PCS-2039
Modelagem e Simulação de Sistemas Computacionais

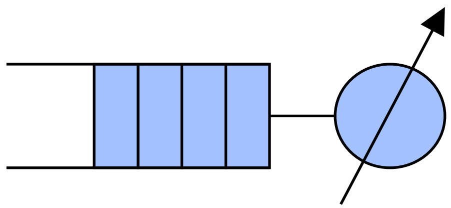
Graça Bressan

gbressan@larc.usp.br

Servidor Dependente de Carga
e
Decomposição Hierárquica

Servidores Dependentes da Carga (SDC)

- Os Servidores Dependentes de Carga (SDC) possuem taxa de serviço $\mu(n)$ variável e dependente do número de usuários no sistema.
- Uma fila com servidor dependente de carga é representada como:



$\mu(n)$.

Servidores Dependentes da Carga (SDC)

- Por exemplo, um sistema com **m** unidades de discos compartilhando uma única fila pode ser representado por uma estação de serviço dependente de carga sendo que a taxa de serviço **μ(n)** varia conforme o número de usuários no sistema até o limite de m.
- A taxa é uma função dependente do número de usuários no sistema de disco. Esta variação pode ser expressa por:

$$\mu(n) = \frac{n}{S} \quad n=1,2,...,m-1 \text{ e}$$

$$\mu(n) = \frac{m}{S} \quad n=m,m+1,...,\infty$$

onde S é o tempo de serviço de cada disco. Este caso é parecido com o sistema M/M/m, porém, a entrada não precisa necessariamente ser exponencial.

AVM com Servidores Dependentes da Carga

- O método da AVM pode ser generalizado para suportar o caso de se ter servidores dependentes da carga. Neste caso, devemos derivar a distribuição do número de usuários no sistema ao invés de somente o número médio de usuários.
- Sejam

$p_i(j \mid n)$	**probabilidade** de se ter **j** usuários na estação **i** dado que o sistema possui **n** usuários;
$\mu_i(j)$	**taxa de serviço** na estação **i** quando o sistema possui **j** usuários.

AVM com Servidores Dependentes da Carga

- O tempo de resposta de um usuário que encontra o dispositivo **i** com **j-1** usuários é dado por:

$$\mathbf{j \ / \ \mu(j)}$$

- A distribuição do tempo de resposta por visita à estação é dado por:

$$R_i(n) = \sum_{j=1}^{n} p_i(j-1 \mid n-1) \frac{j}{\mu_i(j)}$$

AVM com Servidores Dependentes da Carga

- A distribuição do número de usuários na estação i quando o sistema possui **n** usuários é dada por:

$$p_i(j|n) = \frac{X(n)}{\mu_i(j)} p_i(j-1|n-1) \qquad j=1,2,...,n \text{ e}$$

$$p_i(j|n) = 1 - \sum_{k=1}^{n} p_i(k|n) \qquad j=0$$

- O número médio de usuários neste caso é dado por:

$$Q_i(n) = \sum_{j=1}^{n} j p_i(j|n)$$

- É fácil verificar que estas fórmulas se reduzem às dos sistemas com **servidores de capacidade fixa** se substituirmos μ_i**(j)** por **1/S**$_i$, onde **S**$_i$ representa o tempo médio de serviço por visita à estação **i**.

Algoritmo AVM com Servidores Dependentes da Carga

- Entradas:

Z:	Tempo para pensar;
S_i:	Tempo de serviço por visita à estação i;
V_i:	Número de visitas à estação i;
M:	Número de estações no sistema (sem incluir os terminais);
N:	Número de usuários;
$m_i(j)$:	taxa de serviço da estação i quando existem j usuários em i.

- Saídas:

X:	Vazão do sistema;
Q_i:	Número médio de usuários na estação i;
R_i:	Tempo médio de resposta na estação i;
R:	Tempo médio de resposta do sistema;
U_i:	Fator de utilização da estação i;
$P_i(j)$:	Probabilidade de se ter j usuários na estação i.

Algoritmo AVM com Servidores Dependentes da Carga

Inicialização:

Para i = 0 até M faça

{ $Q_i = 0$ para servidor de capacidade fixa (CF) e centros de atraso (CA);

$P_i(0|0) = 1$ para servidores dependentes de carga (SDC);

}

Iterações:

Para n = 1 até N faça

{

Para i = 1 até M faça

{

$R_i = S_i(1+Q_i)$ se CF (capacidade fixa)

$R_i = S_i$ se CA (centro de atraso)

$R_i = \sum_{j=1}^{n} p_i(j-1|n-1)\frac{j}{\mu_i(j)}$ se SDC (servidor dependente de carga)

}

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

$X = n/(Z+R)$

Algoritmo AVM com Servidores Dependentes da Carga

Para i = 1 até M faça
{ Se CF ou CA então $Q_i = XV_iR_i$
Se SDC então
{Para j = n até 1 faça
$P_i(j \mid n) = (X/\mu_i(j)) * P_i(j-1 \mid n-1)$

$$p_i(0 \mid n) = 1 - \sum_{J=1}^{n} p_i(j \mid n)$$

}

}
}

Para i = 1 até M faça
{
$X_i = XV_i$
$U_i = XS_iV_i$ para CF ou CA
$U_i = 1 - P_i(0)$ para SDC

}

Exemplo 1: SDC

- Considere uma rede com duas estações de serviço. A primeira é o Disco B e possui capacidade fixa. A segunda é um SDC. O tempo médio de serviço por visita ao Disco B é de 0,26 segundos. Para cada visita ao SDC um usuário visita 6 vezes o Disco B.
- O tempo médio de serviço por visita ao SDC é dado pela seguinte função:

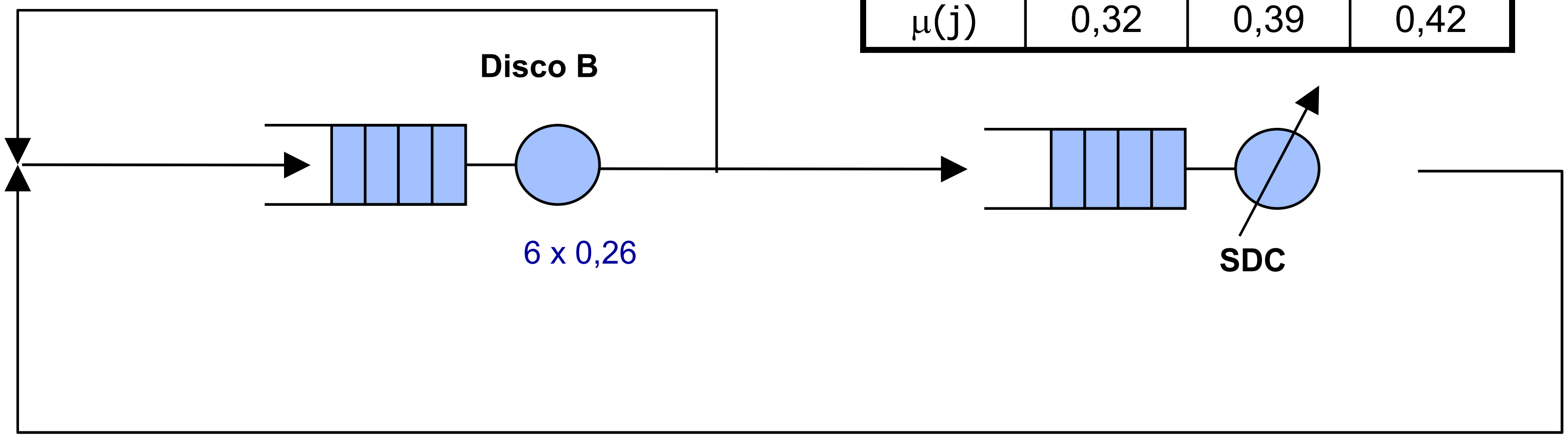
  $\mu(1)$ = 0,32 segundos;

  $\mu(2)$ = 0,39 segundos;

  $\mu(3)$ = 0,42 segundos.

Exemplo 1: SDC

j	1	2	3
$\mu(j)$	0,32	0,39	0,42



Exemplo 1: SDC

- Para analisar esta rede procede-se da seguinte maneira:
  Inicialização:
  $Q_B(0) = 0$ e $P(0|0) = 1$;
  Iteração 1: n=1
  Tempos de resposta dos dispositivos:
  $R_B(1) = S_B[1+Q_B(0)] = 0{,}26$ seg.
  $R_{SDC}(1) = P(0|0)*(1/\mu(1)) = 3{,}13$ seg.
  Tempo de resposta do Sistema:
  $R(1) = R_B(1)V_B + R_{SDC}V_{SDC} = 0{,}26x6 + 3{,}13x1 = 4{,}68$ seg.
  Vazão do Sistema:
  $X(1) = N/R(1) = 1/4{,}68 = 0{,}21$
  Número de usuários e probabilidades:
  $Q_B(1) = X(1)R_B(1)V_B = 0{,}21x0{,}26x6 = 0{,}33$
  $P(1|1) = [X(1)/\mu(1)]P(0|0) = [0{,}21/0{,}32]x1 = 0{,}67$
  $P(0|1) = 1 - P(1|1) = 1 - 0{,}67 = 0{,}33$

Exemplo 1: SDC

Iteração 2: n=2

Tempo de resposta dos dispositivos:

$R_B(2) = S_B[1+Q_B(1)] = 0,26[1+0,33] = 0,35$ seg.

$R_{SDC}(2) = P(0|1)[1/\mu(1)] + P(1|1)[2/\mu(2)] = 0,33x1/0,332 + 0,7x2/0,39 = 4,46$ seg.

Tempo de resposta do sistema:

$R(2) = R_B(2)V_B + R_{SDC}(2)V_{SDC} = 0,35x6+4,46 = 6,54$ seg.

Vazão do Sistema:

$X(2) = N/R(2) = 2/6,54 = 0,31$

Número médio e probabilidades:

$Q_B(2) = X(2)R_B(2)V_B = 0,31x0,35x6 = 0,64$

$P(2|2) = [X(2)/\mu(2)]P(1|1) = [0,31/0,39]x0,67 = 0,52$

$P(1|2) = [X(2)/\mu(1)]P(0|1) = [0,31/0,32]x0,33 = 0,32$

$P(0|2) = 1 - P(1|2) - P(2|2) = 1 - 0,52 - 0,32 = 0,16$

Exemplo 1: SDC

Iteração 3: n=3

Tempo de resposta dos dispositivos:

$R_B(3) = S_B(1+Q_B(2)) = 0{,}26x(1+0{,}64) = 0{,}43$ seg.

$R_{SDC}(3) = P(0|2)[1/\mu(1)] + P(1|2)[2/\mu(2)] + P(2|2)[3/\mu(3)] = 5{,}86$ seg.

Tempo de resposta do Sistema:

$R(3) = R_B(3)V_B + R_{SDC}(3)V_{SDC} = 8{,}42$ seg.

Vazão do Sistema:

$X(3) = N/R(3) = 3/8{,}42 = 0{,}36$

Número de usuários e probabilidades:

$Q_B(3) = X(3)R_B(3)V_B = 0{,}91$

$P(3|3) = [X(3)/\mu(3)]P(2|2) = 0{,}44$

$P(2|3) = [X(3)/\mu(2)]P(1|2) = 0{,}29$

$P(1|3) = [X(3)/\mu(1)]P(0|2) = 0{,}18$

$P(0|3) = 1 - P(1|3) - P(2|3) - P(3|3) = 0{,}09$

Exemplo 1: SDC

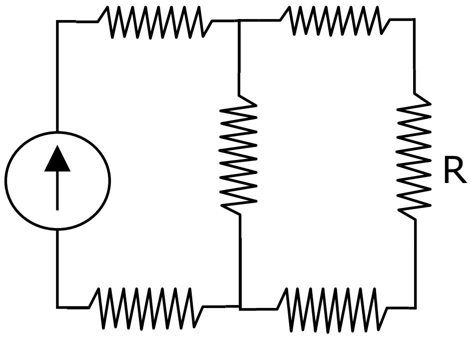
Vazão dos dispositivos para N = 3:
$\mathbf{X_B = XV_B} = 0{,}36x6 = 2{,}16$ jobs/seg
$\mathbf{X_{SDC} = XV_{SDC}} = 0{,}36x1 = 0{,}36$ jobs/seg
Utilização dos dispositivos para N = 3:
$U_B = XS_BV_B = 0{,}36x0{,}26x6 = 0{,}562$
$U_{SDC} = 1 - P(0|3) = 1 - 0{,}09 = 0{,}91$

- As seguintes conclusões podem ser tiradas sobre o sistema:
  - A vazão do sistema é 0,21, 0,31 e 0,36 jobs/segundo com 1, 2 e 3 usuários no sistema respectivamente;
  - O tempo de resposta do sistema é 4,68, 6,54 e 8,42 segundos com 1, 2 e 3 usuários no sistema;
  - O número médio de usuários no Disco B é 0,91 com 3 usuários no sistema;
  - O tempo de resposta do disco B é 0,43 segundos com 3 usuários no sistema;
  - O fator de utilização do disco B é 0,562 com 3 usuários no sistema;
  - As probabilidades de 0, 1, 2 e 3 usuários no SDC com 3 usuários no sistema são 0,09, 0,18, 0,29 e 0,44 respectivamente.

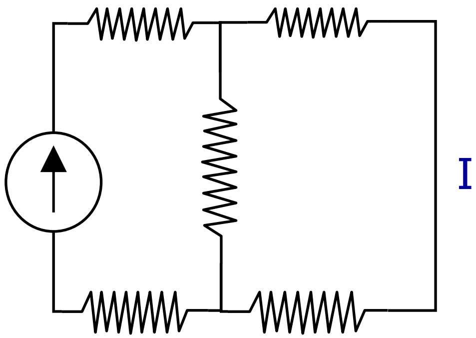
Modelo Equivalente

- Chandy, Herzog e Woo descobriram um método para a determinação do servidor equivalente de uma sub-rede de filas que produz resultados exatos para as redes de filas que obedecem à condições BCMP.
- O servidor equivalente é um servidor com capacidade dependente da carga (SDC). O método é inspirado no teorema de Norton de circuitos elétricos.

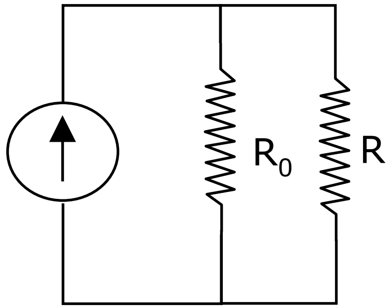
Teorema de Norton de circuitos elétricos:



Rede original



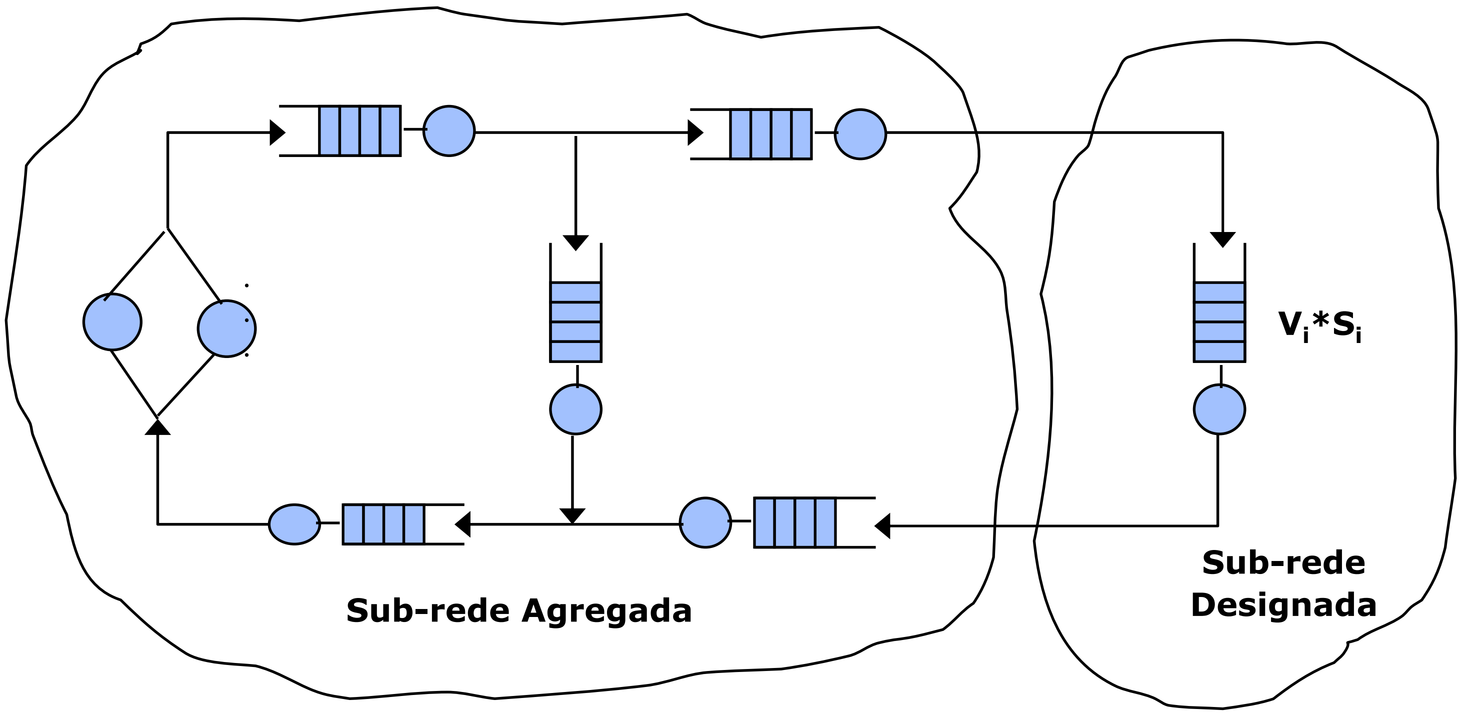
Rede equivalente



Rede encurtada

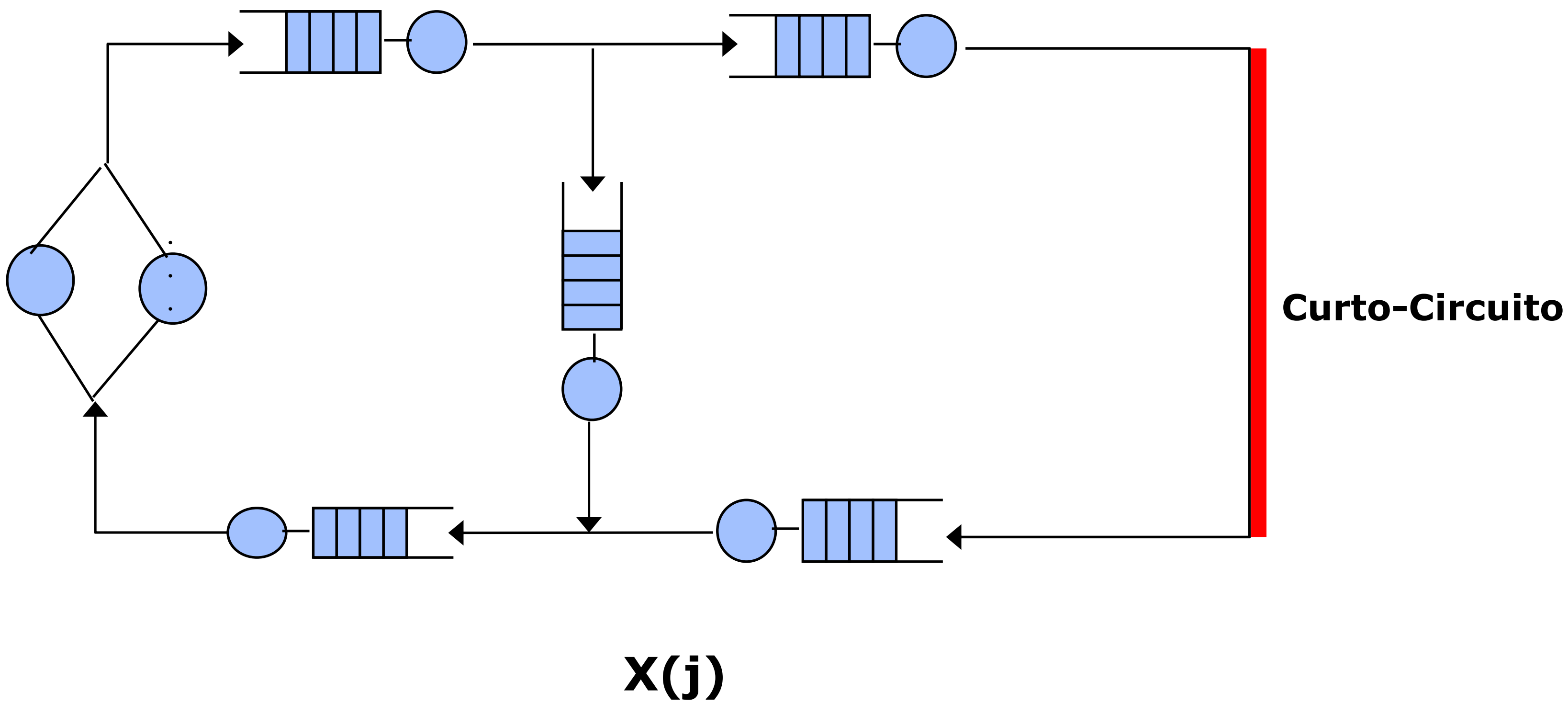
Método de Decomposição

- Dada uma rede de filas, esta rede será dividida em uma sub-rede da qual se deseja calcular o servidor equivalente, chamada de “rede agregada”, e uma sub-rede que permanecerá intocável chamada de “rede designada”.

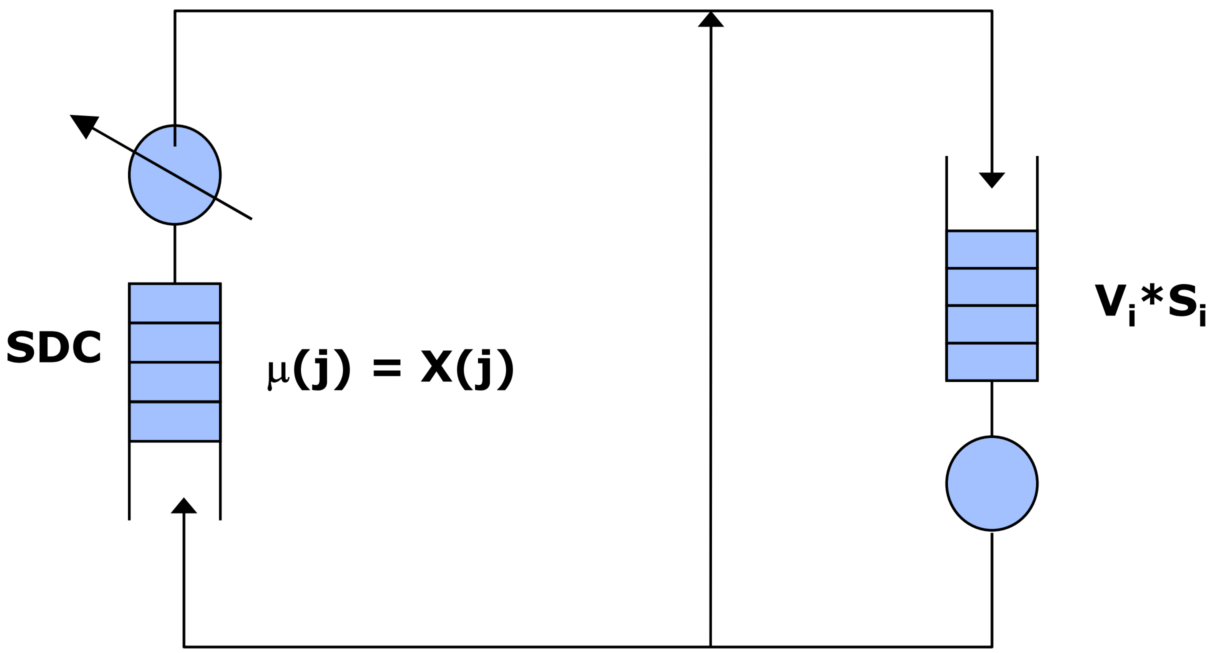


Rede “Curto-Circuitada”

- A vazão desta rede é X(j), calculada pelo método tradicional de redes fechadas



Rede Equivalente



Método de Decomposição

1. Selecione a sub-rede designada. O resto da rede é a sub-rede a ser agregada;
2. Crie o modelo curto-circuitado fazendo o tempo de serviço de todas as filas na sub-rede designada iguais a zero;
3. Resolva o modelo curto-circuitado pelo método AVM ou convolução;
4. Substitua a sub-rede agregada por um servidor dependente da carga. Este servidor possui taxa de serviço $\mu(j)$, igual à vazão da rede curto-circuitada $X(j)$ quando esta possui j usuários;
5. Resolva o modelo equivalente usando o procedimento de cálculo para servidores dependente da carga;
6. Aplique os resultados obtidos no modelo equivalente para a sub-rede designada;
7. Os valores dos parâmetros de performance da sub-rede designada são obtidos dos resultados da rede equivalente;
8. Os valores dos parâmetros de performance dos centros de serviço da sub-rede agregada podem ser obtidos através de probabilidades condicionais.

Desempenho da sub-rede agregada

- Distribuição do número de usuários: probabilidade de se ter **j** usuários na estação **i** da rede agregada, existindo N usuários no sistema, é dada por:

$$P[n_i = j \mid N(\text{sistema})] = \sum_{n=j}^{N} P[n_i = j \mid n(\text{agregado})] * P[n(\text{agregado}) \mid N(\text{sistema})]$$

ou

$$P[n_i = j \mid N(\text{sistema})] = \sum_{n=j}^{N} \{P[n_i = j \mid n(\text{agregado})] * P[n(\text{SDC}) \mid N(\text{sistema})]\}$$

- Número médio de usuários na i-ésima estação

$$Q_i = \sum_{j=1}^{N} jP[n_i = j \mid n(\text{sistema})]$$

Desempenho da sub-rede agregada

- **Vazão:** as vazões dos diversos centros de serviço são proporcionais às suas taxas de visitas:

$$X_i/V_i = X = X_j/V_j$$

onde

X é a vazão do sistema combinado;

V_i e V_j representam o número de visitas de cada usuário aos centros de serviço i e j.

- **Tempo de Resposta:** pode ser calculado usando o resultado de Little's:

$$R_i = Q_i * X_i$$

- **Fator de utilização:** pode ser calculado usando a lei da utilização:

$$U_i = X_i * S_i = X * D_i$$

Exemplo 2: Decomposição Hierárquica

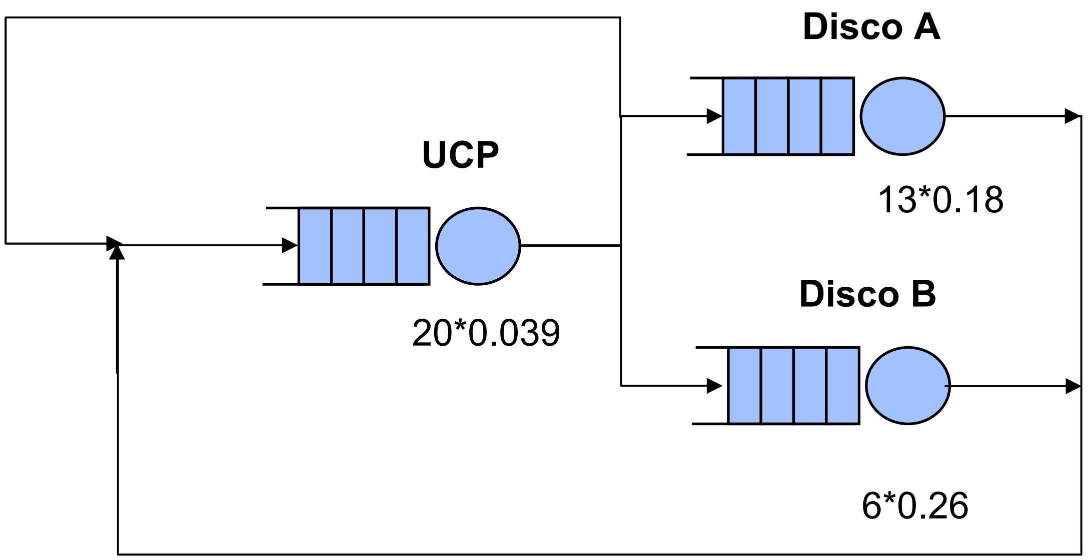
- Considere o modelo do servidor central: 1 UCP e 2 discos com um grau de multiprogramação igual a 3.

  Os tempos médio de serviço são:

  $S_{UCP} = 0.039$, $S_A = 0.18$, $S_B = 0.26$

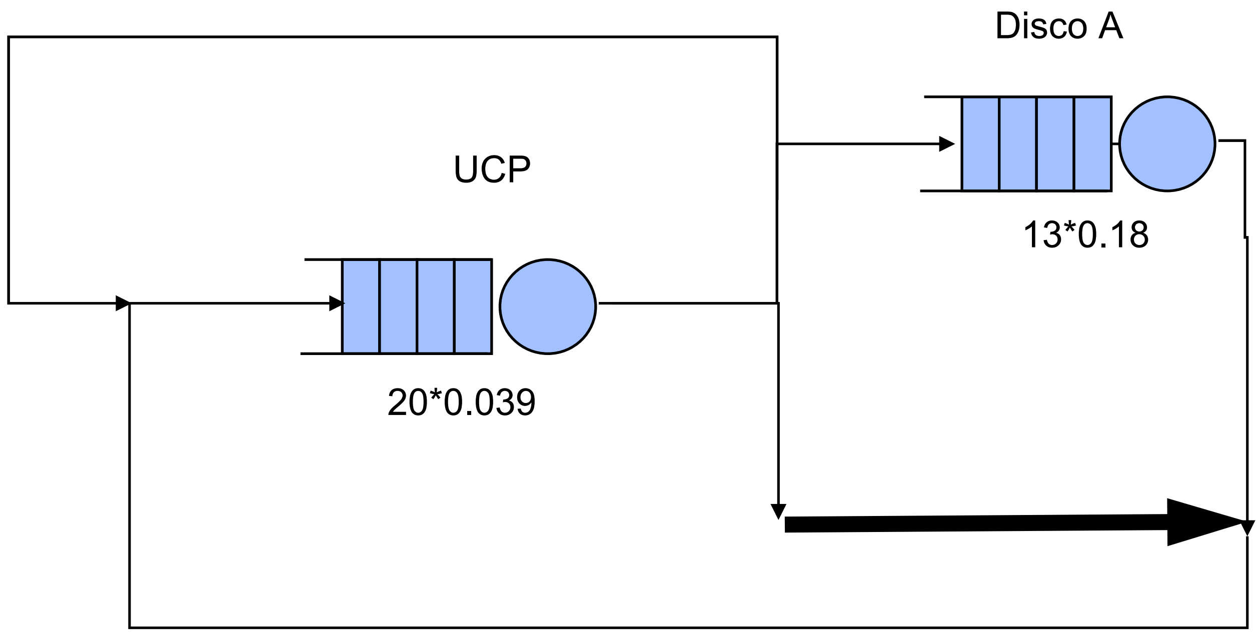
  As taxas de visitas são:

  $V_{UCP} = 20$, $V_A = 13$ e $V_B = 6$



Sistema original: Modelo do servidor central

Exemplo 2: Decomposição Hierárquica

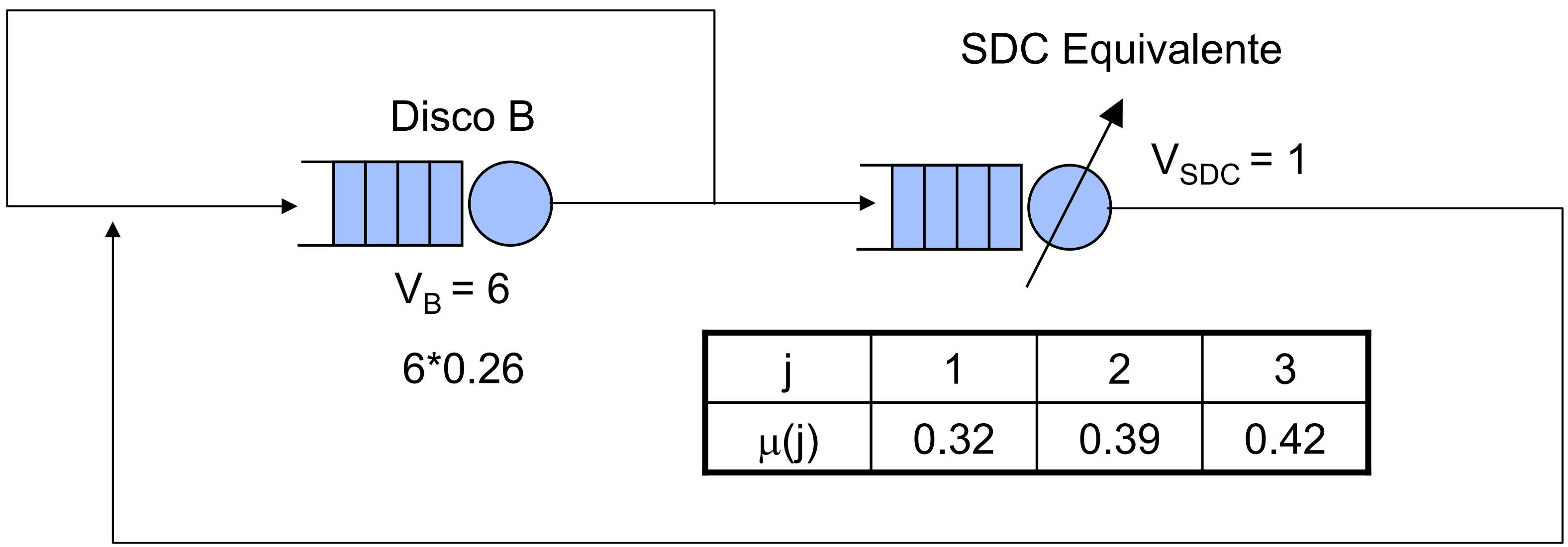


Modelo curto-circuitado

Sub-rede agregada composta pela UCP e Disco A

Sub-rede designada é composta pelo Disco B

Exemplo 2: Decomposição Hierárquica



j	1	2	3
μ(j)	0.32	0.39	0.42

Modelo equivalente para análise do disco B

SDC substitui a UCP e o Disco A

Exemplo 2: Decomposição Hierárquica

- Cálculo da demanda total de serviço em cada centro de serviço no modelo original:

  $D_{UCP} = 20*0.039 = 0.78$

  $D_A = 13*0.18 = 2.34$

  $D_B = 6*0.26 = 1.56$

- Vamos resolver o modelo curto-circuitado usando o método de convolução para calcularmos a distribuição do número de usuários no sistema. Escolhemos, como anteriormente o fator de escala α= 1/ 0.78 que resulta nos seguintes valores:

  $y_{UCP} = 1 \quad y_A = 3$

- Cálculo da constante de normalização G(N):

n	y_{ucp}= 1	y_A= 3	
0	1	1	G(0) = 1
1	1	4	G(1) = 4
2	1	13	G(2) = 13
3	1	40	G(3) = 40

Exemplo 2: Decomposição Hierárquica

- A vazão do sistema para grau de multiprogramação 3 é:

$$X(1) = \alpha^*[G(0)/G(1)] = 0.321$$
$$X(2) = \alpha^*[G(1)/G(2)] = 0.394$$
$$X(3) = \alpha^*[G(2)/G(3)] = 0.417$$

- Desta forma, o servidor equivalente será um centro de serviço com capacidade dependente da carga e com uma taxa de serviço variável igual a:

$$\mu(1) = X(1) = 0.321$$
$$\mu(2) = X(2) = 0.394$$
$$\mu(3) = X(3) = 0.417$$

Exemplo 2: Decomposição Hierárquica

- Probabilidade de j usuários no disco A quando existem n usuários no sistema, no modelo curto-circuitado calculado pelo método de convolução:

$$P(n_A = j \mid n) = \frac{y_A^{\,j}}{G(n)} * [G(n-j) - y_A * G(n-j-1)]$$

n	P(n_A= j | n)			
	j=0	j=1	j=2	j=3
0	1			
1	0.250	0.750		
2	0.077	0.321	0.692	
3	0.025	0.075	0.225	0.675

Exemplo 2: Decomposição Hierárquica

O modelo equivalente é composto pelo disco B e o servidor de capacidade variável SDC. Este modelo já foi resolvido anteriormente e produziu os seguintes resultados:

1. A vazão do sistema é 0.21, 0.31 e 0.36 usuários/seg. com 1, 2 e 3 usuários no sistema;
2. O tempo de resposta é: 4.68, 6.54 e 8.42 para N=1, 2 e 3 usuários respectivamente;
3. O tamanho médio da fila para o disco B com N = 3 é 0.91;
4. O tempo médio de resposta para o disco B com N = 3 é 0.43 segundos;
5. O fator de utilização do disco B com N = 3 é 0.562

Exemplo 2: Decomposição Hierárquica

- Para se obter os parâmetros de desempenho da sub-rede agregada deve-se proceder ao cálculo das probabilidades condicionais, como mostrado a seguir:
  - Da solução do exemplo anterior (exemplo SDC) já foi calculado que a probabilidade de se ter 0, 1, 2 ou 3 usuários no SDC quando se tem 3 usuários no sistema é respectivamente: 0.09, 0.18, 0.29 e 0.44;
  - Estes valores juntamente com os valores da tabela de probabilidades do número de usuários no disco A calculada pelo modelo curto-circuitado, são suficientes para se determinar os parâmetros da sub-rede agregada.

- Cálculo da probabilidade de se ter 0,1,2 e 3 usuários no disco A quando se tem 3 usuários no sistema:

$$\begin{aligned} P(n_A=0|N=3) = \; & P(n_A=0|n=0)^* \, P(n=0|N=3)+ \\ & P(n_A=0|n=1)^* \, P(n=1|N=3)+ \\ & P(n_A=0|n=2)^* \, P(n=2|N=3)+ \\ & P(n_A=0|n=3)^* \, P(n=3|N=3) \\ = \; & 1x0.09 + 0.250{*}0.18 + 0.077{*}0.29 + 0.025{*}0.44 \;=\; 0.166 \end{aligned}$$

Exemplo 2: Decomposição Hierárquica

- De forma similar podemos calcular:

  $P(n_A=1|N=3) = 0.750*0.18+0.231*0.29+0.075*0.44 = 0.233$
  $P(n_A=2|N=3) = 0.692*0.29+0.225*0.44 = 0.\ 3$
  $P(n_A=3|N=3) = 675*0.44 = 0.\ 3$

- O número médio de usuários no disco A pode ser calculado como:

$$Q_A = \sum_{j=1}^{N} jP[n_A = j\,|\,N(\text{sistema})] = 1*0.233 + 2*0.3 + 3*0.3$$

- Similarmente, o número médio de usuários na UCP é calculado, e o seu valor é 0.36 usuários.

- A vazão da UCP e do disco A é calculada pela lei de fluxo forçado:

  $X_{UCP} = X*V_{UCP} = 0.36*20 = 7.2$ usuários/seg.
  $X_A = X*V_A = 0.36*13 = 4.68$ usuários/seg.

Exemplo 2: Decomposição Hierárquica

- Os fatores de utilização da UCP e do disco A são:

$$U_{UCP} = X^*D_{UCP} = 0.36^*0.78 = 0.281$$
$$U_A = X^*D_A = 0.36^*2.34 = 0.843$$

- O tempo médio de resposta é calculado usando-se o resultado de Little:

$$R_{UCP} = Q_{UCP}/X_{UCP} = 0.36/7.2 = 0.05 \text{ seg.}$$
$$R_A = Q_A/X_A = 0.36/4.68 = 0.37 \text{ seg.}$$

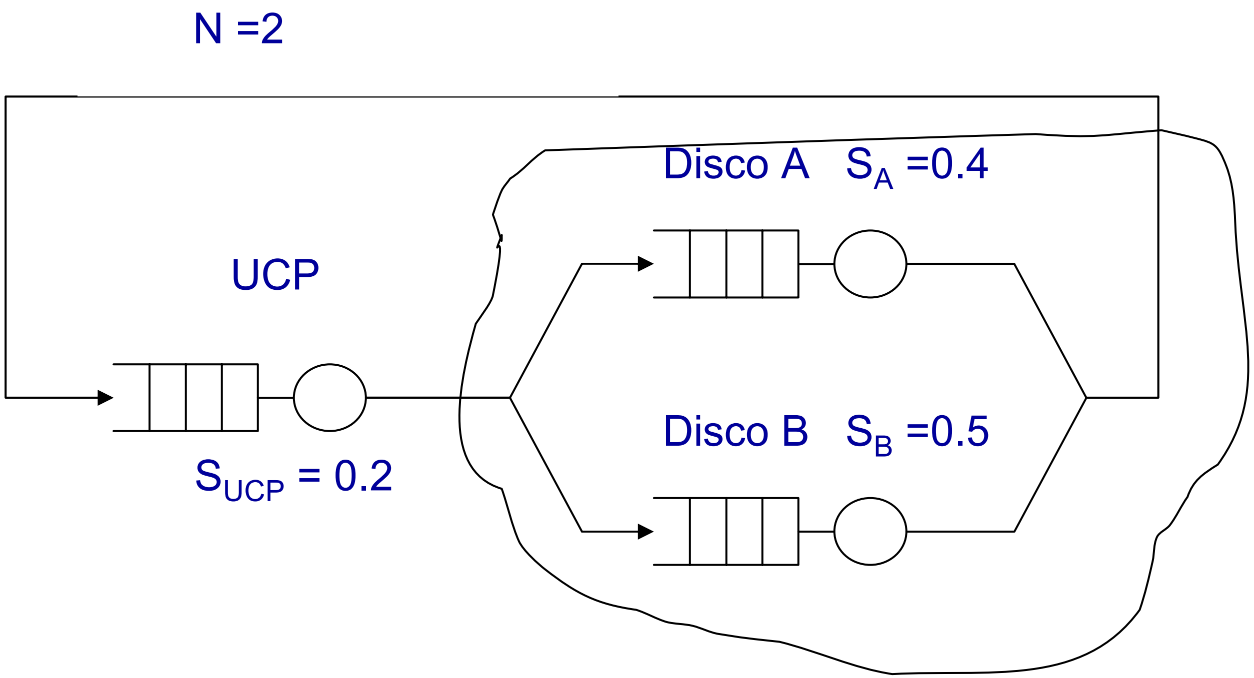
- É importante lembrar que para o modelo equivalente a taxa de visitas usada deve ser $V_B = 6$ e $V_{SDC} = 1$. Se isto não for feito não teremos os resultados corretos.

Conclusão

- O método de decomposição hierárquica produz resultados exatos para redes com probabilidade de estados em forma de produto, isto é, que obedecem a regra BCMP.
- Este método é muito bom para ser aplicado quando se tem uma rede que não possui todos os seus centros de serviço obedecendo a regra BCMP.
- Neste caso, define-se a sub-rede agregada com sendo aquele sub-conjunto que contenha somente os centros de serviço que obedecem a regra BCMP.
- O modelo equivalente, que possui menos centros de serviço que a rede original, então é resolvido por algum método para redes que não obedecem a regra BMCP, ou então por simulação. Este fato, reduz significativamente o tempo de solução pois a rede possui menos elementos.

Exercício 1

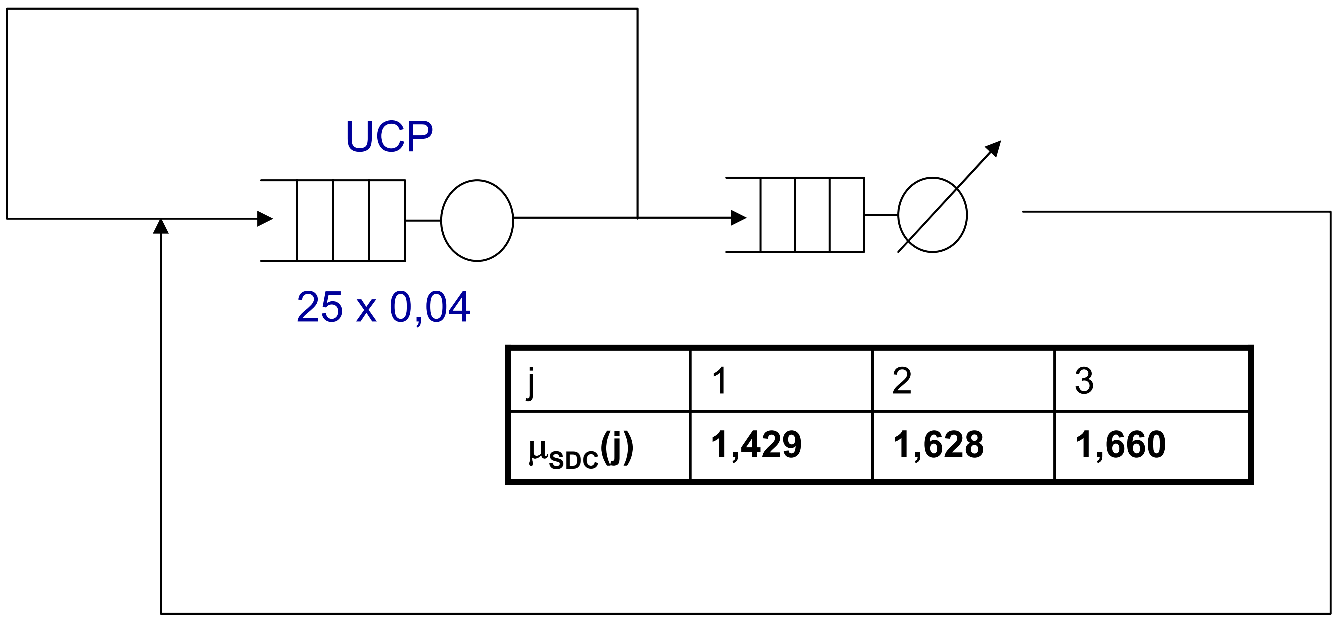
- Seja o seguinte sistema de filas onde a UCP, o disco A e o disco B possuem tempos médios de serviço respectivamente iguais a 0.2, 0.4 e 0.5 seg, taxa de visitas à UCP do disco A e disco B respectivamente 1, 0.5 e 0.5 e grau de multiprogramação 2. Na solução pelo método hierárquico, determine o SDC equivalente aos discos A e B do sistema pelo método de Análise do Valor Médio (MVA).



$V_{UCP} = 1$
$V_A = 0.5$
$V_B = 0.5$

Exercício 2

- Determine a vazão do sistema e o tempo de resposta para o sistema da figura a seguir utilizando AVM dependente de carga. As taxas de serviço $\mu(j)$ do centro de serviço dependente de carga como função do número de programas j no centro de serviço são 1,429, 1,628 e 1,660, respectivamente para j=1, 2, 3.



j	1	2	3
μ_{SDC}(j)	**1,429**	**1,628**	**1,660**

- Resp.: X=0.588, 0.796 e 0.892 para n=1,2,3; R=1.700, 2.506 e 3.365 para n=1,2,3.

Fim do módulo
Servidor Dependente de Carga
e
Decomposição Hierárquica