



Capítulo 6

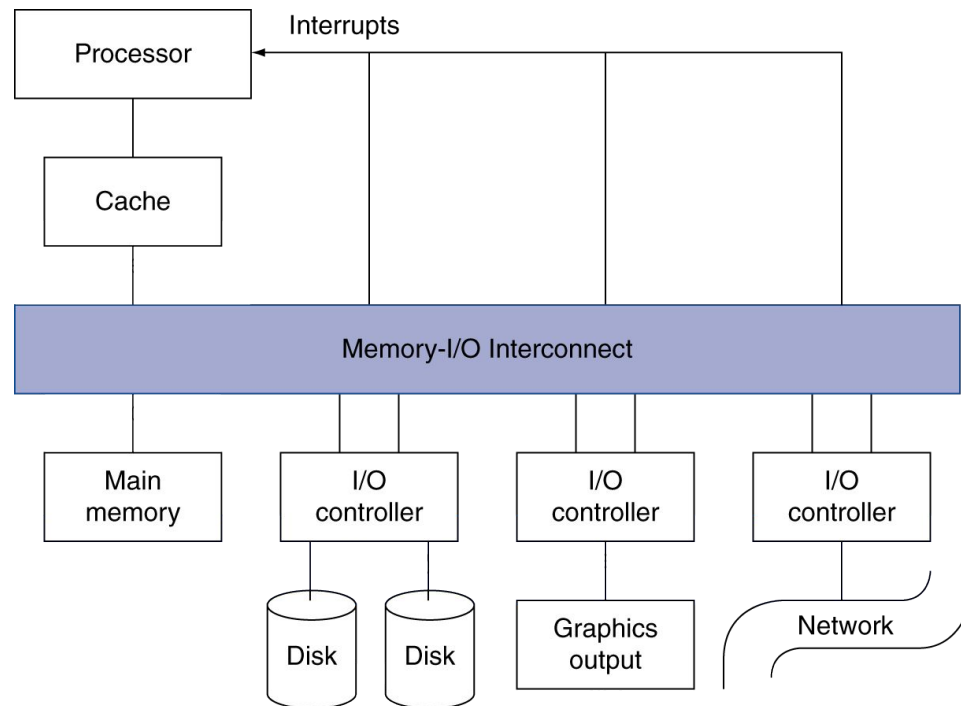
Armazenamento

Introdução

- Dispositivos de E/S podem ser caracterizados por
 - Comportamento:
 - Entrada (somente leitura)
 - Saída (somente escrita, não pode ser lido)
 - Armazenamento (pode ser relido e normalmente reescrito)
 - Parceiro: humano ou máquina
 - Taxa de dados:
 - Taxa de pico em que os dados podem ser transferidos entre o dispositivo de E/S e a memória principal ou processador
 - É útil saber qual é a demanda máxima que o dispositivo pode gerar ao projetar um sistema de E/S

Introdução

- Conexões de E/S
 - Conexão entre processador e memória normalmente chamadas de barramentos (fios paralelos)
 - A maioria das conexões de E/S hoje seja mais próxima de linhas seriais dedicadas



Introdução

Device	Behavior	Partner	Data rate (Mbit/sec)
Keyboard	Input	Human	0.0001
Mouse	Input	Human	0.0038
Voice input	Input	Human	0.2640
Sound input	Input	Machine	3.0000
Scanner	Input	Human	3.2000
Voice output	Output	Human	0.2640
Sound output	Output	Human	8.0000
Laser printer	Output	Human	3.2000
Graphics display	Output	Human	800.0000–8000.0000
Cable modem	Input or output	Machine	0.1280–6.0000
Network/LAN	Input or output	Machine	100.0000–10000.0000
Network/wireless LAN	Input or output	Machine	11.0000–54.0000
Optical disk	Storage	Machine	80.0000–220.0000
Magnetic tape	Storage	Machine	5.0000–120.0000
Flash memory	Storage	Machine	32.0000–200.0000
Magnetic disk	Storage	Machine	800.0000–3000.0000

FIGURE 6.2 The diversity of I/O devices.

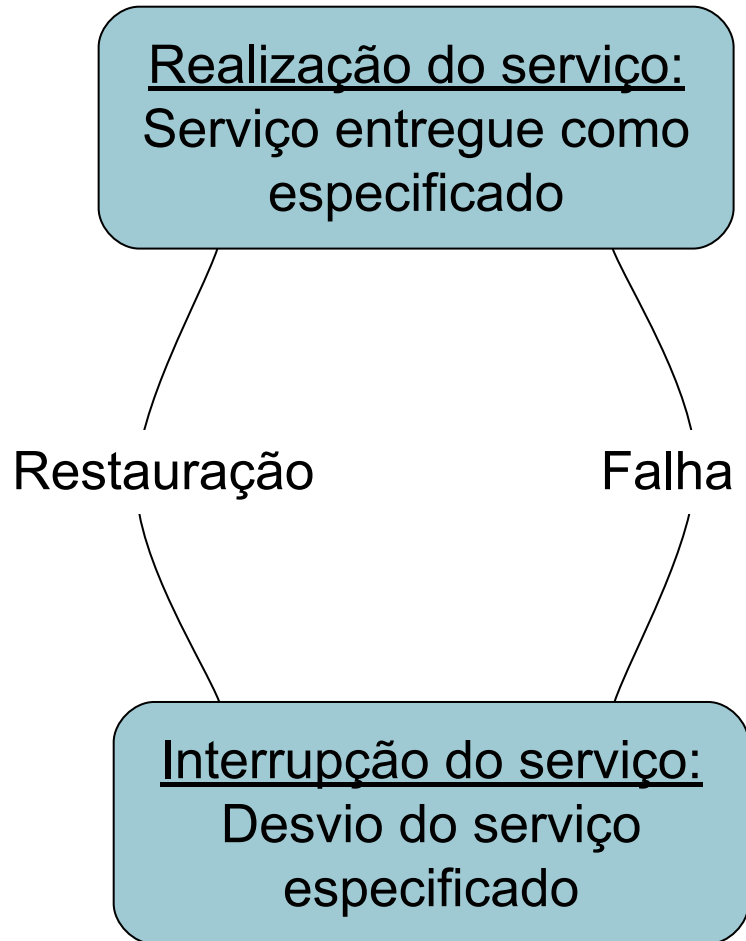
Copyright © 2013 Elsevier Inc. All rights reserved



Características de E/S

- Confiança (Dependability) é importante
 - Principalmente para armazenamento
- Medidas de desempenho
 - Atraso (tempo de resposta)
 - Vazão (Throughput) ou largura de banda
 - Desktops & sistemas embarcados
 - Tempo de resposta & diversidade de dispositivos
 - Servidores
 - Vazão & escalabilidade e capacidade de expansão dos dispositivos

Confiança



- Falha: problema/erro com componente
 - Pode ou não levar a falha do sistema

Medidas de confiabilidade

- **Confiabilidade:** tempo médio entre falhas (MTTF - Mean Time Between Failures)
- Interrupção do serviço: tempo médio para o reparo (MTTR - Mean Time To Repair)
- Tempo médio entre falhas
 - $MTBF = MTTF + MTTR$
- **Disponibilidade** = $MTTF / (MTTF + MTTR)$
- Melhorando a disponibilidade
 - Aumentar MTTF: prevenção de falhas, tolerância a falhas, previsão de falhas
 - Reduzir MTTR: ferramentas e processos aprimorados para diagnóstico e reparo

Medidas de confiabilidade

- Queremos que a disponibilidade seja muito alta
- Uma abreviação é indicar o número de “noves de disponibilidade” por ano
- Um serviço de Internet muito bom hoje oferece 4 ou 5 noves de disponibilidade
 - Com 365 por ano, que significa $365 \times 24 \times 60 = 526.000$ minutos

Um nove:	90%	⇒ 36,5 dias de reparo/ano
Dois noves:	99%	⇒ 3,65 dias de reparo/ano
Três noves:	99,9%	⇒ 526 minutos de reparo/ano
Quatro noves:	99,99%	⇒ 52,6 minutos de reparo/ano
Cinco noves:	99,999%	⇒ 5,26 minutos de reparo/ano

Armazenamento em Disco

■ Armazenamento magnético giratório não volátil



Seek é o processo de posicionar uma cabeça de leitura/gravação na trilha correta de um disco.

Cilindro é usado para se referir a todas as trilhas sob as cabeças em determinado ponto para todas as superfícies.

track

cylinder

sector

Trilha é um dos milhares de círculos concêntricos que compõem a superfície de um disco magnético.

Setor é um dos segmentos que compõem uma trilha em um disco magnético; um setor é a menor quantidade de informação lida ou escrita em um disco.

Setores de Disco e Acesso

- Cada setor registra
 - ID do setor
 - Dados (512 bytes (...2010), 4096 bytes (2011...))¹
 - Código de correção de erros (ECC)
 - Usado para reparar e recuperar dados
 - Campos de sincronização e lacunas (gaps)
- O acesso a um setor envolve
 - Atraso na fila se outros acessos estiverem pendentes
 - Busca (seek): mover as cabeças
 - Atraso de rotação
 - Transferência de dados
 - Tempo da controladora (overhead)



1-<https://www.seagate.com/br/pt/tech-insights/advanced-format-4k-sector-hard-drives-master-ti/>

Exemplo de Acesso ao Disco

- Disco
 - 512B sector, 15,000rpm, 4ms average seek time, 100MB/s transfer rate, 0.2ms controller overhead, idle disk
- Average read time
 - 4ms seek time
 - + $\frac{1}{2} / (15,000/60) = 2\text{ms}$ rotational latency
 - + $512 / 100\text{MB/s} = 0.005\text{ms}$ transfer time
 - + 0.2ms controller delay
 - = 6.2ms
- If actual average seek time is 1ms
 - Average read time = 3.2ms

Exemplo de Acesso ao Disco

■ Disco

- Setores 512B
- Rotação de 15.000 rpm
- Tempo médio de busca (seek) é 4ms
- Taxa de transferência de 100MB/s
- Overhead de controlador é 0.2ms

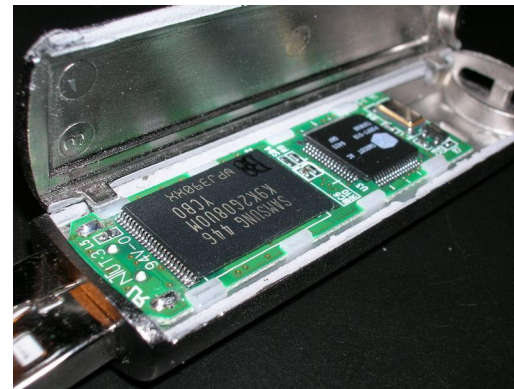
■ Tempo médio de leitura

- 4ms tempo de busca
 - + $\frac{1}{2} / (15.000/60) = 2\text{ms}$ atraso de rotação
 - + $512 / 100\text{MB/s} = 0.005\text{ms}$ tempo de transf.
 - + 0.2ms atraso do controlador
 - = 6.2ms

■ O tempo de busca é o fator de maior impacto!

Armazenamento em Flash

- Armazenamento semicondutor, não-volátil
 - 100× a 1000× mais rápido que disco
 - Menor, com mais baixo consumo de energia, mais robusto
 - Maior custo por GB (entre disco e DRAM)



Tipos de Flash

- Flash NOR: célula de bit como uma porta NOR
 - Acesso aleatório de leitura/gravação
 - Usado para memória de instrução em sistemas embarcados
- Flash NAND: célula de bit como uma porta NAND
 - Mais denso (bits/área), mas bloqueia o acesso de cada vez
 - Mais barato por GB
 - Usado para pen drive USB, armazenamento de mídia, ...
- Bits Flash se desgastam após 100.000 acessos
 - Nivelamento de desgaste: remapear dados para blocos menos usados

Características de Flash

Characteristics	NOR Flash Memory	NAND Flash Memory
Typical use	BIOS memory	USB key
Minimum access size (bytes)	512 bytes	2048 bytes
Read time (microseconds)	0.08	25
Write time (microseconds)	10.00	1500 to erase + 250
Read bandwidth (MBytes/second)	10	40
Write bandwidth (MBytes/second)	0.4	8
Wearout (writes per cell)	100,000	10,000 to 100,000
Best price/GB (2008)	\$65	\$4

FIGURE 6.8 Characteristics of NOR versus NAND flash memory in 2008.

Copyright © 2013 Elsevier Inc. All rights reserved

RAID

- Redundant Array of Inexpensive (Independent) Disks
 - Paralelismo aumenta desempenho
 - Mais disco(s) extra(s) para armazenamento
- Provê um sistema de armazenamento tolerante a falhas
 - Especialmente se discos com falha podem ser “hot swapped”
 - Assume falhas de disco independentes

Hot swapping: Substituição de um componente de hardware enquanto o sistema está em execução.

RAID

Espalhamento dos dados por vários discos, chamado striping, força automaticamente os acessos a vários discos.

Espelhamento dados idênticos são escritos em vários discos, para aumentar a disponibilidade dos dados.

O grupo de discos de dados que **compartilham um disco ou bloco de verificação comum**

O grupo de discos de dados que **compartilham dois discos ou blocos de verificação comum**

RAID 0
(No redundancy)
Widely used

RAID 1
(Mirroring)
EMC, HP(Tandem), IBM

RAID 2
(Error detection and
correction code) Unused

RAID 3
(Bit-interleaved parity)
Storage concepts

RAID 4
(Block-interleaving parity)
Network appliance

RAID 5
(Distributed block-
interleaved parity)
Widely used

RAID 6
(P + Q redundancy)
Recently popular

Data disks

Redundant check disks

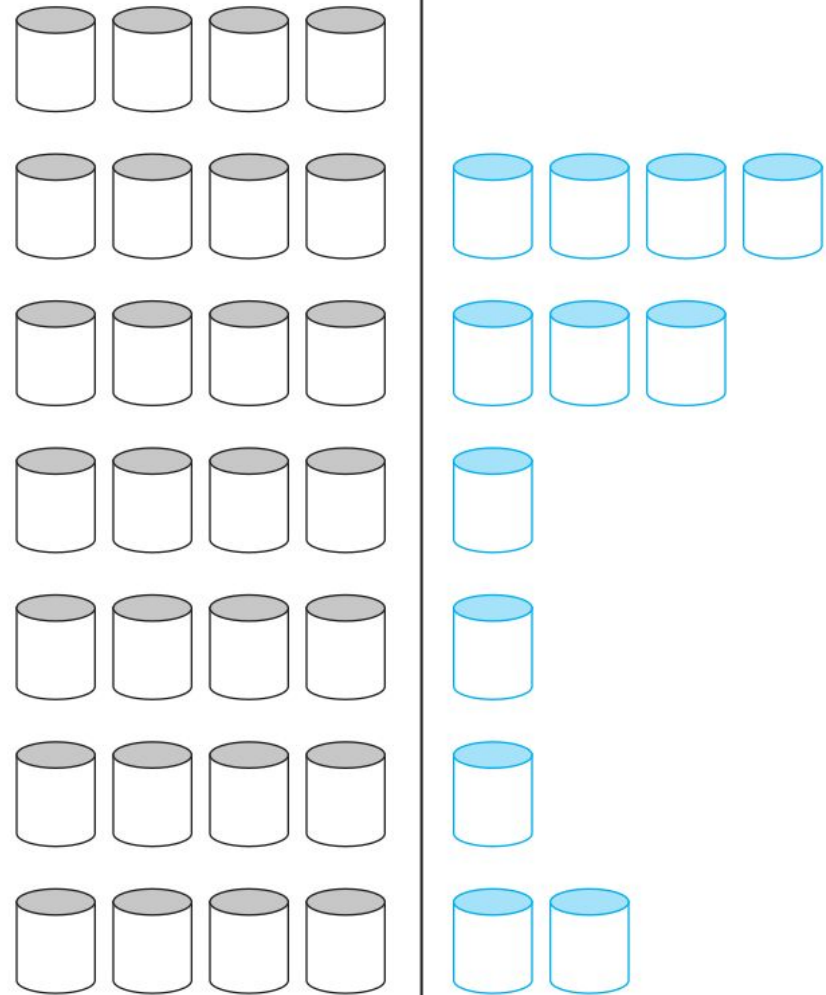


FIGURE 6.12 RAID for an example of four data disks showing extra check disks per RAID level and companies that use each level. Copyright © 2013 Elsevier Inc. All rights reserved

Referências

- Seções 6.1 a 6.4 e 6.9 - Organização e Projeto de Computadores - A Interface Hardware/Software, David A. Patterson & John L. Hennessy, Campus, 4 edição, 2013.