

Aula 1: Classificação da Redes de Filas

Instrutor: Berilhes Borges Garcia

Escreva: Thales Dardengo de Paiva

1 Classificação da Redes de Filas

- Redes Abertas
- Redes Fechadas

1.1 Redes Abertas

Uma rede de filas aberta possui chegada e saída do exterior.

Example: Sistema Servidor-Único.

⊗



Figure 1: Servidor Único

Example: Redes de Filas com Roteamento Probabilístico.

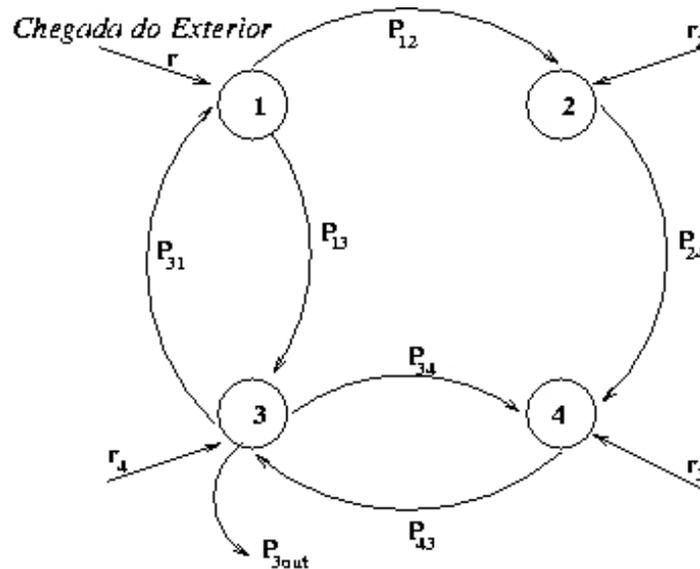


Figure 2: Roteamento Probabilístico entre os servidores

Um pacote que finalizou o serviço no servidor i é a seguir roteado para o servidor j com probabilidade P_{ij} .

Nós podemos permitir que as probabilidades dependam da “classe” do pacote, de modo que nem todos os pacotes tenham de seguir o mesmo esquema de roteamento. ☒

Aplicação:

Example: Redes de Filas com nenhum Roteamento Probabilístico.

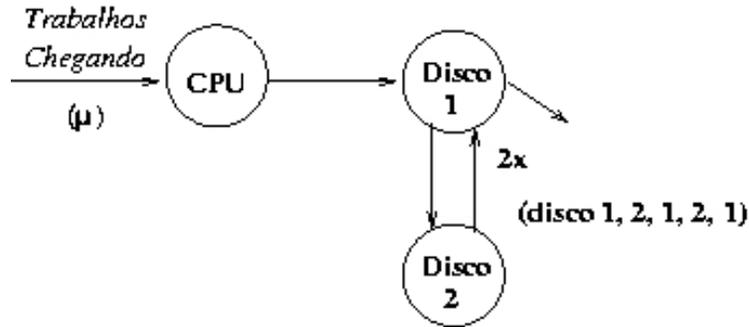


Figure 3: Rede sem Roteamento Probabilístico.

☒

Mais duas Métricas de Desempenho: Throughput e Utilização

- **Throughput:** (É a métrica de desempenho mais utilizada no dia a dia). Todo mundo quer maior throughput! Porque?

Pergunta: Alguém pode me dizer como maximizar o throughput está relacionado com minimizar o tempo de resposta? Por exemplo, qual dos seguintes sistemas tem maior throughput?

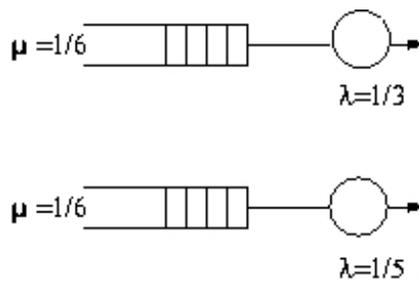


Figure 4: Filas com mesma taxas de médias de chegada, λ, e razões médias de serviço, μ, diferentes.

Resposta: Cenas dos próximos capítulos!

- A **Utilização do Dispositivo i** (U_i) é a fração do tempo que o dispositivo i está ocupado.

$$U_i = \frac{B}{T} \tag{1}$$

O Troughput de i (X_i) é a razão de conclusões no dispositivo i (jobs/seg).

$$X_i = \frac{C}{T} \quad (2)$$

Como X_i está relacionado à U_i ?

Bem,

$$\frac{C}{T} = \frac{C}{B} \cdot \frac{B}{T} \quad (3)$$

Pergunta: Oque é $\frac{C}{B}$?

Resposta:

$$\frac{B}{C} = E\{S\} \quad (4)$$

Logo:

$$\frac{C}{B} = \frac{1}{E\{S\}} = \mu \quad (5)$$

Assim nós temos:

$$X_i = \mu_i \cdot U_i \quad (6)$$

ou, de forma equivalente,

$$U_i = \frac{X_i}{E\{S\}} \quad (7)$$

Esta última formação é conhecida como: **“Lei da Utilização”**.

Example: Rede Servidor-Único. Qual é o throughput?

Pergunta: O que é X ? (Qual o valor de X ?)

Resposta: $X = U \cdot \mu$. Mas qual é o valor de U ? Nós provaremos dentro de breve que $U = \frac{\lambda}{\mu}$. Aqui está uma demonstração intuitiva.

U = Fração do tempo que o servidor está ocupado

U = (Tempo de serviço médio demandado por um job) / (Tempo médio entre chegadas)

$$U = \frac{\frac{1}{\mu}}{\frac{1}{\lambda}} \quad (8)$$

$$U = \frac{\lambda}{\mu} \quad (9)$$

Assim nós temos:

$$X = U \cdot \mu = \frac{\lambda}{\mu} \cdot \mu = \lambda \quad (10)$$

⊠

Dessa forma o throughput não depende da razão de serviço!!!

Ambos tem o mesmo throughput.

Logo no caso do processador mais rápido, o tempo de resposta cai, o tamanho do fila diminui, mas X não muda. Portanto, tempo de resposta menor não está relacionado com uma razão maior.

Pergunta: Explique porque X não muda.

Resposta: Não importa quão alta seja μ , a razão de conclusão é ainda limitada pela razão de chegada. Mudando μ nós alteramos o X máximo possível, mas não o X real. Note que nós estamos assumindo um sistema estável.

Example: Redes de Filas Probabilísticas.

Pergunta: O que é X? Resposta:

$$X = \sum x_i \quad (11)$$

Pergunta: O que é X_i ? Resposta: λ_i denota a razão de chegada total no servidor i . Então:

$$X_i = \lambda_i = x_i + \sum_j \lambda_j P_{ij} \quad (12)$$

Nós temos que resolver estas equações simultaneamente.

Pergunta: Como são os r_i 's restritos?

Resposta: Para a rede alcançar o "equilíbrio", esta deve ter

$$\lambda_i < \mu_i, \forall_i \quad (13)$$

e isto restringe os r_i 's.

⊠

1.2 Redes Fechadas

Redes de Filas fechadas não tem nenhuma chegada ou partida do exterior. Elas podem ser classificadas em duas categorias:

- Interativas
- Batch

- Interativas - Sistemas

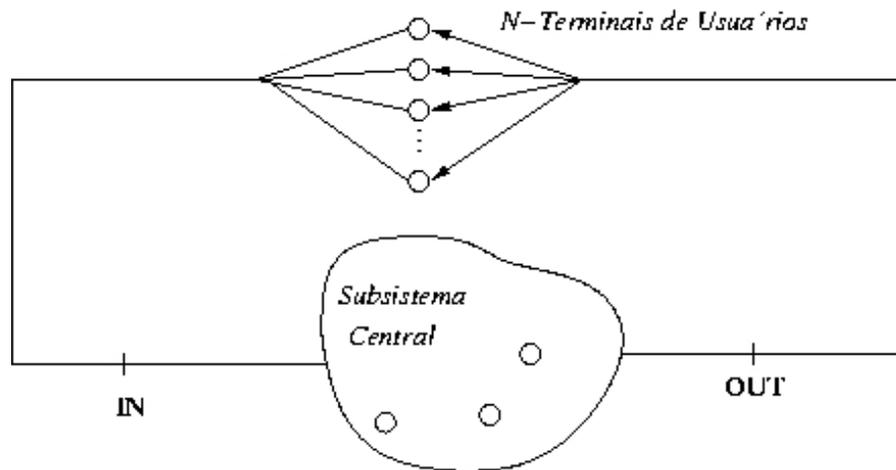


Figure 5: Exemplo de uma Rede Fechada: Sistema Interativo

- Terminais representam usuários, os quais enviam um trabalho para o sub-sistema central e então esperão por uma resposta.

- Usuários não podem enviar o próximo trabalho antes que o trabalho anterior tenha retornado.

- Dessa forma o número de trabalhos é fixo (igual ao número de terminais). Este número é algumas vezes chamado a **carga** ou **NMP** (nível de multiprogramação), não confunda com a utilização do dispositivo.

- Há um tempo de pensamento, Z , o qual é uma variável aleatória representando o tempo, em cada terminal, entre o recebimento de um resultado para um trabalho e o envio do próximo trabalho. Dessa forma, o número de trabalhos no sub-sistema central é no máximo o número de terminais.