Toda la transferencia es ordenada por el procesador central, pero si esta se hiciese de forma directa por él, se dedicaría mucho tiempo en el intercambio; por ello se han diseñado dispositivos de E/S que en la forma más sencilla se denominan unidades de acceso directo a memoria (DMA) y en general se denominan canales de transferencia.



Las **técnicas de E/S** más utilizadas son:

E/S programada: El procesador emite una orden de E/S de parte de un proceso a un módulo de E/S; el proceso espera entonces a que termine la operación, antes de seguir.

E/S dirigida por interrupciones: El procesador emite una orden de E/S de parte de un proceso, continúa la ejecución de las instrucciones siguientes y es interrumpido por el módulo de E/S cuando éste ha completado su trabajo.

Acceso directo a memoria (DMA): Un módulo de DMA controla el intercambio de datos entre la memoria principal y un módulo de E/S. El procesador envía una petición de transferencia de un bloque de datos al módulo de DMA y se ve interrumpido sólo cuando el bloque entero se haya transferido.

2. PRINCIPIOS HARDWARE EN ENTRADAS/SALIDAS

Existe una gran variedad de dispositivos que se pueden conectar a los ordenadores; tanto el hardware como el software de E/S se enfrentan con complicadas tareas para realizar las operaciones de forma segura y eficaz.

Existen tres características que agudizan el problema de entrada-salida.

La primera es la llamada *operación asíncrona*. Los tiempos de transacciones de la CPU con la memoria y los dispositivos de E/S son muy diferentes.

La memoria informa a la CPU con un tiempo fijo dado por unos pocos ciclos de reloj del procesador, por tanto sincronizada. Pero los dispositivos de E/S son muy variables con respecto a los tiempos de respuesta a las peticiones de la CPU, y se dice que actúan asíncronamente respecto a la unidad central.

Como la operación asíncrona es impredecible, el hardware de E/S ha de utilizar unas señales denominadas de permiso de tráfico, para indicar los momentos de la transferencia.

Otra característica a tener en cuenta es la *diferencia de velocidades* entre los dispositivos de E/S, la CPU y la memoria.

Los dispositivos más rápidos se conectan mediante canales especializados utilizando técnicas de acceso directo a memoria, mientras que los más lentos se controlan directamente por la CPU.

La tercera característica pasa por las *conversiones de formato* necesarias entre los periféricos y el sistema, dado que mientras unos requieren transferencias en serie, otros los reciben o transmiten en paralelo.

2.1. Dispositivos de entrada/salida

Los dispositivos externos que tienen que hacer E/S con los ordenadores pueden clasificarse, básicamente, en tres categorías:

Dispositivos legibles por los humanos: apropiados para la comunicación con el usuario, como el teclado y la pantalla de vídeo.

Dispositivos legibles por la máquina: adecuados para comunicarse con equipos electrónicos, como discos y unidades de cinta.

Dispositivos de comunicaciones: apropiados para comunicarse con dispositivos lejanos, como los módems.

Otra división distingue entre dispositivos de bloque y dispositivos de carácter.

Los **dispositivos de bloque** suelen estar asociados a dispositivos de alta velocidad que transmiten datos mediante técnicas de acceso directo a memoria. Los tamaños de bloques son variables y los más comunes son de 128, 512 y 1024 bytes.

Los **dispositivos de tipo carácter** están controlados, generalmente, por la CPU y transmiten carácter a carácter.

2.2. Controladores de dispositivo

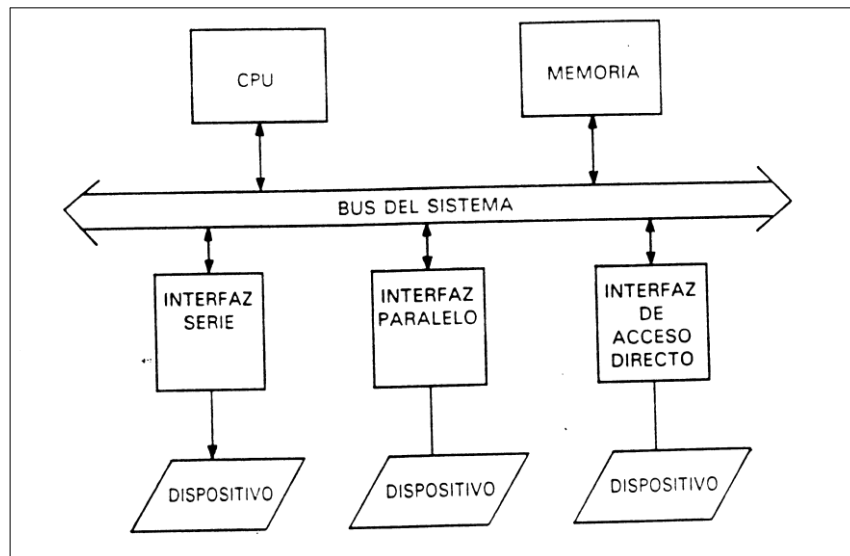
Las unidades de E/S tienen una parte mecánica que es el propio dispositivo y otra parte electrónica que se llama *controlador del dispositivo* y actúa como intermediario entre el ordenador y los dispositivos.

La autonomía de un controlador suele estar limitada a operaciones elementales.

El interés por separar el controlador del periférico es el de poder conectar varios periféricos del mismo tipo a un acoplador único. Así, aunque un sólo periférico puede transmitir información a la vez, los demás pueden realizar, de forma paralela, operaciones que no impliquen entradas o salidas.

Las funciones que realiza el controlador y el periférico dependen de cada tipo de dispositivo.

El controlador suele ser una tarjeta de circuitos impresos que posee un cable que va al dispositivo, y a su vez todos los cables de los dispositivos se unen en un elemento común llamado bus del sistema que va unido a los componentes del ordenador.



Los microordenadores suelen utilizar un sólo bus, tal como aparece en la figura superior, pero en los grandes sistemas se pueden utilizar múltiples buses y computadoras de E/S especializadas que se denominan **canales de E/S**.

Un controlador convierte el flujo de bits en serie transmitidos desde un dispositivo en un bloque de bytes para la CPU y realiza las correcciones de los errores que se puedan cometer en la transmisión.

Para actuar, el controlador dispone de tres *capas funcionales*: la *interfaz con el bus*, el *controlador propiamente dicho* y la *interfaz del dispositivo*.

2.3. Puertos de E/S

La parte intermedia del controlador, se compone de un conjunto de registros dedicados.

Al conjunto de tales registros se les denomina *puerto de E/S*.

Las *categorías funcionales* más importantes encontradas en los puertos de E/S son: *registros de entrada*, *registros de salida*, *registros de estado* y *registros de órdenes*.

Los *registros de entrada y salida* guardan los datos hasta que la CPU o el dispositivo destino esté preparados para recibirlos. Dado que las velocidades de los dispositivos y la CPU son muy dispares, puede suceder que mientras un dispositivo prepara un dato la CPU esté dedicada a otro programa y el dato elaborado tenga que esperar una interrupción para ser recibido, entonces el dato se guardará en un registro de entrada.

Los *registros de órdenes* se encargan de transmitir las órdenes de E/S entre la CPU y un dispositivo de E/S.

2.4. Acceso directo a memoria

La unidad DMA es capaz de imitar a la CPU y, de hecho, es capaz de relevar a la CPU en el control del sistema para transferir los datos con la memoria por el bus del sistema. Normalmente, el módulo de DMA debe usar sólo el bus cuando la CPU no lo necesite, o debe forzar a la CPU a que suspenda temporalmente su operación.

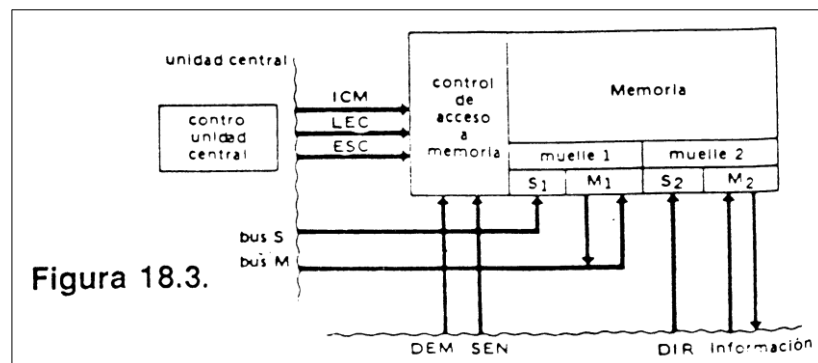
Para actuar necesita una serie de **elementos hardware** entre los que destacan: un *registro que almacene la dirección de memoria desde o hacia donde se produce la transferencia*; un *registro contador* para guardar la longitud del bloque a transferir; un *bit* que indique si la operación es de lectura o escritura y un *bloque de control* que regula el funcionamiento del sistema.

Para realizar una transferencia el procesador iniciará el DMA facilitándole la dirección de memoria, la longitud del bloque, el identificativo del periférico y el sentido de la transferencia.

Después el DMA actúa de forma autónoma preparando la transferencia y solicitando el bus al procesador por cada acceso a memoria. Por último cuando la transferencia se completa, se produce una interrupción para informar al procesador del final de la operación.

Como el bus debe ser compartido por el procesador, para sus tareas propias, y por el DMA, para las transferencias, se buscan formas de aprovechamiento máximo en lo referente a las **transferencias**:

- *Por ráfagas*: La unidad de DMA toma el control de bus en modo exclusivo y hasta su finalización.
- *Por robo de ciclo*: El DMA toma el bus durante un ciclo enviando una palabra cada vez.
- *Transparente*, se aprovechan los ciclos que el procesador está ocioso con respecto al bus.



2.5. Canales de entrada/salida

El canal de E/S es una extensión del concepto de DMA. Un canal de E/S tiene la capacidad de ejecutar instrucciones de E/S, lo que le da un control total sobre dichas operaciones. En un sistema informático que posea tales dispositivos, las instrucciones de E/S se almacenan en memoria principal y serán ejecutadas por un procesador de propósito específico en el mismo canal de E/S.

Canal se puede definir como vía de transmisión de información, tanto en sentido lógico como en sentido físico (línea de conducción), por ejemplo, la unión entre periféricos y ordenador o la conexión entre terminales y, en general, la conexión entre cualquier dispositivo.

Esencialmente, un canal de E/S es un pequeño ordenador con capacidad para ejecutar sus propios programas, denominados programas de canal (almacenados en la memoria del ordenador principal), permitiendo liberar a la CPU de las operaciones de E/S, una vez que éstas han comenzado a ejecutarse y aprovechándose de este tiempo para que dicha unidad pueda ejecutar un programa diferente hasta que el canal haya terminado su operación de E/S ejecutada en paralelo con el trabajo de la unidad central.

Resumiendo, un canal es un microprocesador especializado en operaciones de E/S.

Es evidente que el uso del canal se debe a que no toda la información está almacenada dentro de la memoria principal del ordenador, sino que ésta, está, además, almacenada en las memorias auxiliares o soportes de la información. Por este motivo, la información contenida en estos soportes y leída/escrita por los periféricos ha de ser transmitida hacia o desde el ordenador utilizando los canales como vía física de esa transmisión.

Un canal dispone de una serie de instrucciones que le permiten accionar los controladores y los periféricos a los que está conectado.

Posee una *palabra de estado* que contiene un contador e indicadores de estado de la operación que está realizando.

El programa que lo gestiona dispone de una serie de comandos, cada uno de los cuales realiza una operación sencilla de E/S.

La transferencia consta de las siguientes **operaciones**:

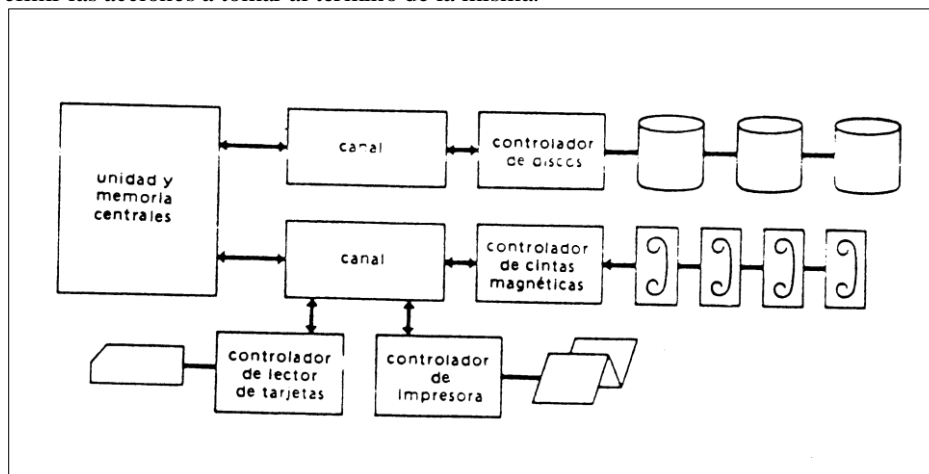
El canal es lanzado mediante una instrucción especial del procesador. Mediante ésta se inicializa la palabra de estado y se indica la dirección donde se encuentra la primera instrucción a realizar; también se señala la dirección del periférico con el que se va a producir la transferencia.

El canal selecciona el periférico y detecta el estado en el cual se encuentra para ver si la transferencia es posible.

Enviar órdenes de control al periférico.

Definir registros en memoria y en el periférico para la transferencia.

Definir las acciones a tomar al término de la misma.



Pues bien los canales, como unión entre periféricos y ordenador cumplen **funciones** tales como:

- a. Servir de vía de comunicación o transferencia de información desde o hacia un periférico y la unidad central de proceso.
- b. Aprovechar con el mayor rendimiento que sea posible los tiempos de ocupación de la unidad central de proceso.

A continuación estudiaremos los canales según la forma de realizar la transferencia de información entre periféricos y la unidad central.

Canales selectores

Tienen como misión principal la de enlazar los dispositivos de E/S rápidos con la CPU.

Estos canales trabajan en la modalidad de transmisión de datos o información denominada **ráfagas**, que consiste en dedicar la actividad del canal en forma de colas de espera de información, de manera que hasta que no termina la operación de transferencia de un bloque de información de un periférico, no

comienza con la transferencia de información de otro, debiendo esperar éste hasta que no se haya terminado la transferencia de información del primero. El canal de E/S elige un dispositivo y realiza la transferencia.

Canales multiplexores

Tienen como misión principal enlazar los dispositivos de E/S lentos con la CPU.

Estos canales trabajan en la modalidad de transmisión de datos o información denominada **multiplexor**, que consiste en repartir la actividad del canal en el tiempo transfiriendo en uno o en otro sentido un solo carácter desde o a cada una de las unidades periféricas lentas a las que sirve como medio de comunicación. Puede manejar la E/S con varios dispositivos al mismo tiempo.

Al estar varios periféricos unidos a la memoria central por el mismo canal, este efectúa una exploración cíclica tomando un carácter a cada uno de los periféricos de entrada y enviando un carácter a cada uno de los periféricos de salida, repitiendo estas operaciones hasta el total de ejecución de las operaciones de entrada y salida. Los caracteres pasan entrelazados.

En general el multiplexor prepara la información en condiciones de ser transmitida y otro (también denominado demultiplexor) identifica y selecciona la información que circula por el canal.

3. PRINCIPIOS DEL SOFTWARE DE ENTRADA/SALIDA

La idea básica del software de E/S consiste en organizar éste como una serie de estratos que oculten las peculiaridades del hardware por un lado y presenten un interfaz sencillo.

Un concepto importante en el diseño del software de E/S es la **independencia del dispositivo**, es decir, servir para cualquier dispositivo sin distinción.

También es importante el **control de errores** que ha de hacerse lo más próximo al hardware como sea posible.

A la hora del diseño de E/S es importante distinguir entre los dispositivos dedicados a un sólo propósito y dispositivos compartibles por varios usuarios a la vez.

Para lograr todos estos objetivos, el software de E/S se estructura en cuatro niveles que describimos a continuación:

3.1. Manejadores de interrupciones

Es el estrato interior más oculto de la estructura. Cuando se emite un comando y se espera una interrupción entonces el manejador de dispositivo queda bloqueado, y cuando se presenta aquella éste se desbloquea.

Los procedimientos de interrupciones comienzan con una orden de grabar para salvar los registros del proceso en ejecución; después se llama a un manejador que controla la interrupción y devuelve el control al proceso cuando se elimina el error o petición que provocó la interrupción.

3.2. Manejadores de dispositivos

Su trabajo consiste en aceptar las solicitudes del software y comprobar que éstas cumplen los requisitos para ser atendidas.

En general cada manejador se hace cargo de un dispositivo o de una clase de dispositivos relacionados entre sí.

Cuando se emite una solicitud, el manejador puede iniciar la ejecución si está libre o por contra guardar la petición en una lista de espera.

Por ejemplo, al solicitar un registro de un disco el manejador del dispositivo verificará que el brazo lector esté sobre la pista adecuada, después buscará el bloque que lo contiene, etc.

3.3. Software de E/S independiente del dispositivo

Existe un conjunto de funciones que son comunes para todos los dispositivos y éstas a pesar de ciertas diferencias se muestran con una interfaz uniforme de cara al usuario.

Entre ellas destacan: la *información de errores*, la *asignación y desasignación de dispositivos*, la *protección* de los mismos, su *nominación*, la *designación del tamaño de los bloques*, etc.

3.4. Software del usuario

Una pequeña parte del software de E/S no reside en el S.O. como tal, sino que está grabado en bibliotecas que los usuarios manejan mediante programas enlazados a las mismas.

También existen otros procedimientos de E/S en el ámbito de usuario, por ejemplo el manejo de dispositivos por cola de impresión que se utiliza en sistemas multiprogramados.

La idea consiste en grabar un directorio especial con los programas que van a ir al dispositivo y un programa manejador de dicho directorio que selecciona el siguiente proceso cada vez que el dispositivo queda libre. Así se evita que los usuarios mantengan el dispositivo abierto para su uso un tiempo indefinido mientras otros procesos están esperando la utilización de dicho periférico.

4. E/S CONTROLADA POR PROGRAMA

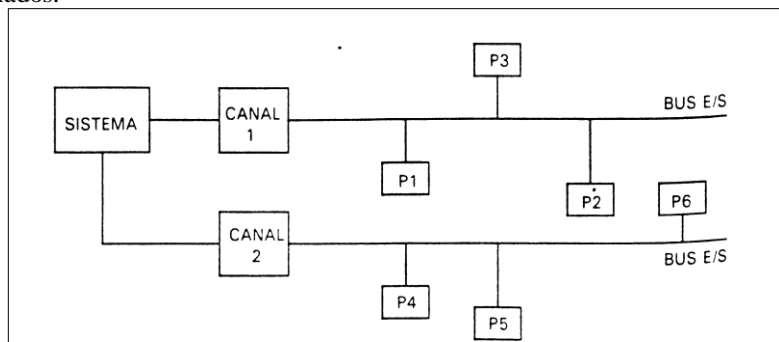
Los dispositivos de E/S sincronizan sus actividades con la CPU mediante protocolos de intercambio. Mediante el bus del sistema y los registros de estado y de órdenes se proporcionan las señales para dicha sincronización.

Para comenzar con una operación de E/S es imprescindible tener conocimiento sobre el estado en que se encuentra el dispositivo. Una práctica común es inicializar cada dispositivo antes de que participe en actividades de E/S.

Cuando la CPU quiere transmitir o recibir datos lanza una orden de arranque; entonces en el punto de E/S establece un bit de estado en la posición de ocupado. Cuando el dato ha sido transmitido el bit cambia a modo de preparado para indicar que el dispositivo puede ser utilizado para una nueva operación.

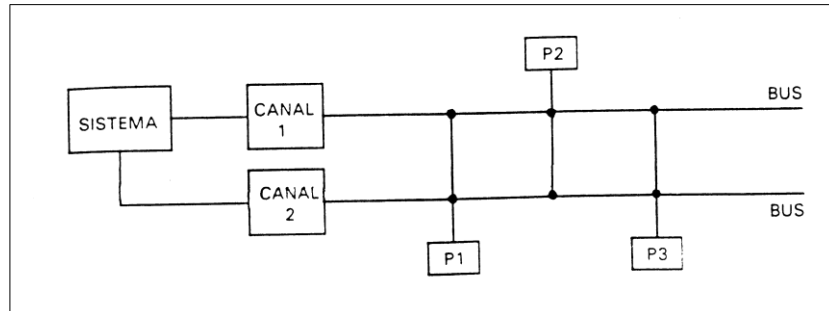
En general, a cada canal se le conectan varios dispositivos periféricos y aunque las formas de conexión son muy variadas hay dos que se utilizan con mayor frecuencia:

En la **conexión en serie** cada canal tiene asignada una conexión fija a un grupo de periféricos determinados.



Para conectar con un periférico hay que hacerlo por su canal, y en caso de que esté ocupado por otro periférico, hay que esperar, por tanto se deben establecer prioridades.

En la **conexión en red** cada canal se conecta con todos los periféricos permitiendo transferencias alternativas para los periféricos. Si un canal está ocupado se transmite por otro.



Como cada canal está conectado a varios periféricos ha de repartir el tiempo entre ellos.

Si atiende a los periféricos que tiene conectados repartiendo el tiempo en forma circular se dice que es un *multiplexor*.

Si por el contrario cuando un periférico comienza una transferencia retiene el canal correspondiente hasta el final de la misma, hablaremos de un canal *selector*.

En el método de E/S controlada por programa la CPU repite continuamente un ciclo de comprobación de estado del dispositivo hasta que se encuentre preparado para una nueva transferencia.

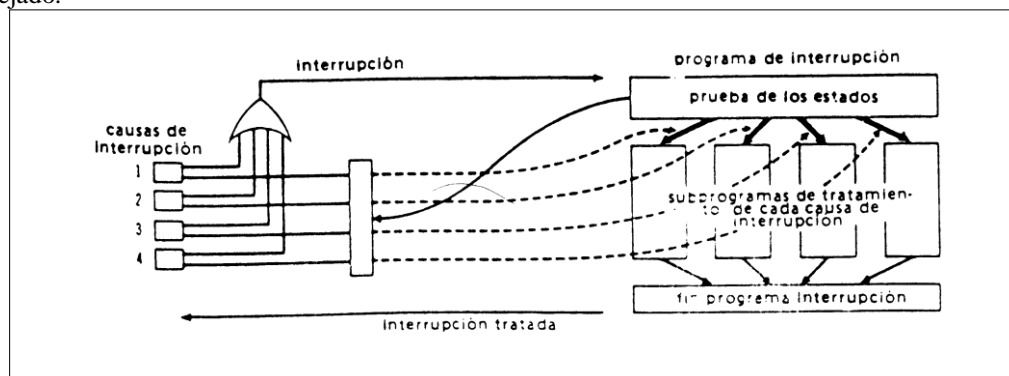
Para efectuar una salida se repite el mismo bucle. Dicho bucle se le suele denominar de *espera activa* y se utiliza frecuentemente en sistemas monousuario.

Un método alternativo llamado de *escrutinio*, consiste en utilizar el bucle de comprobación de la CPU a los diferentes dispositivos y pasar al programa de E/S del primer dispositivo preparado.

El escrutinio se puede hacer comenzando cada vez con el primer dispositivo o bien de forma cíclica para evitar prioridades.

5. INTERRUPCIONES

Las interrupciones son un mecanismo hardware para sincronizar la CPU con los sucesos externos que son asíncronos. La idea consiste en que la interfaz fuerce a la CPU a dejar su actividad para atender a una rutina de E/S y luego haga que la unidad central retorne al proceso en curso en el punto donde lo había dejado.



Para implementar dicha actividad es necesario un registro que guarde el contador del programa en el instante en que se produce la interrupción para que la CPU pueda reanudar el proceso justo en el punto en que fue detenido.

Cuando existen varios dispositivos se puede realizar un escrutinio después de la petición de interrupción; en el caso de existir varios dispositivos preparados habría que tener en cuenta su prioridad para ser atendidos.

6. SPOOL

Un sistema de spool se puede subdividir en varios componentes que a su vez pueden agruparse en dos funciones: entrada y salida.

El *spool de entrada* ha de realizar dos operaciones principales: leer cada registro y reservarlo en un dispositivo auxiliar de acceso directo y proporcionar acceso a la copia del dispositivo auxiliar cuando sea solicitado.

La primera operación se realiza independientemente del proceso a ejecutar.

El acceso a la copia se produce como respuesta a una llamada específica del proceso de ejecución.

Los sistemas de llamada e interrupción han de funcionar coordinados y para ello el dispositivo auxiliar esta dividido en secciones las cuales quedan referenciadas en una tabla del spool de entrada.

Estas secciones pasan por varios estados: disponible, entrada, retención y ejecución.

Aunque las características del spool son buenas no se utilizan en todos los sistemas debido a su complejidad y al consumo de buena parte del área del S.O., además de necesitar una elevada cantidad de espacio en el dispositivo de almacenaje auxiliar.

7. BIBLIOGRAFÍA

Stallings, W. *Sistemas operativos* Prentice Hall, 2ed., 1997

Tanenbaum, Andrew S. *Sistemas operativos: Diseño e implementación* Prentice Hall, 1990