Análise do Valor Médio

1 Análise do Valor Médio para Redes Abertas de Filas

Hipótese

Rede Aberta de Filas em que todas as estações de serviço possuem servidores com capacidade fixa de processamento (CF) ou servidores que não apresentam espera em filas (infinitos servidores – IS).

Para todos os servidores de capacidade fixa, tem-se:

$$R_i = S_i(1+Q_i)$$

Um usuário ao chegar na estação i, em média, irá encontrar Q_i usuários à sua frente. Portanto, ele deverá esperar o atendimento de Q_i usuários, que levam em média S_i , antes de começar o seu próprio atendimento que levará também S_i .

1

Assumindo que o fluxo é balanceado, tem-se:

$$X = \lambda$$

A vazão da i-ésima estação é dada por:

$$X_i = X V_i$$

A utilização da estação i é:

$$U_i = X_i S_i = X V_i S_i = \lambda D_i$$

O número médio de usuários na estação i é:

$$Q_i = X_i R_i = X_i S_i (1+Q_i) = U_i (1+Q_i)$$
 ou

$$Q_i = U_i/(1-U_i)$$

Note que esta equação é análoga à do sistema M/M/1.

Substituindo-se este na equação inicial $\ de \ R_i \ tem-se$:

$$R_i = S_i/(1-U_i)$$
.

No caso de infinitos servidores tem-se:

$$R_i = S_i$$

e

$$Q_i = R_i X_i = S_i X V_i = X D_i = U_i$$

1.1 Algoritmo AVM para Redes Abertas de Filas

Entradas:

X: Taxa externa de chegada, Vazão do sistema;

S_i: Tempo de serviço por visita à estação i;

V_i: número de visitas à estação i;

M: número de estações na rede;

Saídas:

Q_i: número médio de usuários na estação i;

R_i: Tempo de resposta na estação i;

R: Tempo de resposta do Sistema;

U_i: Utilização da estação i;

N: Número médio de usuários no Sistema.

Cálculos:

Demanda total de serviço:

$$D_i = S_i V_i$$

Utilização dos dispositivos:

$$U_i = XD_i$$

Vazão dos Dispositivos:

$$X_i = XV_i$$

Tempo de resposta dos Dispositivos:

 $R_i = S_i/(1-U_i)$ Servidores de capacidade fixa;

 $R_i = S_i$ Infinitos servidores

Número médio de usuários nos dispositivos:

 $Q_i = U_i/(1-U_i)$ Servidores de capacidade fixa

 $Q_i = U_i$ Infinitos Servidores

Tempo médio de Resposta do Sistema:

$$R = \sum_{i=1}^M R_i V_i$$

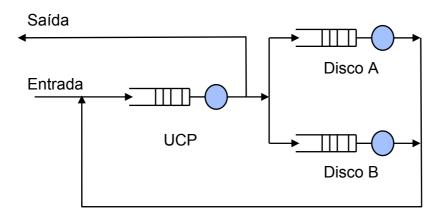
Número médio de usuários no Sistema:

$$N = \sum_{i=1}^M Q_i$$



Exemplo 1: Modelo de um Servidor de Arquivos

O modelo representa um sistema cliente/servidor com 6 clientes fazendo requisições ao servidor de arquivos.



O servidor de arquivos possui uma UCP e dois discos. Foram realizadas medidas que produziram os seguintes dados:

Intervalo de observação: 3 600 segundos Número de requisições dos clientes: 10 800 Tempo de UCP ocupada: 1 728 segundos Tempo de Disco A ocupado: 1 512 segundos Tempo de Disco B ocupado: 2 592 segundos Número de visitas ao disco A: 75 600 acessos Número de visitas ao disco B: 86 400 acessos

Baseando-se nessas medidas podem-se calcular as seguintes grandezas:

Vazão do Sistema:

X = 10800/3600 = 3 requisições por cliente

Visitas às estações por requisição de cliente:

 $V_A = 75600/10800 = 7 \text{ visitas}$ $V_B = 86400/10800 = 8 \text{ visitas}$ $V_{UCP} = 1 + 7 + 8 = 16 \text{ visitas}$

Demandas das estações por requisição de cliente:

 $D_{UCP} = 1728/10800 = 0,16 \text{ segundos}$ $D_A = 1512/10800 = 0,14 \text{ segundos}$

 $D_B = 2592/10800 = 0.24$ segundos (Gargalo do sistema)

Taxa de serviço das estações:

 $S_{UCP} = 0.16/16 = 0.01 \text{ segundos}$ $S_A = 0.14/7 = 0.02$ segundos

$$S_B = 0.24/8 = 0.03$$
 segundos

Cálculo do fator de utilização de cada dispositivo do sistema:

$$U_{UCP} = XD_{UCP} = 3x0,16 = 0,48$$

 $U_A = XD_A = 3x0,14 = 0,42$

$$U_B = XD_B = 3x0,24 = 0,72$$

Cálculo do tempo de resposta de cada dispositivo:

$$R_{UCP} = S_{UCP}/(1-U_{UCP}) = 0.01/(1-0.48) = 0.0192$$
 segundos

$$R_A = S_A/(1 - U_A) = 0.02/(1 - 0.42) = 0.0345$$
 segundos

$$R_B = S_B/(1 - U_B) = 0.03/(1 - 0.72) = 0.107$$
 segundos

Cálculo do tempo de resposta do Sistema:

$$R = \sum_{i=1}^{M} R_i V_i = 16x0,0192+7x0,0345+8x0,107 = 1,406 \text{ segundos}$$

Este modelo pode ser utilizado para responder diversas questões, tais como:

Qual é o novo desempenho do sistema se:

- a) Tivermos 8 clientes?
- b) Usarmos um Cache para o disco B com uma taxa de acerto de 50%, embora ele aumente a sobrecarga da UCP em 30% e o tempo de serviço do disco B em 10%?
- c) Utilizarmos um servidor com um disco (disco A) e dirigirmos todos os acessos a ele?

Respostas:

a) Sistema com 8 clientes:

Supondo que o cliente médio possua as mesmas características, pode-se dizer que a taxa de chegada vai crescer na proporção de 8:6.

$$X = 3 * (8/6) = 4 \text{ requisições/segundo}$$

Os outros parâmetros do sistema podem ser obtidos:

Utilização das estações:

$$U_{UCP} = XD_{UCP} = 4*0.16 = 0.64$$

$$U_A = XD_A = 4*0,14 = 0,56$$

$$U_B = XD_B = 4*0.24 = 0.96$$

Tempos de respostas das estações:

$$R_{UCP} = S_{UCP}/(1 - U_{UCP}) = 0.01/(1 - 0.64) = 0.0278$$
 segundos

$$R_A = S_A/(1 - U_A) = 0.02/(1 - 0.56) = 0.0455$$
 segundos

$$R_B = S_B/(1 - U_B) = 0.03/(1 - 0.96) = 0.75$$
 segundos

Tempo de Resposta do Sistema:

$$R = 16*0,0278+7*0,0455+8*0,75 = 6,76 \text{ segundos}$$

Conclusão: Se o número de clientes vai de 6 para 8 o tempo de resposta do sistema degrada 6,76/1,406 = 4,8 vezes.

Modelagem e Simulação de Sistemas de Computacionais

b) A segunda alteração implica numa mudança do número de visitas ao disco B:

V_B = 0,5*8 = 4 visitas ao disco B com Cache e taxa de acerto de 50%

 $S_{UCP} = 1.3*0.01 = 0.013 ==> D_{UCP} = 0.208$ segundos (Sobrecarga de 30% na

 $S_B = 1,1*0,03 = 0,033 ==> D_B = 4*0,033 = 0,132$ segundos (Aumento de 10%).

A análise do sistema modificado é a seguinte:

$$U_{UCP} = XD_{UCP} = 3*0,208 = 0,624$$

$$U_A = XD_A = 3*0,14 = 0,42$$

$$U_B = XD_B = 3*0,132 = 0,396$$

$$R_{UCP} = S_{UCP}/(1 - U_{UCP}) = 0.013/(1 - 0.624) = 0.0346$$
 segundos

$$R_A = S_A/(1 - U_A) = 0.02/(1 - 0.42) = 0.0345$$
 segundos

$$R_B = S_B/(1 - U_B) = 0.033/(1 - 0.396) = 0.0546$$
 segundos

Tempo de Resposta do Sistema:

$$R = 16*0,0346+7*0,0345+4*0,0546 = 1,013 \text{ segundos}.$$

Conclusão: Se utilizarmos Cache para o disco B o tempo de resposta do sistema melhorará em (1,406 - 1,013)/1,406 = 28%.

c) Terceira modificação: implica em ajustes em V_A e V_B :

$$V_B = 0$$

$$V_A = 7 + 8 = 15$$

 $D_{UCP} = 0.16$ segundos (como antes)

 $D_A = 15*0,02 = 0,3 \text{ segundos}$

$$U_{UCP} = XD_{UCP} = 3*0.16 = 0.48$$

$$U_A = XD_A = 3*0.3 = 0.90$$

$$R_{UCP} = S_{UCP}/(1 - U_{UCP}) = 0.01/(1 - 0.48) = 0.0192$$
 segundos

$$R_A = S_A/(1 - U_A) = 0.02/(1 - 0.90) = 0.2$$
 segundos

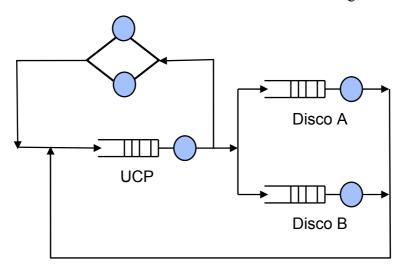
$$R = 16*0.0192 + 15*0.2 = 3.31$$
 segundos

Conclusão: Neste caso o tempo de resposta degrada por um fator de 3,31/1,406 = 2,35.



2 Análise do Valor Médio Rede Fechada

Consideremos a rede fechada de filas de um sistema "timesharing" com N usuários:



O tempo de resposta da i-ésima estação é dado por:

$$R_i(N) = S_i[1 + Q_i(N-1)]$$

onde Q_i(N-1) é o número médio de usuários na estação i quando existem N-1 usuários no sistema.

O desempenho do sistema para N usuários pode ser calculado em função do desempenho com N-1 usuários.

Em função das leis operacionais pode-se calcular os parâmetros do sistema. Conhecendose o tempo de resposta de cada estação pode-se calcular o tempo de resposta do sistema por:

$$R(N) = \sum_{i=1}^{M} V_i R_i(N)$$

A vazão do sistema é dada por:

$$X(N) = N/(R(N) + Z)$$

A vazão individual de cada estação é dada por:

$$X_i(N) = X(N)^*V_i$$

O número médio de usuários em cada estação é dada por:

$$Q_i(N) = X_i(N)^*R_i(N) = X(N)^*V_i(N)^*R_i(N)$$

As equações acima são válidas para servidores de capacidade fixa.

Para estações com infinitos servidores (SI), o tempo de resposta da estação é dado por:

$$R_i(N) = S_i$$



As outras equações da vazão e do número de usuários são válidas.

2.1 Algoritmo AVM para Redes Fechadas de Filas

Entradas:

N: número de usuários;

Z: tempo de pensamento;

M: número de estações na rede;

S_i: tempo médio de serviço por visita na estação i

V_i: número de visitas à estação i.

Saídas:

X: Vazão do sistema;

Q_i: Número médio de usuários na estação i

R_i: Tempo de resposta da estação i

R: Tempo de resposta do sistema

U_i: Utilização da estação i.

Algoritmo AVM para Rede Fechada

```
Inicialização:
```

De i = 1 até M faça
$$Q_i = 0$$
;

Iterações:



Exemplo 2

Considere a rede fechada do modelo de "timesharing" do servidor central. Cada requisição do usuário faz 10 acessos de E/S ao disco A e 5 ao disco B. O tempo de serviço por acesso ao disco A e B é respectivamente 300 e 200 milisegundos. Cada requisição à UCP leva 2 segundos e o usuário pensa 4 segundos.

Os parâmetros do sistema são:

$$S_A = 0.3$$
 $V_A = 10 ==> D_A = 3$
 $S_B = 0.2$ $V_B = 5 ==> D_B = 1$
 $D_{UCP} = 2$ $V_{UCP} = V_A + V_B + 1 = 16 ==> S_{UCP} = 0.125$
 $Z = 4$ $N = 20$

Inicialização:

Número de usuários: N = 0

Número de usuários nas estações:

$$Q_{UCP} = 0, Q_A = 0, Q_B = 0$$

Iteração 1:

Número de usuários:

$$N = 1$$

Tempo de resposta dos dispositivos:

$$R_{UCP} = S_{UCP} (1+Q_{UCP}) = 0.125(1+0) = 0.125$$

Tempo de resposta dos dispositivos(cont.):

$$R_A = S_A (1+Q_A) = 0.3(1+0) = 0.3$$
 segundos
 $R_B = S_B (1+Q_B) = 0.2(1+0) = 0.2$ segundos

Tempo de Resposta do Sistema:

$$R = R_{UCP}V_{UCP} + R_AV_A + R_BV_B = 0,125x16 + 0,3x10 + 0,2x5 = 6 \text{ segundos}$$

Vazão do Sistema:

$$X = N/(R+Z) = 1/(6+4) = 0.1$$

Número de usuários nas estações:

$$Q_{UCP} = XR_{UCP}V_{UCP} = 0.1x0.125x16 = 0.2$$

 $Q_A = XR_AV_A = 0.1x0.3x10 = 0.3$
 $Q_B = XR_BV_B = 0.1x0.2x5 = 0.1$

Iteração 2:

Número de usuários:

$$N = 2$$

Tempo de resposta das estações:

$$R_{UCP} = S_{UCP}(1+Q_{UCP}) = 0.125(1+0.2) = 0.15$$
 segundos

Tempo de resposta das estações:

Modelagem e Simulação de Sistemas de Computacionais

$$R_A = S_A(1+Q_A) = 0.3(1+0.3) = 0.39$$
 segundos
 $R_B = S_B(1+Q_B) = 0.2(1+0.1) = 0.22$ segundos

Tempo de Resposta do Sistema:

$$R = R_{UCP}V_{UCP} + R_AV_A + R_BV_B = 0,15x16 + ,39x10 + 0,22x5 = 7,4 \text{ segundos}$$

Vazão do Sistema:

$$X = N/(R+Z) = 2/(7,4+4) = 0,175;$$

Número de usuários nas estações:

$$Q_{UCP} = XR_{UCP}V_{UCP} = 0,175x0,15x16 = 0,421$$

$$Q_A = XR_AV_A = 0,175x0,39x10 = 0,684$$

$$Q_B = XR_BV_B = 0,175x0,22x5 = 0,193$$

A iterações seguintes produziram os resultados mostrados a seguir:

Iteração no.	R_{UCP}	R_A	R_B	R	Χ	Q_{UCP}	Q_A	Q_B
1	0,125	0,3	0,2	6,0	0,1	0,2	0,3	0,1
2	0,15	0,39	0,22	7,4	0,175	0,421	0,684	0,193
20	0,373	4,854	0,3	56,016	0,333	1,991	16,177	0,5

AVM Resumo

O método AVM é aplicável somente se a rede possuir solução na forma de produto, ou seja, satisfizer as condições de:

- Fluxo balanceado;
- Eventos únicos;
- Dispositivos homogêneos.

Adicionalmente, deve-se ter também

- Servidores de capacidade fixa (CF), ou
- Infinitos servidores (IS), e
- Em ambos os casos, com tempo de serviço distribuído exponencialmente.



3 AVM Aproximada

O algoritmo de valor médio é iterativo e para se calcular o desempenho do sistema com N usuários deve-se conhecer o desempenho do sistema com N-1 usuários. Para simplificar este processo e evitar ter que calcular o desempenho para 0,1,...,N,utiliza-se uma aproximação.

A aproximação se baseia no fato de que se aumentarmos o número de usuários no sistema, o número de usuários em cada estação irá aumentar na mesma proporção, isto é:

$$(Q_i(N)/N) = a_i$$
 constante para qualquer N, ou $(Q_i(N-1)/(N-1)) = Q_i(N)/N)$ ou $Q_i(N-1) = [(N-1)/N]Q_i(N)$

As equações do método AVM podem ser então re-escritas como a seguir:

Capacidade Fixa:

$$R_i(N) = S_i(1 + [(N-1)/N]Q_i(N))$$

Infinitos servidores:

$$\begin{aligned} R_i(N) &= S_i \\ R(N) &= \sum_{i=1}^M V_i R_i(N) \\ X(N) &= N/(Z + R(N)) \\ Q_i(N) &= X(N) V_i R_i(N) \end{aligned}$$

Observe que Q_i(N) é recalculado ao final de cada iteração. Quando a diferença entre $Q_i(N)$ no começo e no final da iteração for menor que um certo ε o algoritmo deve parar. O valor inicial de Q_i(N) não influencia o valor final. Ele interfere somente no número de iterações do algoritmo.

AVM Aproximada: Algoritmo de Schweitzer

Inicialização:

Iterações:



```
R_i = S_i^*[1+[(N-1)/N]^*Q_i]
                                        {CF} ou
                                        {SI}
R = 0:
De i = 1 até M faça R = R + V_i * R_i
X = N/(R+Z);
Aprox = max_i\{|Q_i - X^*V_i^*R_i|\}
De i = 1 até M faça
        Q_i = X^*V_i^*R_i
        X_i = X^*V_i
        U_i = X^*S_i^*V_i
}
```

Exemplo 3:

Considere novamente o exemplo do sistema de "timesharing";

Vamos analisá-lo com o método AVM aproximado quando o sistema possui 20 usuários; O critério de parada do algoritmo é quando a máxima alteração no número de usuários for menor que 0,01.

Os parâmetros do Sistema são:

```
N = 20 usuários
S_A = 0.3segundos V_A = 10 ==> D_A = 3

S_B = 0.2segundos V_B = 5 ==> D_B = 1
D_{UCP} = 2
V_{UCP} = V_A + V_B + 1 = 16
S_{UCP} = 0,125segundos
Z = 4segundos
```

Inicialização:

Para inicializar as filas assume-se que todos os 20 usuários estão igualmente distribuídos nas 3 filas do sistema:

$$Q_{UCP} = Q_A = Q_B = 20/3 = 6,67$$
 usuários

Iteração 1:

Tempo de resposta dos dispositivos:

$$\begin{aligned} R_{UCP} &= S_{UCP}(1 + (19/20)^*Q_{UCP}) = 0.125(1 + 0.95^*6.67) \\ &= 0.92 segundos \\ R_A &= S_A(1 + (19/20)^*Q_A) = 0.3 \ (1 + 0.95^*6.67) \\ &= 2.20 \ segundos \end{aligned}$$

Tempo de resposta dos dispositivos:

$$R_B = S_B(1+(19/20)^*Q_B) = 0.2 (1+0.95*6.67) = 1.47 \text{ segundos}$$



Tempo de resposta do Sistema:

$$R = R_{UCP}V_{UCP} + R_AV_A + R_BV_B = 0.92x16 + 2.20x10 + 1.47x5$$

= 44 segundos

Vazão do Sistema:

$$X = N/(R+Z) = 20/(44+4) = 0,42usuários/segundos$$

Número de usuários nos dispositivos:

$$Q_{UCP} = XR_{UCP}V_{UCP} = 0,42x0,92x16 = 6,11$$

$$Q_A = XR_AV_A = 0.42x2.20x10 = 9.17$$

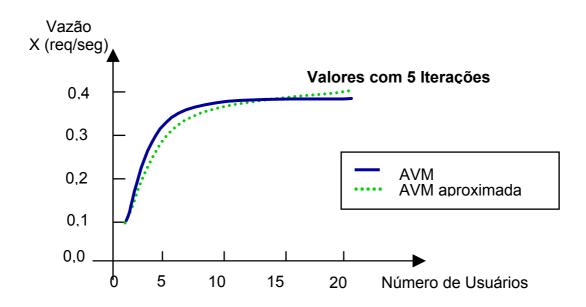
$$Q_B = XR_BV_B = 0.42x1.47x5 = 3.06$$

Variação máxima no número de usuários:

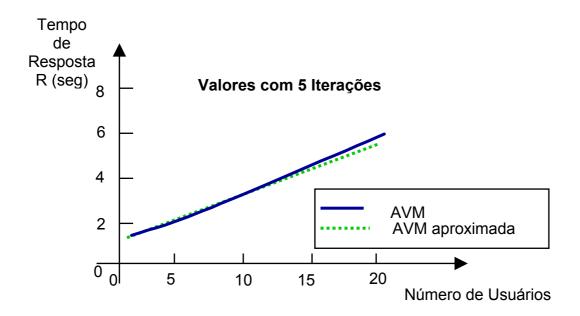
$$\Delta Q = \max\{0,56; 2,5; 3,61\} = 3,61 > 0,01$$

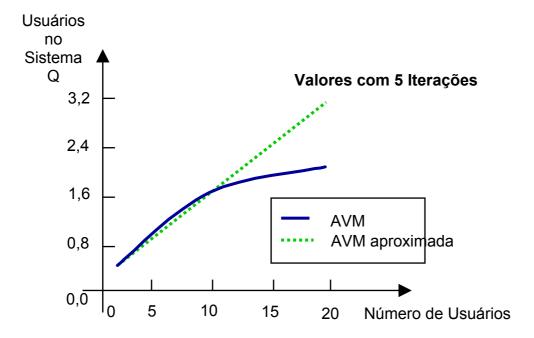
Este processo continua até que $\Delta Q < 0.01$, o que ocorre na 16 iteração.

Iterações	s R _{UCP}	R_A	R_B	R	Χ	Q_{UCP}	Q_A	Q_B
1	0,92	2,20	0,29	44	0,42	6,11	9,17	3,06
5	0,56	4,36	1,47	54,23	0,34	3,10	14,97	0,56
10	0,37	4,90	0,32	56,35	0,33	1,94	16,25	0,48
16	0,34	4,98	0,29	56,63	0,33	1,78	16,42	0,48









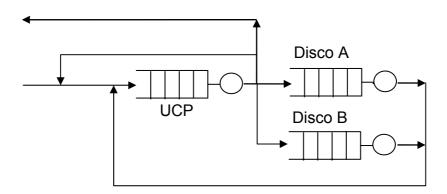
Bibliografia

[1] Jain, R., "The Art of Computer Systems Performance Analysis", John Wiley & Sons Inc, ISBN: 0-471-50336-3, 1991, 685 p.

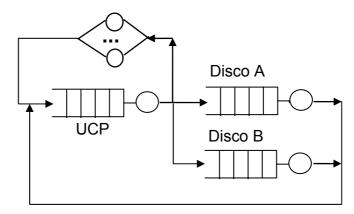


5 Exercícios

1) Um sistema de processamento transacional pode ser modelado pela rede de filas aberta como a da figura. As transações chegam a uma taxa de 0.8 transação por segundo, utilizam 1 segundo de UCP e fazem 20 acessos ao disco A e 4 acessos ao disco B. O total de visitas à UCP é 25 e o tempo médio de servico dos discos são 30 e 25 ms, respectivamente.



- a) Utilizando o AVM determine o número médio de transações no sistema, o tempo médio de resposta e qual dispositivos é gargalo do sistema.
- b) O que pode ser feito para melhorar o desempenho do sistema? Resp.: a) 5,01 e 6,26
- 2) Em um sistema de timesharing com 2 discos (para usuários e sistema), após o uso da UCP, a probabilidade de um programa utilizar o disco A é de 0,80, de utilizar o disco B é de 0,16 e de utilizar os terminais é de 0,04. O tempo que o usuário fica pensando é de 5 segundos, o tempo de serviço dos discos A e B é 30 e 25 ms, e o tempo médio de serviço por visita à UCP é 40 mseg.



Utilizando o AVM, determine a vazão e o tempo de resposta do sistema para N=1,...,5 usuários interativos.

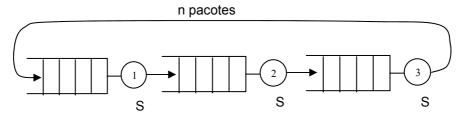
Resp.:

N		F	₹		Vazão		Q	
	UCP	Disco A	Disco B	Sistema	X	UCP	Disco A	Disco B
1	0.040	0.030	0.025	1.700	0.149	0.149	0.090	0.015
2	0.046	0.033	0.025	1.904	0.290	0.333	0.189	0.029
3	0.053	0.036	0.026	2.149	0.420	0.559	0.299	0.043
4	0.062	0.039	0.026	2.443	0.537	0.838	0.419	0.056
5	0.074	0.043	0.026	2.795	0.641	1.179	0.546	0.068

3) Repita o exercício anterior utilizando a aproximação de Schweitezer do MVA com N=5 usuários. Use o valor inicial de 5/3 para cada comprimento de fila e pare após 5 iterações. Resp.:

Iter		F	₹		Vazão		Q	
_	UCP	Disco A	Disco B	Sistema	X	UCP	Disco A	Disco B
1	0.293	0.220	0.183	12.467	1.145	8.397	5.038	0.840
2							•	
3							•	
4							•	
5	0.507	0.099	0.288	14.767	1.012	12.826	2.003	0.112

4) Uma rede de computadores com 2 enlaces (2-hop) e janela de controle de fluxo igual a n pode ser representada pela rede de fila fechada da figura a seguir:



Assumindo que a rede é balanceada, isto é, cada computador leva o mesmo tempo de serviço S para processar um pacote e que cada pacote faz uma visita a cada fila, determine a vazão e o tempo de resposta como função de n, para n=1,...,5, utilizando o MVA. Mostre que R(n)=(n+2)S; X(n)=n/[(n+2)S] $n\ge 1$.

5) Repita o exercício anterior para uma rede de computadores com h enlaces. Tal rede terá h+1 filas conectadas em anel. Escreva uma expressão para o poder da rede, o qual é definido como a razão entre a vazão e o tempo de resposta. Determine o tamanho da janela n que dê maior poder da rede. Mostre que R(n)=(n+h)S; X(n)=n/[(n+h)S] n≥1 Sendo Potência = X/R mostre que a potência máxima ocorre n=h.