

AULA 17 - Sistema de Entrada e Saída

Dentre as inúmeras funções do SO, também é de sua responsabilidade controlar os dispositivos de E/S de um computador. Para isso, o sistema operacional pode enviar comandos aos dispositivos, tratar interrupções e erros, além de fornecer uma interface entre os dispositivos externos e o sistema. Os dispositivos podem ser do seguinte tipo:

1. armazenamento: discos magnéticos, discos ópticos, fitas, etc.
2. transmissão: placas de rede, modem, etc.
3. interface homem-máquina: monitor, teclado, mouse, etc.

Basicamente, todo dispositivo se comunica com o sistema computacional enviando sinais via cabo ou através do ar. Para isso, os dispositivos utilizam um ponto de conexão chamado **porta** (serial ou paralela). Um conjunto de feios condutores é chamado de **barramento**. Podemos definir um barramento como uma estrutura de interconexão que possui um protocolo bem definido para a transmissão das mensagens. Este protocolo especifica o que podem ser enviado pelos fios condutores.

Todo dispositivo de E/S está conectado ao sistema computacional através de uma **controladora**. Esta controladora é um dispositivo eletrônico que pode operar uma porta, um barramento ou um dispositivo. Em outras palavras, a controladora é um módulo responsável pela interface que é feita entre o sistema e os dispositivos. Dentre as suas principais funções estão:

1. controle e temporização.
2. comunicação como processador.
3. comunicação com os dispositivos.
4. armazenamento temporário de dados.
5. detecção de erros.

São exemplos de controladoras:

1. controladora gráfica: interface entre monitor e barramento do sistema (PCI).
2. controladora de memória: interface entre memória e barramento do sistema.

3. controladora de disco IDE: interface entre discos e barramento.
4. controladora de disco SCSI: interface entre discos SCSI e barramento.

Enquanto algumas controladoras são bem simples (controladora de uma porta serial) outras podem ser bastante complexas. Uma controladora SCSI geralmente é implementada como uma placa de circuito separada e contém um microprocessador, um micro código e unidades de armazenamento (memória) para permitir o processamento de mensagens no protocolo SCSI. Outros dispositivos têm sua controladora embutida, como por exemplo, os discos IDE (impressa em uma de suas superfícies).

A Figura 1 (Tanenbaum) apresenta as taxas de dados de dispositivos comuns. Com o passar do tempo e com a constante evolução tecnológica, estes dispositivos tendem a ficarem mais rápidos.

Device	Data rate
Keyboard	10 bytes/sec
Mouse	100 bytes/sec
56K modem	7 KB/sec
Telephone channel	8 KB/sec
Dual ISDN lines	16 KB/sec
Laser printer	100 KB/sec
Scanner	400 KB/sec
Classic Ethernet	1.25 MB/sec
USB (Universal Serial Bus)	1.5 MB/sec
Digital camcorder	4 MB/sec
IDE disk	5 MB/sec
40x CD-ROM	6 MB/sec
Fast Ethernet	12.5 MB/sec
ISA bus	16.7 MB/sec
EIDE (ATA-2) disk	16.7 MB/sec
FireWire (IEEE 1394)	50 MB/sec
XGA Monitor	60 MB/sec
SONET OC-12 network	78 MB/sec
SCSI Ultra 2 disk	80 MB/sec
Gigabit Ethernet	125 MB/sec
Ultrium tape	320 MB/sec
PCI bus	528 MB/sec
Sun Gigaplane XB backplane	20 GB/sec

Figura 1. Taxas típicas de transferências de dados para dispositivos, redes e barramentos.

A Figura 2 ilustra como o monitor e o teclado é conectado ao barramento de sistema. A controladora de vídeo (placa de vídeo) é responsável por criar a interface entre o monitor e o sistema computacional. Deve ser um canal de comunicação com um taxa de transferência mais rápida que o canal de comunicação (porta serial) usado pelo teclado.

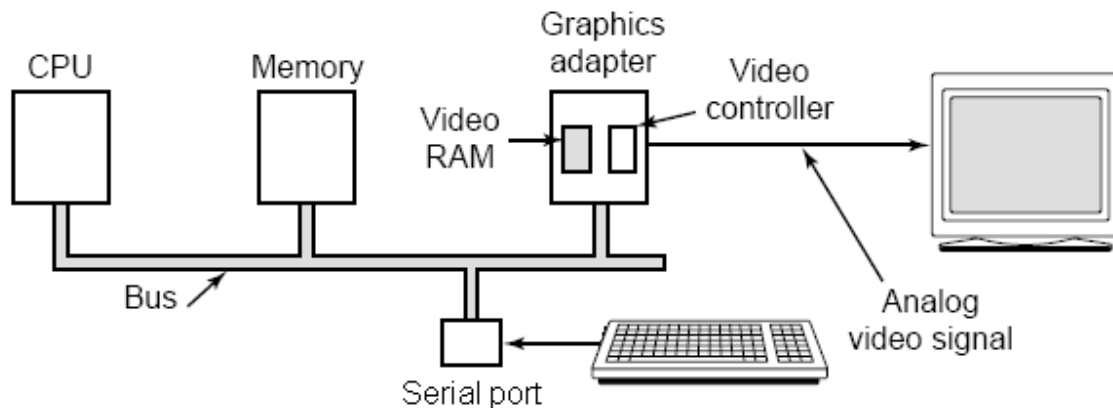


Figura 2. Exemplos de como os dispositivos são conectados ao barramento de sistema.

A pergunta é: Como o processador pode emitir comandos e dados para uma controladora a fim de realizar uma transferência de I/O?

As controladoras possuem registradores internos capazes de armazenar dados e sinais de controle emitidos pela CPU. Dessa forma, a comunicação se estabelece através de operações de leitura e escrita de bits nesses registradores. Algumas controladoras também permitem que seja feita programação de suas funções através desses registradores.

Existem três técnicas diferentes que podem ser utilizadas durante a realização de operações de E/S, são elas: *E/S Programada*, *E/S dirigida por Interrupção* e *Acesso Direto à Memória (DMA)*. A seguir comentamos o funcionamento de cada uma dessas técnicas.

E/S Programada

Neste modo de comunicação o processador, além de executar o programa, possui controle total sobre as operações de E/S. Este controle inclui desde a detecção do estado do dispositivo, o envio de comandos para o módulo de E/S (leitura ou escrita) e a transferência de dados. Por isso, toda vez que o programa em execução realiza alguma operação de I/O, o processador tem que

interromper sua execução para tratar da operação solicitada. Com isso, a execução do programa fica interrompida até que a operação seja finalizada. Como podemos imaginar, se o processador for mais rápido que o módulo de E/S ocorrerá um desperdício de processamento, ocasionado pela espera que foi comentada.

E/S dirigida por Interrupção

Neste modo de comunicação o processador não fica esperando que a operação de I/O seja finalizada para continuar a execução do programa. Ao invés disso, ele simplesmente emite um sinal para a controladora (módulo de I/O) com a operação solicitada e continua executando outras instruções do programa. Quando a controladora tiver finalizado a operação solicitada pela CPU ela emite um sinal de interrupção ao processador indicando o término da operação. Nesse instante o processador acessa o módulo de I/O para obter o resultado da operação.

Devemos ressaltar que tanto na E/S Programada quanto na dirigida por Interrupção, o processador sempre é o responsável por obter dados da memória principal (operação de saída memória => dispositivo) ou por armazenar dados na memória principal (operação de entrada dispositivo => memória).

DMA – Acesso Direto à Memória

Esta técnica é uma opção mais interessante de realizar as operações de E/S, pois a transferência de dados entre o módulo de E/S e a memória principal é feita diretamente sem a necessidade de envolver o processador.

Dentre as principais desvantagens dos métodos anteriores podemos citar:

1. a taxa de transferência de E/S é limitada pela velocidade com que o processador pode servir um dispositivo.
2. o processador tem que se ocupar em gerenciar a transferência dos dados.

Para evitar estes problemas desenvolveu-se um módulo adicional que está conectado ao barramento do sistema. Este módulo é chamado de DMA. Esse controlador é capaz de imitar o processador e transferir dados diretamente de e para a memória através do barramento do sistema. Assim, podemos verificar que este módulo pode disputar o uso do barramento com a CPU.

Portanto, o gerenciamento da operação de I/O fica a cargo do DMA. Suponha que um programa solicite uma leitura de um arquivo em disco. O DMA será o responsável por controlar a operação e armazenar os dados lidos do disco direto na memória principal, no endereço transmitido pela CPU durante a solicitação. Ao finalizar a transferência, o DMA emite um sinal de interrupção ao processador indicando o término da operação. Nesse instante, o processador busca os dados direto na memória, poupando tempo de acesso à controladora do disco e, conseqüentemente, melhorando o desempenho do sistema computacional. Dessa forma, o processador só se envolve no início e no fim da operação.

Interrupções

A um sinal ou evento capaz de interromper a seqüência normal de execução de instruções damos o nome de *interrupção*. Os tipos mais comuns de interrupções estão listados a seguir:

1. *interrupção de software*: ocasionada por alguma condição resultante da execução de uma instrução, como por exemplo um *overflow* ou uma divisão por zero.
2. *interrupção de relógio*: gerada pelo *clock* do processador, destinada a executar operações periódicas.
3. *interrupção de E/S*: ocasionada por um controlador de E/S para indicar a conclusão de uma operação ou a ocorrência de erros.
4. *interrupção de falha de hardware*: provocada por falha do *hardware*, como um erro de paridade em memória ou a queda de energia.

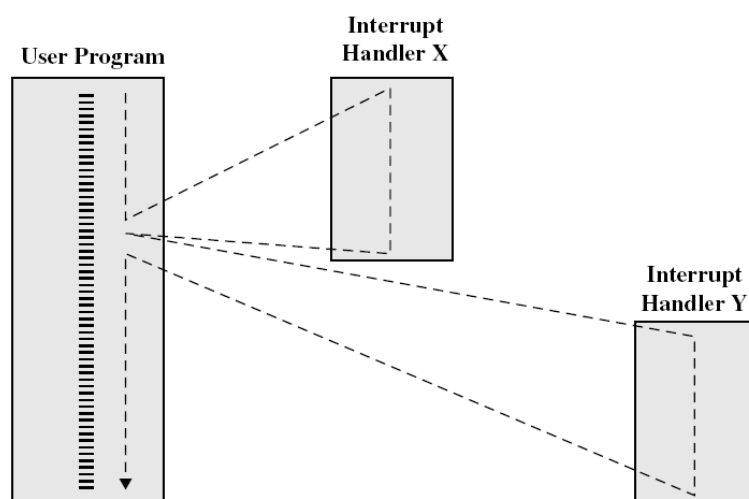


Figura 3. Esquema de processamento de uma interrupção sequencial.

O principal objetivo das interrupções é melhorar o desempenho da CPU (Ex. Leitura de dado em disco).

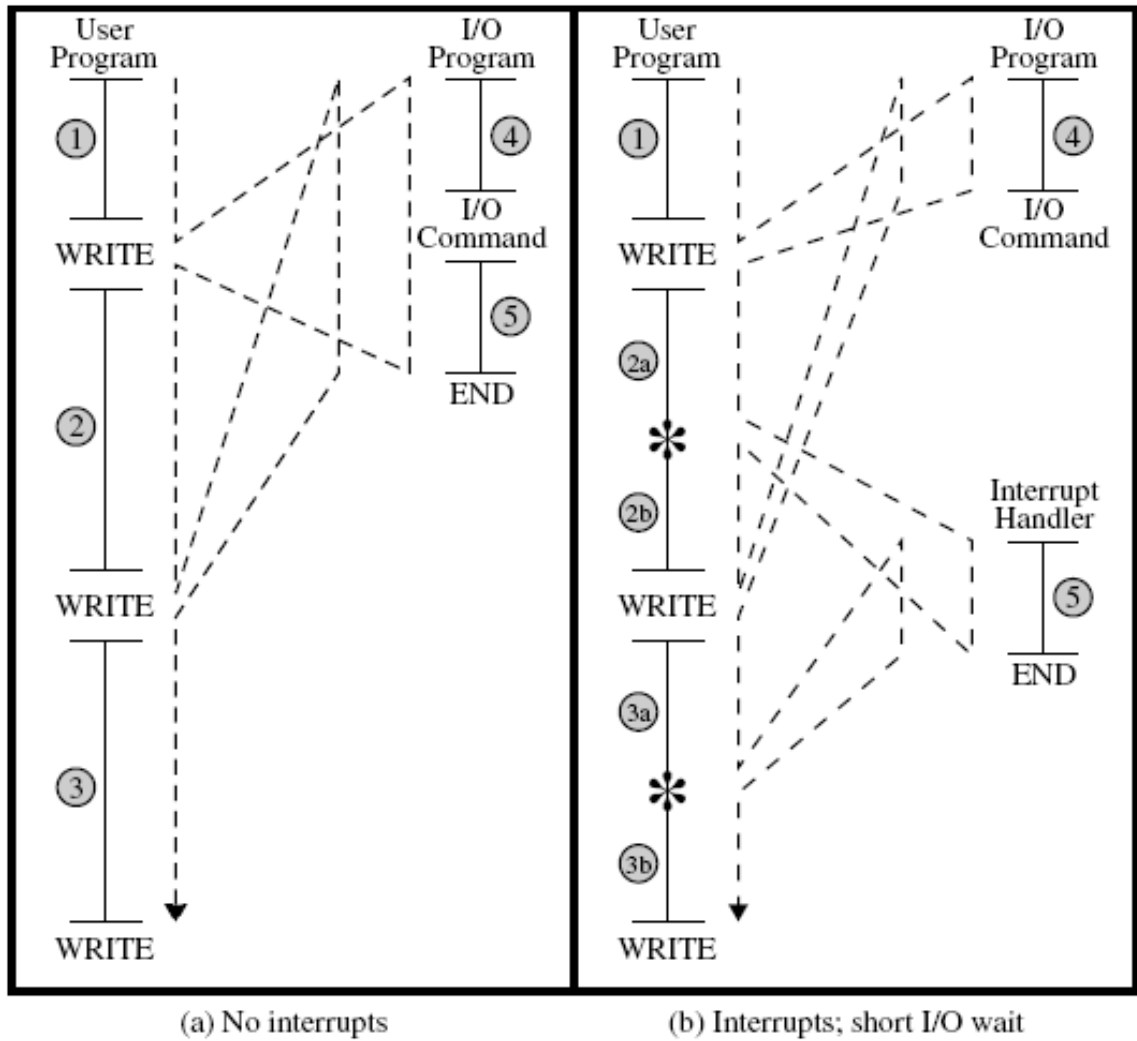


Figura 4. Fluxo de controle de um programa com e sem interrupção.

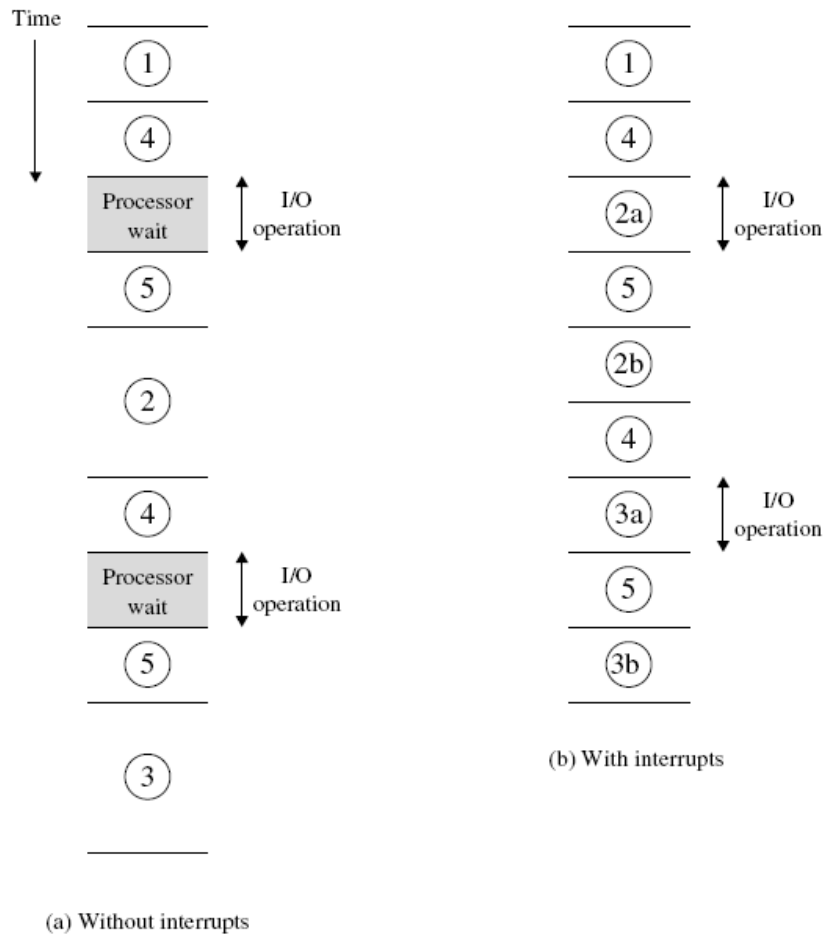


Figura 5. Diagrama de tempo de execução.

Exercícios

1. Quais são as principais funções das controladoras dos dispositivos de E/S?
2. Cite as controladoras que você possui na sua máquina.
3. O que é uma interrupção?
4. Explique como funcionam os mecanismos de entrada e saída: programada, dirigida por interrupção e com DMA.
5. Como as interrupções podem ajudar a melhorar o desempenho dos processadores?

Bibliografia

- [1] STALLINGS, W. *Arquitetura e Organização de Computadores*, 5ª Edição, Prentice Hall, São Paulo, 2002. (Cap. 6)
- [2] SILBERSCHATZ, A. et al. *Sistemas Operacionais: conceitos e aplicações*. Campus, 2001. (Cap. 12)
- [3] TANENBAUM, A. *Sistemas Operacionais Modernos*. 2ª Edição. Prentice Hall, 2003. (Cap. 5)