Análise das Redes de Filas pelo VALOR MÉDIO

(Chapter 34 – Raj Jain).

Este capítulo estende os resultados apresentados no Capítulo 11 (Leis Operacionais).

1. Análise de Redes de Filas Abertas

Redes de Filas Abertas podem ser usadas como modelo que representa sistemas de processo de transações como sistemas de reservas de companhias aéreas ou sistemas bancários. Nestes sistemas, a taxa de chegada das transações não é dependente da carga no sistema computacional. As chegadas das transações são modeladas como um processo de Poisson com taxa de chegada λ.

Vamos assumir que todos os dispositivos no sistema computacional podem ser modelados como:

a) Centro de Serviço com Capacidade Fixa (por exemplo, servidor único com distribuição exponencial do tempo de serviço); ou

b) Centros de Atraso (infinitos servidores com distribuição exponencial do tempo de serviço), ou seja, servidores que não apresentam espera em filas. (Terminais).

Para os Serviços com Capacidade Fixa (CF), quando um usuário chega na <u>estação i</u>, em média, ele irá encontrar <u>Q_i</u> usuários à sua frente. Portanto, ele deverá esperar o atendimento de <u>Q_i</u> usuários, onde cada um leva, em média, um tempo de serviço S_i. (inclusive o usuário que acabou de chegar).

Dessa forma, o tempo de resposta médio para o usuário que chega à <u>estação i</u> será:

$$R_i = S_i(1+Q_i) \quad (I)$$

~~Considerando o fluxo balanceado~~

Nesta equação, assume-se que o serviço é <u>sem memória</u>; se fosse avaliar o tempo de resposta de um usuário <u>sem considerar</u> a condição sem memória, seria necessário saber o tempo que o usuário que está sendo atendido já consumiu.

A <u>equação anterior</u>, em conjunto com as <u>Leis Operacionais</u>, são suficientes para calcular o <u>Valor Médio</u> dos parâmetros de desempenho do sistema.

Assumindo-se Fluxo Balanceado, a Vazão do Sistema (X) é igual à taxa de chegada:

$$X = \lambda$$

A Vazão do iésimo dispositivo será:

$$X_i = X V_i \text{ (Leis Operacionais)}$$

A taxa de Utilização é dada por:

$$U_i = S_i . X_i = S_i . X . V_i \text{ (Leis Operacionais).}$$

$$U_i = X . D_i \Rightarrow \underline{U_i = \lambda . D_i}$$

O tamanho da fila do iésimo dispositivo, de acordo com a "Lei de Little" é: (nº de usuários no dispositivo)

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

$$\underline{Q_i = U_i . (1 + Q_i)}$$

ou $\boxed{Q_i = \dfrac{U_i}{1 - U_i}}$ (idêntica ao valor obtido nas filas M/M/1)

$$\Rightarrow R_i = S_i (1 + Q_i) = S_i \left(1 + \frac{U_i}{1 - U_i} \right)$$

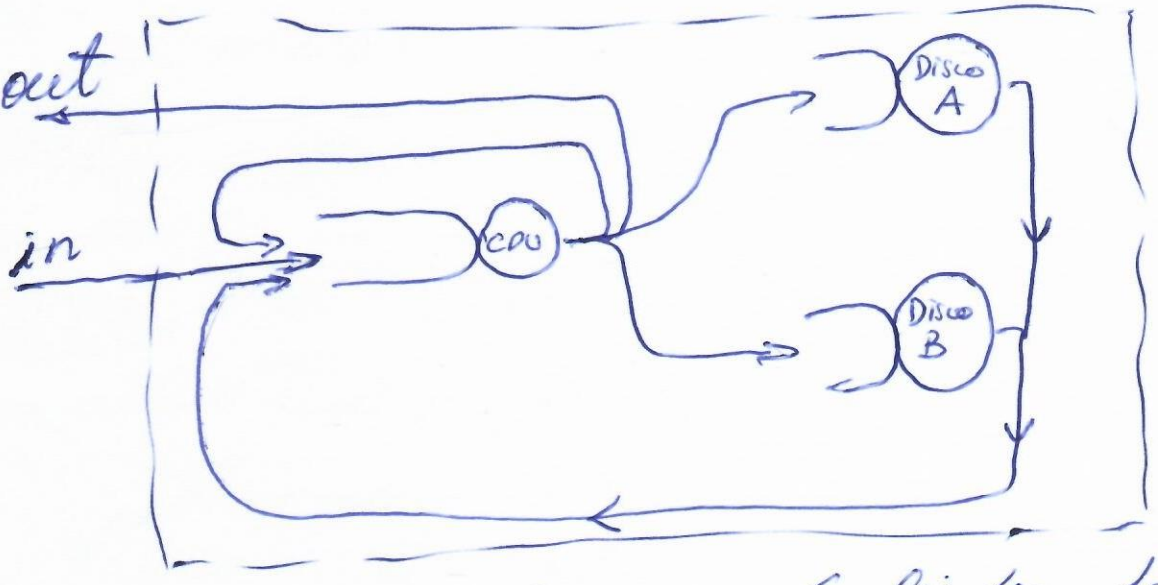
$$\boxed{R_i = \frac{1}{1 - U_i} . S_i}$$

Num Centro de Atraso, a equação (I) anterior é modificada pois não há espera na fila: $\boxed{R_i = S_i}$ (II).

$$\Rightarrow Q_i = X_i \cdot R_i = X_i \cdot S_i = U_i$$

(neste caso, o Q_i representa o número de usuários recebendo serviço).

Exemplo 1: Considere o seguinte sistema.



Medidas realizadas com 6 clientes fazendo requisições ao servidor produziram os seguintes dados:

Tempo de Observação: 3600 seg
Número de Requisições: 10800
Tempo de Ocupação da CPU: 1728 seg
Tempo de Ocupação do Disco A: 1512 seg
Tempo de Ocupação do Disco B: 2592 seg.

Número de Visitas ao Disco A = 75600
Número de Visitas ao Disco B = 86.400

Desta forma, tem-se:

Vazão do Sistema: $X = \frac{10.800}{3600} \Rightarrow \underline{X = 3 \text{ seg/h}}$

$V_A = \frac{75.600}{10.800} \rightarrow \underline{V_A = 7}$ visitas ao Disco A por requisição de cliente

$V_B = \frac{86400}{10.800} \rightarrow \underline{V_B = 8}$ visitas ao Disco B por requisição de cliente.

$V_{CPU} = 1 + V_A + V_B = 1 + 7 + 8 \Rightarrow \underline{V_{CPU} = 16}$

$D_{CPU} = \frac{1728}{10.800} \Rightarrow D_{CPU} = 0{,}16$ seg de CPU por requisição de cliente.

$D_A = \frac{1512}{10.800} \rightarrow D_A = 0{,}14$ seg de Disco A por requisição de cliente.

$D_B = \frac{2592}{10.800} \Rightarrow D_B = 0{,}24$ seg de Disco B por requisição de cliente.

$S_{CPU} = \frac{0{,}16}{16} \rightarrow \underline{S_{CPU} = 0{,}01 \text{ seg}}$ tempo de serviço na CPU por visita.

$S_A = \frac{0{,}14}{7} \rightarrow \underline{S_A = 0{,}02 \text{ seg}}$ tempo de serviço no Disco A por visita.

$S_B = \frac{0{,}24}{8} \rightarrow \underline{S_B = 0{,}03 \text{ seg}}$ tempo de serviço no Disco B por visita.

Pode-se calcular agora o <u>Fator de Utilização</u>. ⑥

$(U_i = X . D_i)$

$U_{CPU} = X . D_{CPU} = 3 . 0{,}16 \rightarrow \underline{U_{CPU} = 0{,}48}$

$U_A = X . D_A = 3 . 0{,}14 \longrightarrow \underline{U_A = 0{,}42}$

$U_B = X . D_B = 3 . 0{,}24 \longrightarrow \underline{U_B = 0{,}72}$

Em seguida, pode-se calcular o <u>Tempo de Resposta</u>,

$$R = \sum_{i=1}^{3} V_i . R_i \quad e \quad R_i = S_i (1 + Q_i) = \frac{1}{1 - U_i} . S_{CPU}$$

Para CPU: $R_{CPU} = S_{CPU} (1 + Q_{CPU}) = \frac{1}{1 - U_{CPU}} . S_{CPU}$

$$R_{CPU} = \frac{1}{1 - 0{,}48} . 0{,}01 \rightarrow \underline{R_{CPU} = 0{,}0192 \text{ seg}}$$

$$R_A = \frac{1}{1 - U_A} . S_A = \frac{1}{1 - 0{,}42} . 0{,}02 \rightarrow \underline{R_A = 0{,}0345 \text{ seg}}$$

$$R_B = \frac{1}{1 - U_B} . S_B = \frac{1}{1 - 0{,}72} . 0{,}03 \longrightarrow \underline{R_B = 0{,}107 \text{ seg}}$$

$$\Rightarrow R = 16 . 0{,}0192 + 7 . 0{,}0345 + 8 . 0{,}107 \Rightarrow \underline{\underline{R = 1{,}406 \text{ seg}}}$$

a) Qual o novo desempenho do Sistema se tivermos 8 clientes?

$$X = 3 \cdot \frac{8}{6} \rightarrow X = 4 \text{ req/h}$$

Os Fatores de Utilização serão:

$$U_{CPU} = X \cdot D_{CPU} = 4 \cdot 0{,}16 \rightarrow U_{CPU} = 0{,}64$$

$$U_A = X \cdot D_A = 4 \cdot 0{,}14 \rightarrow U_A = 0{,}56$$

$$U_B = X \cdot D_B = 4 \cdot 0{,}24 \rightarrow U_B = 0{,}96$$

Os tempos de resposta serão:

$$R_{CPU} = \frac{1}{1 - U_{CPU}} \cdot S_{CPU} = \frac{1}{1 - 0{,}64} \cdot 0{,}01 \rightarrow R_{CPU} = 0{,}0278 \text{ seg}$$

$$R_A = \frac{1}{1 - U_A} \cdot S_A = \frac{1}{1 - 0{,}56} \cdot 0{,}02 \rightarrow R_A = 0{,}0455 \text{ seg}$$

$$R_B = \frac{1}{1 - U_B} \cdot S_B = \frac{1}{1 - 0{,}96} \cdot 0{,}03 \rightarrow R_B = 0{,}75 \text{ seg}$$

$$R = 16 \times 0{,}0278 + 7 \cdot 0{,}0455 + 8 \cdot 0{,}75 \rightarrow R = 6{,}76 \text{ seg}$$

(4,8 vezes maior do que para 6 clientes)

b) Qual o novo desempenho do Sistema se usarmos um Cache para o Disco B com uma taxa de acerto de 50%, embora ele aumente a sobrecarga da CPU em 30% e o tempo de serviço do disco B em 10%? (8)

. Como não precisa ir ao Disco B com o ~~Cache~~ Cache em 50% $\Rightarrow v_B = \frac{8}{2} \rightarrow \underline{v_B = 4}$ visitas ao disco B por requisição de cliente.

. 30% de sobrecarga na CPU $\rightarrow S_{CPU} = 0{,}01 \cdot 1{,}3$

$$\Rightarrow \underline{S_{CPU} = 0{,}013 \text{ seg}}$$

. 10% em tempo de Disco B $\Rightarrow S_B = 0{,}03 \cdot 1{,}1 \Rightarrow \underline{S_B = 0{,}033 \text{ seg}}$

Temos que $D_i = V_i \cdot S_i$

$$\begin{cases} D_{CPU} = V_{CPU} \cdot S_{CPU} = 16 \cdot 0{,}013 \rightarrow \underline{D_{CPU} = 0{,}208 \text{ seg}} \\ D_B = V_B \cdot S_B = 4 \cdot 0{,}033 \longrightarrow \underline{D_B = 0{,}132 \text{ seg}} \end{cases}$$

Agora pode-se calcular a Taxa de Utilização:

$$U_{UCP} = X \cdot D_{UCP} = 3 \cdot 0{,}208 \Rightarrow \underline{U_{UCP} = 0{,}624}$$

$$U_A = X \cdot D_A = 3 \cdot 0{,}14 \Rightarrow \underline{U_A = 0{,}42}$$

$$U_B = X \cdot D_B = 3 \cdot 0{,}132 \Rightarrow \underline{U_B = 0{,}396}$$

Em seguida, pode-se calcular o Tempo de Resposta.

$$R_i = \frac{1}{1-U_i} \cdot S_i$$

$$R_{CPU} = \frac{1}{1-U_{CPU}} \cdot S_{CPU} = \frac{1}{1-0{,}624} \cdot 0{,}013 \Rightarrow \underline{R_{CPU} = 0{,}0346 \text{ seg}}$$

$$R_A = \frac{1}{1-U_A} \cdot S_A = \frac{1}{1-0{,}42} \cdot 0{,}02 \Rightarrow \underline{R_A = 0{,}0345 \text{ seg}}$$

$$R_B = \frac{1}{1-U_B} \cdot S_B = \frac{1}{1-0{,}396} \cdot 0{,}033 \Rightarrow \underline{R_B = 0{,}0546 \text{ seg}}$$

$$R = 16 \cdot 0{,}0346 + 7 \cdot 0{,}0345 + 4 \cdot 0{,}0546$$

$$\Rightarrow \underline{\underline{R = 1{,}013 \text{ seg}}} \text{ (melhorou } 28{,}8\%)$$

$$\left(\frac{1{,}406 - 1{,}013}{1{,}406}\right).$$

c) Qual o novo desempenho do Sistema se utilizarmos um Servidor com um Disco A e dirigirmos todos os acessos a ele?

$$\Rightarrow V_B = 0 \Rightarrow V_A = 7 + 8 = 15$$

$$\underline{D_{UCP} = 0{,}16 \text{ seg}}$$

$$D_A = V_A \cdot S_A = 15 \cdot 0{,}02 \rightarrow \underline{D_A = 0{,}3 \text{ seg}}$$

A Taxa de Utilização:

$$U_{UCP} = X \cdot D_{UCP} = 3 \cdot 0{,}16 \rightarrow \underline{U_{UCP} = 0{,}48}$$

$$U_A = X \cdot D_A = 3 \cdot 0{,}3 \rightarrow \underline{U_A = 0{,}9}$$

E ~~o~~ Tempo de Resposta:

$$R_{CPU} = \frac{1}{1-U_{CPU}} \cdot S_{CPU} = \frac{1}{1-0{,}48} \cdot 0{,}01 \quad R_{CPU} = 0{,}0192\,seg$$

$$R_A = \frac{1}{1-U_A} \cdot S_A = \frac{1}{1-0{,}9} \cdot 0{,}02 \Rightarrow R_A = 0{,}2\,seg$$

$$R = 0{,}0192 \times 16 + 0{,}2 \times 15 \Rightarrow$$

$$R = 3{,}31\,seg \quad (\text{piorou } 135\%!!)$$

$$\left(\frac{3{,}31 - 1{,}406}{1{,}406}\right)$$

2. Análise de Redes de Filas Fechadas

Essa análise de Valor Médio aplica-se a Redes com uma variedade de disciplinas de Serviços e de distribuição de tempo.

Vamos iniciar nossa análise para Centros de Serviço com Capacidade Fixa.

Seja essa Rede de Fila fechada com N usuários.

O tempo de resposta é dado por: (para cada dispositivo)

$$R_i(N) = S_i\,[1 + \underbrace{Q_i(N-1)}]$$

Dado o desempenho ~~para~~ para (N-1) usuários, esta equação junto com as "Leis Operacionais" podem calcular o desempenho com N usuários.

Quando um usuário chega ao iésimo dispositivo ele encontra $Qi(N-1)$ usuários à frente (incluindo o que está sendo atendido) e espera por $Qi(N-1).Si$ antes de receber seu serviço.

Assume-se que o serviço é sem memória.
O Tempo de Resposta é dado por:

$$R(N) = \sum_{i=1}^{M} Vi \cdot Ri(N)$$

Conforme já foi visto nas Leis Operacionais:

$$X(N) = \frac{N}{R(N)+Z}$$

A Vazão do dispositivo i é dada por:

$$Xi(N) = X(N) \cdot Vi$$

De acordo com a "Lei de Little" tem-se:

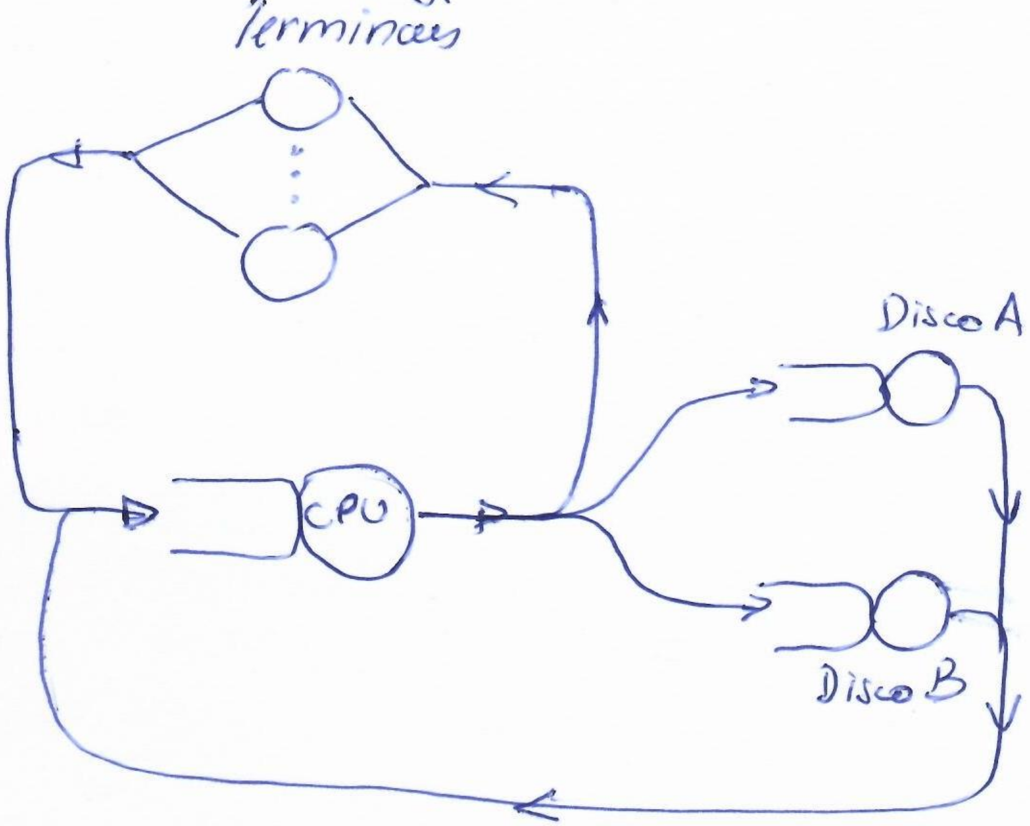
$$Qi(N) = Xi(N) \cdot Ri(N) = X(N) \cdot Vi \cdot Ri(N)$$

Por outro lado, se trata-se de um Centro de Atraso, não há tempo de espera antes do serviço a ser realizado, e o tempo de resposta é igual ao tempo de Serviço $\Rightarrow$ $Ri(N) = Si$

Todo o processo inicia-se com $N=0$ e vai aumentando interativamente.

Exemplo 2

Seja o Sistema de Rede de Filas fechada apresentado a seguir: (Servidores com Timesharing).



Cada requisição do Usuário gera 10 requisições para o Disco A e 5 requisições para o Disco B.

O tempo de serviço por visita ao Disco A é de 300 ms e ao Disco B é de 200 ms.

Cada requisição leva no total 2 segundos de CPU e o Usuário pensa em média 4 segundos

$S_A = 300\,ms = 0{,}3\,seg$ $\quad V_A = 10 \rightarrow D_A = V_A . S_A$

$D_A = 0{,}3 . 10 \longrightarrow D_A = 3\,seg.$

$S_B = 0{,}2 \quad V_B = 5 \rightarrow D_B = V_B . S_B = 0{,}2 . 5 \rightarrow D_B = 1\,seg$

$D_{CPU} = 2\,seg \rightarrow D_{CPU} = V_{CPU} . S_{CPU} \rightarrow S_{CPU} = \frac{2}{V_{CPU}}$

$V_{CPU} = V_A + V_B + 1 = 10 + 5 + 1 = 16$

$\Rightarrow \quad S_{CPU} = \frac{2}{16} \longrightarrow \underline{S_{CPU} = 0{,}125 \, seg}$

$\underline{Z = 4}; \ \underline{N = 20};$

Calculando o desempenho de forma interactiva tem-se:

Iniciação:

- $\underline{N = 0}: Q_{CPU} = 0 \quad Q_A = 0 \quad Q_B = 0$

- $\underline{N = 1}$

Tempo de Resposta:

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

$R_{CPU}(1) = 0{,}125(1 + 0) = 0{,}125 \, seg$

$R_A(1) = 0{,}3(1 + 0) = 0{,}3 \, seg$

$R_B(1) = 0{,}2(1 + 0) = 0{,}2 \, seg.$

$R(1) = 0{,}125 . 16 + 0{,}3 . 10 + 0{,}2 . 5 \quad \underline{R(1) = 6 \, seg}$

Vazão do Sistema:

$X = \frac{N}{R + Z} = \frac{1}{6 + 4} \longrightarrow \underline{X = 0{,}1 \, req/seg}$

Número Médio de Usuário

$$Q_i = X_i . R_i = X . V_i . R_i$$

$$\begin{cases} Q_{CPU} = 0,1 . 16 . 0,125 = \underline{0,2} \\ Q_A = 0,1 . 10 . 0,3 = \underline{0,3} \\ Q_B = 0,1 . 5 . 0,2 = \underline{0,1} \end{cases}$$

- **N=2**

Tempo de Resposta:

$$R_i(2) = S_i\,(1 + Q_i(1))$$

$$R_{CPU}(2) = 0,125(1+0,2) = 0,15 \text{ seg.}$$

$$R_A(2) = 0,3\,(1+0,3) = 0,39 \text{ seg.}$$

$$R_B(2) = 0,2\,(1+0,1) = 0,22 \text{ seg.}$$

$$R(2) = 0,15 \cdot 16 + 0,39 \cdot 10 + 0,22 \cdot 5 \Rightarrow \underline{R(2) = 7,4 \text{ seg}}$$

Vazão do Sistema

$$X = \frac{N}{R+Z} = \frac{2}{7,4+1} \longrightarrow \underline{X = 0,175 \text{ req/seg}}$$

Número Médio de Usuários

$$Q_{CPU} = 0,175 . 16 . 0,15 \rightarrow \underline{Q_{CPU} = 0,421}$$

$$Q_A = 0,175 . 10 . 0,39 \rightarrow \underline{Q_A = 0,684}$$

$$Q_B = 0,175 . 5 . 0,22 \rightarrow \underline{Q_B = 0,193}$$

Fazendo-se iterativamente até $N=20$

tem-se:

$$R(20) = 56{,}016 \text{ seg}$$

$$VAZÃO(20) = 0{,}333 \text{ reg/seg}$$

$$\begin{cases} Q_{CPU} = 1{,}991 \\ Q_A = 16{,}177 \\ Q_B = 0{,}5 \end{cases}$$