

Ficha 6 – Periféricos

Exercícios resolvidos

1. Um disco magnético roda a 10000 RPM e tem 516 setores (de 0,5 kB) por cilindro. O seu tempo de busca mínimo é 4 ms e o tempo de busca médio é 10 ms. Sabe-se ainda que a taxa de transferência é de 100 MB s^{-1} e a latência do controlador é de 5 ms. Calcule o tempo necessário para transferir um ficheiro de 600 kB, assumindo que os respetivos setores estão dispostos em disco da maneira mais favorável possível (inicialmente, a cabeça de leitura está numa posição aleatória).

1. O ficheiro ocupa $\frac{600 \text{ kB}}{0,5 \text{ kB}} = 1200$ setores. Uma vez que cada cilindro tem uma capacidade de 516 setores, o ficheiro está distribuído por 3 cilindros ($516 + 516 + 168$) o que implica duas mudanças de cilindro durante a transferência.

O tempo de acesso é a soma das seguintes parcelas: tempo de busca médio, tempo de rotação, tempo de transferência, tempo de mudança de cilindro e latência do controlador do disco. Ou seja:

$$10\text{ms} + 0,5 \times \frac{60\text{s}}{10000} + \frac{600\text{kB}}{100\text{MB s}^{-1}} + 2 \times 4\text{ms} + 5\text{ms}$$

$$= 10\text{ms} + 3\text{ms} + 6\text{ms} + 8\text{ms} + 5\text{ms} = 32\text{ms}$$

2. Um sistema possui um CPU que opera a 4 GHz. Este sistema possui ainda um disco que transfere dados para o processador em grupos de 4 palavras (2 bytes cada) e tem uma taxa de transferência de dados de 40 MB s^{-1} . [Considere $\text{kB} = 10^3 \text{ B}$, $\text{MB} = 10^6 \text{ B}$.]

- Se pretendermos utilizar *polling* como técnica de gestão de periféricos e sabendo que uma operação de *polling* consome 400 ciclos de relógio, qual é a fração de tempo de CPU consumida?
- Se em vez de *polling* fosse utilizada a técnica de interrupções, qual seria a fração de tempo de CPU consumida? Admita que o *overhead* de cada transferência, incluindo o atendimento da interrupção, é de 600 ciclos de relógio e que o disco está sempre potencialmente ocupado.
- Comparando os resultados das alíneas anteriores, em que situações é que a técnica de interrupções pode ser vantajosa?

2.

- Dados transferidos por acesso: $4 \times 2\text{B} = 8 \text{ B}$
Acessos por segundo:

$$\frac{40 \text{ MB s}^{-1}}{8 \text{ B}} = \frac{40 \times 10^6}{8} = 5 \times 10^6$$

Então, o número de ciclos consumidos pela operação é: $5 \times 10^6 \times 400 = 2 \times 10^9$

Tendo em consideração o número de ciclos consumidos por segundo, a percentagem média de tempo de CPU gasto é:

$$\frac{2 \times 10^9}{4 \times 10^9} = 0,5 = 50 \%$$

- b) Dados transferidos por acesso: $4 \times 2B = 8B$

Acessos por segundo:

$$\frac{40 \text{ MB s}^{-1}}{8B} = \frac{40 \times 10^6}{8} = 5 \times 10^6$$

Então, o número de ciclos consumidos pela operação é: $5 \times 10^6 \times 6 \times 10^2 = 3 \times 10^9$

Tendo em consideração o número de ciclos consumidos por segundo, a percentagem média de tempo de CPU gasto é:

$$\frac{3 \times 10^9}{4 \times 10^9} = 0,75 = 75 \%$$

- c) No segundo cenário, o custo (em ciclos de relógio) do *overhead* de cada transferência, incluindo o atendimento da interrupção, é superior ao custo de uma operação de *polling*. A técnica baseada em interrupções é mais vantajosa quando o periférico não está sempre potencialmente ocupado. Para este caso concreto, a técnica de interrupções é mais vantajosa quando o disco está ocupado menos de 66,7 % do tempo.

$$\frac{X \times 3 \times 10^9}{4 \times 10^9} = 0,5 \iff X \times \frac{3}{4} = 0,5 \iff X = \frac{0,5}{0,75} = 66,7 \%$$

3. Um sistema informático, possui um processador capaz de executar 2×10^9 ciclos/s, e dois discos com uma taxa de transmissão de 100 MB s^{-1} cada. Uma transferência por DMA tem um custo de início de 1800 ciclos e de finalização de 200 ciclos. Assumindo que os discos têm continuamente informação a enviar ao CPU (pior cenário), determine qual é o tamanho mínimo da transferência de DMA por forma a que o custo total de acesso aos discos seja, no máximo, 1% do tempo de CPU. Considere $\text{kB} = 10^3 \text{ B}$ e $\text{MB} = 10^6 \text{ B}$.

3. Seja D o tamanho da transferência de DMA para um disco.

$$\text{Tempo de uma transferência de DMA: } t = \frac{D}{100 \times 10^6} = \frac{D}{10^8}$$

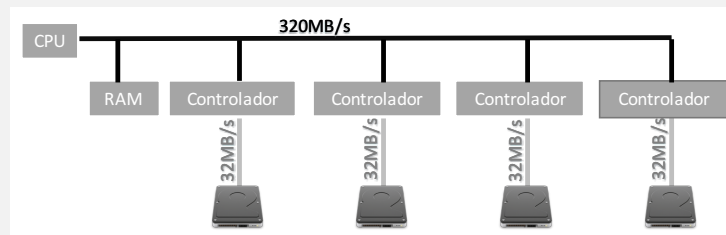
$$\text{Ciclos necessários por transferência: } N = 1800 + 200 = 2000 = 2 \times 10^3$$

$$\text{Ciclos necessários por unidade de tempo: } C = \frac{N}{t} = 2 \times 10^3 \times \frac{10^8}{D} = \frac{2 \times 10^{11}}{D}$$

$$\text{Percentagem do tempo de CPU: } \frac{C}{2 \times 10^9} = 0,01 \iff \frac{2 \times 10^{11}}{2 \times 10^9} = 0,01 \times \iff D = 10^4$$

O tamanho mínimo de uma transferência de DMA é de 10 kB para um disco. Havendo dois discos iguais será proporcionalmente reduzida, ou seja, 5 kB.

4. Considere o computador indicado na figura e que tem as seguintes características:



- O CPU opera a 2GHz;
- O barramento de memória possui uma taxa de transferência de 320 MB s^{-1} ;
- Ligados ao barramento de memória estão 4 controladores, cada um com o seu disco;
- O acesso aos discos é feito com uma largura de banda de 32 MB s^{-1} ; o tempo médio de busca mais a latência de rotação é 6 ms;
- O acesso aos discos é feito em blocos de 128 kB, guardados em setores consecutivos;
- Em cada acesso, o programa do utilizador e o sistema operativo gastam, respetivamente, 1 milhão e 3 milhões de ciclos de relógio.

Determine qual dos recursos (CPU, barramento de memória e discos) limita o desempenho expresso em blocos processados por unidade de tempo. [Considere $\text{kB} = 10^3 \text{ B}$, $\text{MB} = 10^6 \text{ B}$.]

4. CPU:

Tratamento de 1 bloco:

$$\frac{1 \times 10^6 + 3 \times 10^6}{2 \times 10^9} = 2 \text{ ms}$$

Nº blocos/s: 500

Barramento de Memória:

$$\frac{320 \text{ MB s}^{-1}}{128 \text{ kB}} = 2500 \text{ blocos/s}$$

Discos:

Por disco:

$$6 \text{ ms} + \frac{128 \text{ kB}}{32 \text{ MB s}^{-1}} = 10 \text{ ms}$$

O que corresponde a 100 blocos por disco por segundo.

No total temos então:

$$4 \times 100 \text{ blocos/s} = 400 \text{ blocos/s}$$

Como o CPU consegue processar 500 blocos/s e o barramento de memória suporta a transferência de 2500 blocos/s, são os discos (com 400 blocos/s) que limitam o desempenho.

Exercícios propostos

5. As seguintes medidas foram efetuadas em dois computadores semelhantes, mas ligados a redes de tipo diferente:

Caraterística	Rede tipo A	Rede tipo B
Taxa máxima de transmissão	$1,125 \text{ MB s}^{-1}$	10 MB s^{-1}
Latência de interconexão	$21 \mu\text{s}$	$56 \mu\text{s}$
<i>Overhead</i> de SW para envio	$200 \mu\text{s}$	$207 \mu\text{s}$
<i>Overhead</i> de SW para receção	$241 \mu\text{s}$	$360 \mu\text{s}$

A taxa máxima de transmissão representa a quantidade máxima de informação que pode ser transferida por unidade de tempo. A latência de interconexão é o atraso introduzido pela ligação física (cabo) e pelo processamento realizado pela placa de rede (em *hardware*). O envio e a receção de dados possuem ainda um *overhead* resultante do *software* que gere estas operações.

- Determinar, para cada tipo de rede, a latência computador-a-computador para uma mensagem de 250 bytes.
- Comentar a seguinte afirmação: “Por vezes a comunicação entre dois computadores é mais rápida utilizando uma rede do tipo A em vez de uma rede tipo B”.

6. Considerar dois sistemas diferentes destinados ao processamento de entrada/saída (E/S) de dados: o sistema A pode suportar 1500 operações de E/S por segundo e o sistema B pode suportar 1000 operações de E/S por segundo.

Os sistemas usam o mesmo processador, capaz de executar 12,5 MIPS (milhões de instruções por segundo). Assumir que cada transferência envolve 5 operações de E/S e que cada E/S requer 10000 instruções.

Determinar a taxa máxima de transferências por segundo (débito) de cada um dos sistemas.

7. Pretende-se enviar de um computador portátil para outro computador 100 ficheiros, com cerca de 40 MB cada, através de uma ligação sem fios a 5 Mbit s^{-1} . Antes de iniciar a transferência a energia na bateria é de 100 kJ. A placa de rede sem fios consome 5 W, enquanto transmite, e os restantes recursos do portátil consomem 35 W. Antes da transferência de cada ficheiro são necessários 10 s para escolher o ficheiro a enviar.

Calcular quantos ficheiros completos se conseguem enviar antes de o computador ficar sem energia. Nota: $1 \text{ W} = 1 \text{ J s}^{-1}$.

8. Um modelo de discos magnéticos tem as seguintes características:

- capacidade de cada setor: 512 bytes;
- velocidade de rotação: 10000 RPM;
- tempo de busca médio: 6 ms;
- taxa de transferência: 50 MB s^{-1} ;
- atraso introduzido pelo controlador: 0,2 ms.

- Calcule o tempo médio de acesso ao disco para ler ou escrever um setor.

- b) Após utilização verificou-se que o tempo de busca médio é afinal 25 % do valor anunciado. Calcule quanto melhora o tempo médio de acesso ao disco.

9. Assumir que um disco magnético tem as seguintes características:

- 10000 RPM, 416 setores (de 512 bytes) por cilindro;
- tempo de busca mínimo 2 ms e médio 8 ms;
- taxa de transferência 80 MB s^{-1} .

Ignorar os atrasos introduzidos pelo controlador.

- a) Calcule o tempo necessário para transferir ficheiros de 200 KiB e de 400 KiB, assumindo que os respetivos setores estão dispostos em disco da maneira mais favorável possível.
Nota: Tenha em consideração a capacidade total de um cilindro.
- b) Suponha agora que o ficheiro de 200 KiB ocupa setores escolhidos aleatoriamente. Determine o respetivo tempo de transferência.

10. Um programa executa repetidamente as seguintes tarefas:

1. Lê do disco um bloco de dados de 4 kB;
2. Processa estes dados;
3. Escreve o resultado (outro bloco de dados de 4 kB) num local aleatório do disco.

Os blocos são contíguos e localizam-se aleatoriamente numa pista do disco. O disco possui uma velocidade de rotação de 10000 RPM, o tempo médio de busca é 8 ms, a taxa de transferência é de 50 MB s^{-1} e o atraso introduzido pelo controlador do disco é 2 ms. O processamento dos dados demora 20 milhões de ciclos de relógio, sendo a frequência do sinal de relógio de 5 GHz. Assumir que nenhum outro programa está a usar o disco, ou o processador, e que não há sobreposição entre a operação do disco e o processamento dos dados.

Calcular o desempenho do sistema em blocos processados por segundo.

11. Considere um computador com dois discos:

- Um disco magnético tradicional que roda a 3000 RPM, tem uma latência média de busca de 30 ms, uma latência de controlador de 10 ms e transfere dados a 20 MB/s .
- Um disco SSD que usando memória *Flash* consegue ler e escrever blocos de 4 kB a uma taxa de 10 MB/s , com uma latência de acesso desprezável.

Pretende-se projetar um controlador inteligente para alternar entre os discos consoante for mais eficaz. Justifique as suas respostas com cálculos.

- a) Considerando uma transferência de 40 KiB contíguos, qual é o tempo total de acesso para o disco SSD?
- b) Qual é o tempo de acesso da mesma transferência de 40 KiB contíguos no caso do disco magnético? Considere que após começar a transferência, eventuais mudanças de pista demoram um tempo negligenciável.
- c) Calcule o tamanho da transferência que o controlador inteligente deve usar como valor limite para decidir trocar entre o disco SSD e o disco magnético (i.e., o limiar de decisão).

12. A execução de um programa demora 90 s, dos quais 80 s são gastos pelo CPU, correspondendo os restantes 10 s ao tempo de E/S. Se durante cinco anos o tempo de CPU for melhorado 50 % em cada ano, permanecendo o tempo de E/S constante, determine qual será a melhoria do tempo de execução do programa após esses cinco anos.

13. Pretende-se determinar o impacto de usar *polling* para realizar transferências de dados com três periféricos diferentes. Assuma que a operação de *polling* consome 800 ciclos de relógio e que o CPU opera a 2 GHz. Determine a fracção de tempo de CPU consumida por *polling* em cada um dos três casos seguintes, assumindo que a operação é feita com uma periodicidade tal que não ocorre perda de dados e que os periféricos estão sempre potencialmente ocupados.

- a) Um rato deve ser examinado 50 vezes por segundo para garantir que nenhum movimento do utilizador é perdido.
- b) Um leitor de disquetes transfere dados para o processador em unidades de 16 bits e tem uma taxa de transferência de dados de 50 kB s^{-1} .
- c) Um disco duro transfere grupos de 4 palavras (4 bytes cada) a uma taxa de 8 MB s^{-1} .

14. Suponha que são usados o mesmo CPU e disco duro do exercício anterior, mas desta vez com E/S realizada por interrupção. O *overhead* de cada transferência, incluindo o atendimento da interrupção, é de 1000 ciclos de relógio. Determine a fracção média de tempo de CPU consumida se o disco duro transferir dados durante 5 % do tempo. Compare com o resultado do exercício anterior e comente.

15. Suponha que usamos o mesmo sistema dos problemas anteriores, mas desta vez usando *Direct Memory Access* (DMA). Assuma que o estabelecimento da transferência de DMA consome 2500 ciclos e o tratamento da interrupção de fim de DMA consome 500 ciclos. Se a transferência média for de 8 KiB, determine a fracção do tempo de CPU consumida se o disco estiver ativo 100 % do tempo. Ignore o impacto da eventual contenção no barramento entre CPU e controlador de DMA.

16. Considere um computador com um CPU de 2 GHz e uma câmara de vídeo ligada por um barramento Firewire capaz de transferir blocos de 16 KiB a 800 MB/s. Pretende-se usar a estratégia de *polling* para transferir dados da câmara para memória.

- a) Considere um cenário em que o CPU tem de gastar 99 % do seu tempo em outras tarefas que não *polling*. Sabendo que uma operação de *polling* consome 200 ciclos, qual é o número máximo de operações de *polling* que podem ser realizadas num segundo?
- b) Considerando apenas as restrições impostas pelo barramento Firewire, qual é o número máximo de blocos de 16 KiB que podem ser lidos da câmara de vídeo num segundo?
- c) Com base nos cálculos anteriores, quantas operações de *polling* devem ser realizadas (por segundo) de forma a evitar perdas de dados e minimizar a ocupação de CPU?

17. Considerar um computador com as seguintes características:

- O CPU executa 3000 MIPS;
- O barramento de acesso a memória possui uma taxa de transferência de 1000 MB s^{-1} ;
- O programa do utilizador e o sistema operativo precisam de, respetivamente, 200000 e 100000 instruções por cada operação de E/S;

- O sub-sistema de armazenamento de dados é composto por N controladores de barramento SCSI Ultra32 com taxa de transferência de 320 MB s^{-1} , suportando cada um até 7 discos;
- O acesso a discos é feito com uma largura de banda de 75 MB s^{-1} e o tempo médio de busca mais a latência de rotação é 6 ms.

Considerar que a leitura dos discos se faz em blocos de 64 KiB situados em setores sequenciais na mesma pista. Assumir que não ocorrem conflitos de acesso a discos.

- a) Relativamente aos componentes fixos do sistema (CPU e barramento de memória), mostre que é o CPU que limita o desempenho do computador (*bottleneck*) calculando a taxa máxima de operações de E/S permitida por ambos.
- b) Calcule o tamanho mínimo de bloco que levaria a que fosse o barramento de memória o limitador do desempenho.
- c) Determine quantos discos e controladores são necessários para atingir a taxa de E/S calculada na alínea a), procedendo da seguinte forma:
 - i. Calcule o tempo de acesso a disco e o número de acessos por unidade de tempo;
 - ii. Determine o número de discos;
 - iii. Verifique que a largura de banda do controlador de disco é suficiente para suportar a taxa média de transferência de 7 discos e determine o número de controladores.

18. Considerar um sistema computacional com as seguintes características:

- o programa de um utilizador faz continuamente leituras de blocos de 64 KiB, requerendo 2 milhões de ciclos de relógio para processar cada bloco;
- o sistema operativo dispende 1 milhão de ciclos por cada operação de E/S;
- o barramento de acesso a memória possui uma taxa de transferência de 640 MB s^{-1} ;
- o CPU opera a 3 GHz;
- o acesso a discos é feito com uma largura de banda de 64 MB s^{-1} e o tempo médio de busca mais a latência de rotação é 9 ms;
- existem 20 discos ligados ao barramento, cada um com o respetivo controlador (assumir que cada disco pode ser controlado de forma independente e ignorar conflitos no acesso aos discos).

Identificar o recurso (CPU, barramento de memória ou discos) que limita o desempenho.

Fim do enunciado

Soluções:

5.

- a) $t_A = 684 \mu\text{s}$; $t_B = 648 \mu\text{s}$
b) A é mais rápida que B se $n^\circ \text{ bytes} \leq 204$

6.

Limite de A: 250 transf./s (CPU); limite de B: 200 transf./s (E/S)

7.

34

8.

- a) $t_{\text{acesso}} = 9,21 \text{ ms}$
b) melhoria $\simeq 2$

9.

- a) $t_{200k} = 13,5 \text{ ms}$; $t_{400k} = 18 \text{ ms}$
b) $t_{200k} = 4,4025 \text{ s}$

10.

33 blocos/s

11.

- a) $t_{\text{acesso}} = 4 \text{ ms}$
b) $t_{\text{acesso}} = 52 \text{ ms}$
c) 1 MB

12.

melhoria = 7,2

13.

- a) 0,002%
b) 1%
c) 20%

14.

1,25%

15.

0,15%

16.

- a) 100000 acessos/s

- b) 50000 blocos/s
- c) 50000 acessos/s

17.

- a) CPU: 10000 acessos/s; barramento de memória: 15625 acessos/s
Desempenho é limitado pelo CPU.
- b) se bloco > 100 kB CPU deixa de ser o *bottleneck*
- c)
 - i. $t_{\text{acesso}} = 6,9 \text{ ms}$; 144 acessos/s
 - ii. 70 discos
 - iii. 10 controladores

18. CPU: 1000 blocos/s; barramento de memória: 10000 blocos/s; discos: 2000 blocos/s.
Desempenho é limitado pelo CPU.