

Análise Operacional de Redes de Fila

Prof. Fabrício B. Gonçalves

Instituto Federal Fluminense
Campus Bom Jesus do Itabapoana

Agenda



1 Introdução

Agenda



- 1 Introdução
- 2 Classificação de Redes de Filas

Agenda



- 1 Introdução
- 2 Classificação de Redes de Filas
- 3 Variáveis Operacionais

Agenda



- 1 Introdução
- 2 Classificação de Redes de Filas
- 3 Variáveis Operacionais
- 4 Leis Operacionais para Redes de Filas

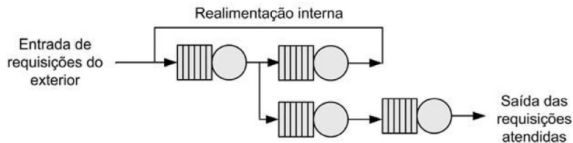
Introdução

Introdução



Uma rede de filas é uma coleção de centros de serviços e clientes. Uma rede de filas caracteriza-se por ser um sistema que contém duas ou mais filas conectadas ou não entre si. As requisições chegam ao sistema por uma ou mais entradas do exterior, trafegam de uma fila para outra, e saem por uma ou mais saídas.

Introdução



Classificação de Redes de Filas

Classificação de Redes de Filas



Uma rede de filas pode ser classificada quanto:

- Tipo;
- Carga;
- Número de cargas.

Tipos de Filas

Classificação de Redes de Filas

Tipo de Filas



Quanto ao tipo, uma rede de filas pode ser classificada como:

- Filas abertas;
- Filas fechadas;
- Filas mistas.

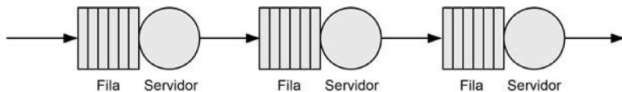
Classificação de Redes de Filas

Tipo de Filas: Filas Abertas



Filas Abertas

São aquelas que admitem chegadas de requisições do exterior e saídas das requisições processadas para o exterior.



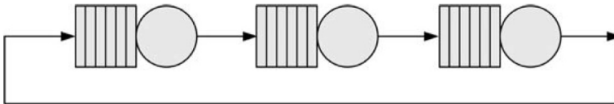
Classificação de Redes de Filas

Tipo de Filas: Filas Fechadas



Filas Fechadas

Essas filas não admitem entradas nem saídas de requisições. Nelas, o número de requisições é sempre constante.



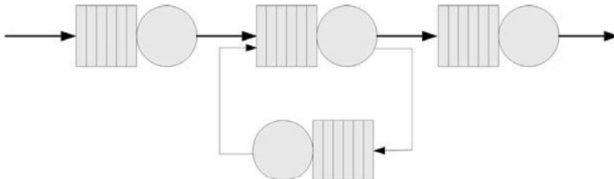
Classificação de Redes de Filas

Tipo de Filas: Filas Mistas



Filas Mistas

Filas mistas possuem partes abertas e fechadas ao mesmo tempo. Nesse caso, parte do sistema se comporta de forma fechada e parte do sistema se comporta de forma aberta.



Tipos de Cargas

Classificação de Redes de Filas

Tipos de Cargas



A carga de requisições que trafega em uma rede pode ser classificada como:

- **Cargas abertas:** Uma rede com carga aberta permite que o sistema receba um número ilimitado de clientes. Por exemplo, um roteador recebe infinitos pacotes para encaminhamento. Cargas de trabalho do tipo transação são especificadas pelo parâmetro de taxa de chegadas (λ) de novos clientes ao sistema. A população de clientes varia com o tempo e os clientes devem sair do sistema após o término do serviço. Modelos de filas com carga do tipo transação são chamados de modelos abertos.

Classificação de Redes de Filas

Tipos de Cargas



Continuação:

- **Cargas fechadas:** Um sistema de carga fechada permite apenas a entrada de um número de clientes no sistema limitado e conhecido. Por exemplo, cargas interativas e batch.
 - *Carga em lote (batch):* a variável N indica o número de clientes de trabalhos ativos no sistema, possui população finita, e os clientes com serviço terminado saem do sistema e são substituídos, instantaneamente, por outro grupo de trabalhos em espera.
 - *Carga de terminal:* N indica o número de clientes ativos e Z indica o tempo de pensar.

Modelos de filas fechados usam cargas fechadas. Filas mistas usam em conjunto cargas fechadas e cargas abertas.

Tipos de Número de Cargas

Classificação de Redes de Filas

Tipos de Número de Cargas



Uma rede de filas pode ter um só tipo de requisição (ou clientes) ou mesmo diversos tipos de requisições ao mesmo tempo. As cargas podem ser:

- **Carga única:** As requisições que trafegam na rede de filas são de um mesmo tipo. Todas as requisições possuem as mesmas características de taxa de chegadas, tempo de serviço, prioridade etc. Por exemplo, um sistema de processamento de consultas que aceita apenas um tipo de consulta SQL; ou uma rede de comunicação que aceita apenas pacotes IP de serviço de “melhor esforço”.
- **Múltiplas cargas:** O sistema de filas pode aceitar diversos tipos de cargas ao mesmo tempo. Por exemplo, o sistema de bancos de dados de uma empresa aceita consultas de cadastros, inserção de registros etc. Cada tipo de carga pode ter um tempo diferente para processamento, prioridade, taxa de chegadas de usuários etc. Um exemplo de múltiplas cargas é o caso de uma rede de filas mista, onde a carga é do tipo aberta e outra do tipo fechada.

Variáveis Operacionais



Variáveis Operacionais

As variáveis operacionais utilizadas nos estudos de redes de filas devem incluir os parâmetros de qual servidor elas se referem, já que uma rede de filas terá mais de um servidor.

São definidas as seguintes variáveis operacionais para o sistemas de redes de filas:

- A_i : número de chegadas de transações ao servidor i .
- C_{ij} : número de transações que saem do servidor i e vão para o servidor j .
- C_{io} : número de transações que saem do servidor i e saem do sistema.
- C_o : número de transações que deixam o sistema.
- V_i : taxa relativa de visitas ao servidor i . Equivale ao número médio de visitas ao dispositivo por requisição:

$$V_i = \frac{C_i}{C_o}$$

Variáveis Operacionais



Continuação:

- D_i : demanda de serviço no servidor i . Diferentemente do tempo de serviço, que representa o tempo de serviço de processamento de uma requisição, a demanda representa o tempo de serviço total para todas as visitas da mesma requisição ao servidor i .

$$D_i = V_i S_i$$

- X ou X_0 : taxa de processamento ou vazão do sistema.
- B_i : tempo ocupado do servidor i .
- S_i : tempo de serviço do servidor i .
- R_i : tempo de resposta do servidor i .

Leis Operacionais para Redes de Filas



Leis Operacionais para Redes de Filas

As leis operacionais para uso de sistemas de redes em filas são:

- Lei do fluxo forçado.
- Lei do tempo de resposta.
- Lei da demanda de serviço.

Lei do Fluxo Forçado

Leis Operacionais para Redes de Filas

Lei do Fluxo Forçado



Esta lei mostra que o fluxo (vazão) em todas as partes do sistema deve ser proporcional.

$$X_i = \frac{C_i}{T} = \frac{C_i}{T} \times \frac{C_o}{C_o} = \frac{C_i \times C_o}{T \times C_o} = \frac{C_i}{C_o} \times \frac{C_o}{T} = v_i X_o$$

Leis Operacionais para Redes de Filas

Lei do Fluxo Forçado



Durante um período de 10 segundos, 40 requisições foram atendidas por um servidor de arquivos. Cada requisição requer dois acessos ao disco. O tempo médio de serviço no disco é de 30 milissegundos. Qual é a utilização média do disco nesse período?

Valores fornecidos no problema:

$$T_{\text{observado}} = 10 \text{ segundos}$$

$$C_i = 40 \text{ requisições}$$

$$V_i = 2$$

$$S_i = 30 \text{ ms} = 0,03 \text{ segundo}$$

Obtém-se

$$X_o = \frac{C_i}{T} = \frac{40}{10} = 4 \text{ requisições/segundo}$$

Como é desejado saber a utilização do disco, mas os dados fornecidos oferecem dados sobre a vazão do sistema pela Lei do Fluxo Forçado é possível obter a vazão do disco, e com ela, obter a utilização:

$$X_i = X_o V_i = 4 \times 2 = 8 \text{ req/s}$$

$$U_i = X_i S_i = 8 \times 0,03 = 0,24 = 24\%$$

A utilização do disco é de 24%.

Leis Operacionais para Redes de Filas

Lei do Fluxo Forçado



Um sistema distribuído tem um servidor de impressão com velocidade de 60 páginas por segundo. Durante um período de tempo observado de 20 minutos, observou-se que o servidor imprimiu 500 páginas. Sabendo que toda requisição que chega ao sistema para impressão deve gerar cinco páginas impressas, qual foi o índice de tarefas completadas no sistema?

Valores fornecidos no problema:

$$T_{\text{observado}} = 10 \text{ minutos} = 20 \times 60 = 1200 \text{ segundos}$$

$$C_i = 500 \text{ páginas}$$

$$V_i = 5$$

$$S_i = \frac{1}{\mu} = \frac{1}{60} = 0,017 \text{ segundo}$$

Obtém-se

$$X_i = \frac{C_i}{T} = \frac{500}{1200} = 0,42 \text{ impressão/segundo}$$

Como o problema oferece dados de um item do sistema (o servidor), mas espera-se obter os valores de vazão do sistema (composto de outros itens fora o servidor), pode-se usar a Lei do Fluxo Forçado para esse cálculo.

$$X_i = X_o V_i \Rightarrow X_o = \frac{X_i}{V_i} = \frac{0,42}{5} = 0,084 \text{ requisição/segundo}$$

Lei do Tempo de Resposta

Leis Operacionais para Redes de Filas

Lei do Tempo de Resposta



O tempo de resposta total do sistema (o tempo de processamento de uma requisição, desde a entrada no sistema de filas até sua saída) depende do tempo de resposta de cada fila versus o número de visitas que a mesma requisição faz ao servidor daquela fila. Portanto, para um sistema com K sistemas de fila:

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

Juntando a Lei do Tempo de Resposta com o Teorema do Tempo de Resposta, tem-se:

$$R = \sum_{i=1}^K \frac{D_i}{1 - U_i}$$

Leis Operacionais para Redes de Filas

Lei do Tempo de Resposta



Considere o sistema de computação da Figura 4.6, que recebe transações a $\lambda = 5$ transações/segundo.

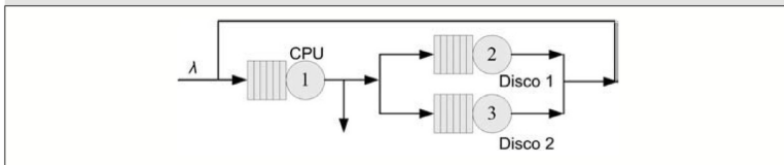


Figura 4.6 Exemplo de rede de fila.

Foram obtidos os seguintes dados:

$$S_1 = 12 \text{ ms/visita}$$

$$V_1 = 6 \text{ visitas/transação}$$

$$S_2 = 30 \text{ ms/visita}$$

$$V_2 = 1 \text{ visita/transação}$$

$$S_3 = 25 \text{ ms/visita}$$

$$V_3 = 4 \text{ visitas/transação}$$

Calcule o tempo médio de resposta do sistema.

Leis Operacionais para Redes de Filas

Lei do Tempo de Resposta



Para calcular o tempo de resposta do sistema, é necessário obter o tempo de resposta de cada elemento do sistema, como a CPU e os dois discos. Portanto, sabendo apenas os valores de λ , V_i e S_i , pode-se calcular a demanda e utilização de cada elemento, e assim, calcular o tempo de resposta do sistema, através da segunda opção da Lei do Tempo de Resposta.

$$D_1 = V_1 S_1 = 6 \times 0,012 = 0,072 \text{ segundo}$$

$$D_2 = V_2 S_2 = 1 \times 0,03 = 0,03 \text{ segundo}$$

$$D_3 = V_3 S_3 = 4 \times 0,025 = 0,1 \text{ segundo}$$

Para calcular a utilização de cada elemento, usa-se a Lei do Fluxo Forçado e assume-se que o tempo de serviço de cada disco, com múltiplas visitas, é sua demanda:

$$U_1 = X \cdot D_1 = 5 \cdot 0,072 = 0,36$$

$$U_2 = X \cdot D_2 = 5 \cdot 0,03 = 0,15$$

$$U_3 = X \cdot D_3 = 5 \cdot 0,1 = 0,5$$

$$R = R_1 + R_2 + R_3 = \frac{D_1}{1 - U_1} + \frac{D_2}{1 - U_2} + \frac{D_3}{1 - U_3}$$

$$R = \frac{0,072}{1 - 0,36} + \frac{0,03}{1 - 0,15} + \frac{0,1}{0,5} = 0,11 + 0,035 + 0,2 = 0,345 \text{ s}$$

O tempo médio de resposta do sistema é de 0,345 segundo.

Lei de Demanda de Serviço

Leis Operacionais para Redes de Filas

Lei de Demanda de Serviço



A demanda de serviço, anteriormente definida como $D_i = V_i D_i$, pode ser relacionada a vazão do sistema e sua utilização. Combinando as Leis do Fluxo Forçado e da Utilização, tem-se que:

$$D_i = V_i S_i = \frac{X_i}{X_o} \times \frac{U_i}{X_i} = \frac{U_i}{x_i}$$

Leis Operacionais para Redes de Filas

Lei de Demanda de Serviço



Um sistema de computação foi monitorado durante 60 minutos. Nesse período, 120 transações foram processadas, e foram obtidos valores para a utilização da CPU e dos dispositivos de E/S. Calcule a demanda de serviço para cada dispositivo.

Dispositivo	U_i
CPU	80%
Disco 1	20%
Disco 2	25%
Disco 3	30%

Valores fornecidos no problema:

$$T_{\text{observado}} = 60 \text{ minutos} = 3600 \text{ segundos}$$

$$C_o = 120 \text{ transações}$$

Obtém-se:

$$X_o = \frac{C_o}{T} = \frac{120}{3600} = 0,033 \text{ transação/segundo}$$

Para calcular a demanda de cada dispositivo, usa-se a Lei da Demanda de Serviço:

$$D_{\text{CPU}} = \frac{U_{\text{CPU}}}{X_o} = \frac{0,8}{0,033} = 24,24 \text{ s}$$

$$D_1 = \frac{U_1}{X_o} = \frac{0,2}{0,033} = 24,24 \text{ s}$$

$$D_2 = \frac{U_2}{X_o} = \frac{0,25}{0,033} = 7,58 \text{ s}$$

$$D_3 = \frac{U_3}{X_o} = \frac{0,3}{0,033} = 9,1 \text{ s}$$

Leis Operacionais para Redes de Filas

Lei de Demanda de Serviço



Um sistema de computação possui dois tipos de carga, uma interativa e uma em lote. O sistema tem 20 terminais ativos e tempo de pensar de 15 segundos. A taxa de chegadas de requisições da carga interativa é de 2,5 transações por segundo. Em determinado disco i do sistema, a utilização total foi de 90% e o consumo de recursos foi de 0,4 segundo para a carga interativa e 0,2 segundo para a carga em lote. Qual o tempo médio de resposta da carga interativa?

Valores fornecidos no problema:

$N = 20$ terminais

$Z = 15$ segundos

$\lambda_{\text{interativo}} = 2,5$ transações/segundo

$U_i = 0,9$

$D_{i, \text{interativa}} = 0,4$ segundo

$D_{i, \text{lote}} = 0,2$ segundo

Para o cálculo do tempo de resposta da carga interativa, deve-se usar a Lei do Tempo de Resposta Interativo (Capítulo 3):

$$R = \frac{N}{X_e} - Z$$

Leis Operacionais para Redes de Filas

Lei de Demanda de Serviço



Neste problema tem-se os valores de N e Z , e precisa-se de X_o . Temos o valor da utilização do disco i e os valores de demanda de cada carga nesse dispositivo.

Portanto:

$$D_i = \frac{U_i}{X_o}$$

$$\text{Como } U_{i, \text{ interativa}} + U_{i, \text{ lote}} = U_i$$

$$D_{i, \text{ interativa}} X_{o, \text{ interativa}} + D_{i, \text{ lote}} X_{o, \text{ lote}} = U_i$$

Pela Lei do Fluxo Forçado, tem-se

$$0,4 \times X_{o, \text{ interativa}} + 0,2 \times 2,5 = 0,9$$

$$X_{o, \text{ interativa}} = 1 \text{ transação/segundo}$$

Portanto:

$$R = \frac{20}{1} - 15 = 5 \text{ s}$$

O tempo de resposta do sistema é de 5 segundos.

Análise Operacional de Redes de Fila

Prof. Fabrício B. Gonçalves

Instituto Federal Fluminense
Campus Bom Jesus do Itabapoana