Teoria de Filas - Resumo e exercícios

Pedroso

7 de março de 2013

1 Introdução

A Teoria de filas é uma das abordagens mais utilizadas no estudo de desempenho e dimensionamento de sistemas de comunicação de dados. Muita atenção deve ser dada aos processos de chegada e atendimento. Boas referências podem ser encontradas em http://web2.uwindsor.ca/math/hlynka/qonline.html.

2 Definições básicas

Considere a Figura 1. Elementos chegam a uma fila com uma taxa (ou ritmo) de chegadas dada por λ. Os elementos são atendidos por M servidores com uma taxa de atendimento dada por μ.

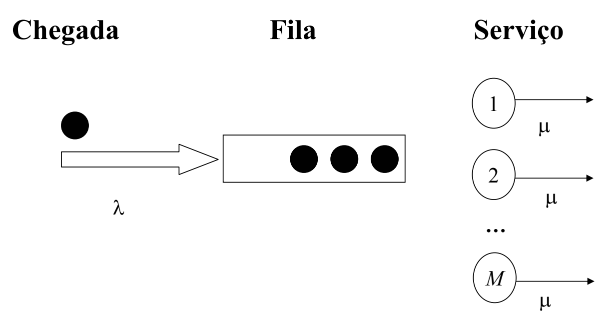

Figura 1: Seqüência de chamadas do sistema RPC

Podem-se definir as seguintes variáveis:

TA Tempo médio de atendimento, dado por $TA = 1/\mu$;

IC Intervalo médio entre chegadas, dado por $IC = 1/\lambda$;

TF Tempo médio gasto na fila;

TS Tempo médio gasto no sistema, dado por $TS = TF + TA$;

NA Número médio de elementos sendo atendidos;

NF Número médio de elementos na fila;

NS Número médio de elementos no sistema, dado por $NS = NF + NA$;

2.1 Leis de Little

As leis de Little aplicam-se a todos os sistemas de fila, independentemente do processo de chegada e atendimento:

$$NF = TF.\lambda \tag{1}$$

$$NS = TS.\lambda \tag{2}$$

2.2 Ocupação do Sistema

Define-se como ocupação (ou utilização) do sistema:

$$\rho = \frac{\lambda}{M.\mu} \tag{3}$$

Sendo que para que o sistema seja estacionário $\rho < 1$. Esta relação é fundamental para o estudo de sistemas de filas. Caso $\rho \geq 1$ a fila aumenta indefinidamente.

A ocupação do sistema não possui unidade. Significa a parcela do tempo em que os servidores estão atendendo requisições. Logo, o tempo livre dos servidores pode ser dado por $1 - \rho$.

Exercício 1: Considere um sistema onde chegam requisições a um servidor. Abaixo estão os valores para os intervalos entre chegadas e tempo de atendimento para cada requisição:

Intervalo entre Chegadas: 2,3 ; 0,5; 2,9; 2,2; 2,0; 2,2; 2,5; 2,6; 4,8; 1,3; 2,5; 1,4; 1,2; 2,1; 0,5.

Tempo de Atendimento: 4,1; 1,7; 6,9; 3,8; 1,8; 1,7; 1,3; 0,4; 3,4; 1,6; 0,8; 0,5; 3,5; 2,8; 0,1.

Determine:

a. O ritmo médio de chegadas.

b. O ritmo médio de atendimentos.

c. O tempo médio de espera na fila neste sistema (apenas observe a dinâmica da fila, não é necessário assumir nenhum modelo).

d. A ocupação do sistema.

e. O número médio de elementos na fila e no sistema.

□

Exercício 2: Suponha uma operadora de cartões de crédito. Muitos comerciantes ainda utilizam a modalidade de conexões discadas para acessar o sistema. Sabendo-se que o tempo médio de atendimento é de 20 segundos (tempo para enviar os dados necessários para processar a transação) e que o sistema tem uma taxa de chegada de 1000 conexões por minuto, determine a quantidade *mínima* de modems. □

3 Processos de Chegada e Atendimento

A identificação dos processos de chegada e atendimento permite a utilização de resultados bem estabelecidos para determinação dos valores de TF (e conseqüentemente de TS, NF e NS) para sistemas de filas. A identificação de tais processos utiliza os clássicos testes de aderência, estudados em probabilidade.

4 Notação de Kendall

É uma notação padrão para classificar sistemas de filas de acordo com as diferentes configurações possíveis.

```
A/B/C/K/P/Z
```

A	Distribuição do intervalo entre chegadas
B	Distribuição do tempo de serviço
C	Número de servidores
K	Número máximo de clientes no sistema (valor default ∞)
P	Tamanho da população (valor default ∞)
Z	Disciplina da fila (valor default FIFO)

As variáveis A e B podem assumir os seguintes valores:

M Distribuição exponencial (Markoviano);

D Determinístico;

E_k Distribuição de Erlang (k = shape parameter);

G Geral (qualquer distribuição)

Exemplo 3: $D/M/n$ descreve uma fila com o intervalo entre chegadas determinístico (sempre o mesmo intervalo), tempo de atendimento exponencial e n servidores. $\square$

5 Resultados Clássicos

5.1 Modelo M/M/1

$$NF = \frac{\lambda^2}{\mu(\mu-\lambda)} \quad NS = \frac{\lambda}{\mu-\lambda}$$
$$TF = \frac{\lambda}{\mu(\mu-\lambda)} \quad TS = \frac{1}{\mu-\lambda}$$

Probabilidade de existirem n clientes no sistema: $P_n = \big(1 - \frac{\lambda}{\mu}\big)\big(\frac{\lambda}{\mu}\big)^n$

5.2 Modelo M/M/c

$$TF = TA.\frac{1-A}{(1-\rho)(1-\rho A)}, \quad A = \frac{\displaystyle\sum_{i=0}^{M-1} \frac{(M\rho)^i}{i!}}{\displaystyle\sum_{i=0}^{M} \frac{(M\rho)^i}{i!}}$$

5.3 Modelo M/M/c com Perda de Chamada

Neste modelo o tamanho da fila é igual a zero. Caso todos os servidores estejam ocupados, as requisições são perdidas. A probabilidade de obter-se o sistema ocupado neste caso é dada por

$$\Pr[Ocupado] = \frac{\dfrac{(M\rho)^M}{M!}}{\displaystyle\sum_{i=0}^{M} \frac{(M\rho)^i}{i!}}$$

Em anexo pode ser encontrado um gráfico com a probabilidade de perda em função de ρ.

5.4 Modelo M/G/1

A equação que fornece o número médio de tarefas no sistema M/G/1, conhecida por equação de Pollaczek-Khinchin, é dada pela Equação 4,

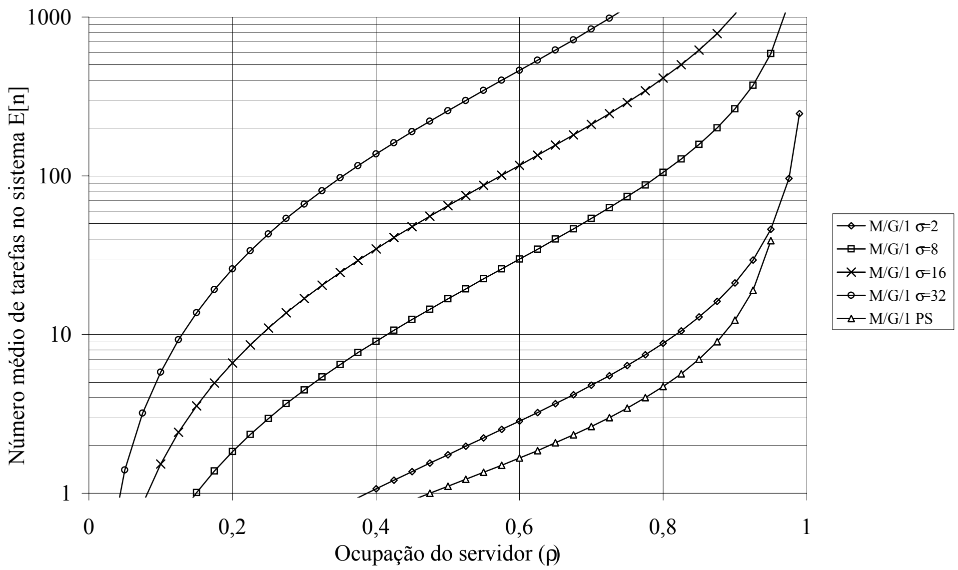

Figura 2: Número de tarefas no sistema com desvio padrão de $\sigma = 0, 2, 8, 16, 32$ no sistema *M/G/1* e no sistema *M/G/1 PS*

$$E[n] = \frac{\rho^2}{2.(1-\rho)}.\left[1 + \frac{\sigma^2}{TA^2}\right] + \rho, \quad \rho < 1 \tag{4}$$

onde ρ é a ocupação do sistema, dado pela razão entre a taxa de chegada λ e a taxa de atendimento μ, $\rho = \lambda/\mu$. O desvio padrão do tempo de serviço é representado por σ e TA indica o tempo médio de serviço. O tempo total de atendimento pode ser calculado utilizando-se as leis operacionais de Little [**?**].

Segundo [**?**], a expressão que calcula o número de elementos no sistema *M/G/1 PS* é dada por $E[n] = \rho/(1-\rho)$. O tempo médio de resposta pode ser obtido com as leis de Little.

Demais modelos podem ser encontrados em [**?**].

6 Exercícios Selecionados

Exercício 3: Foi observado o comportamento de um certo servidor de banco de dados durante um período de tempo onde o sistema era estacionário.

Durante o período de 1 minuto de observação, o tempo livre do sistema (idle time) foi de 10 segundos. A taxa média de chegada foi de 5 requisições por segundo.

Utilizando o modelo M/M/1, determine:

1. A utilização do sistema;
2. O número de médio requisições sendo processadas ;
3. O tempo médio de resposta;
4. A probabilidade do número de tarefas no sistema ser maior que 10.

Já é conhecido o fato de que os tempos de resposta em servidores seguem distribuições de cauda pesada. Recalcule os ítens acima considerando o sistema como uma fila M/G/1, aplicando valores crescentes para o desvio padrão do tempo de serviço. O que ocorre com o tempo médio de resposta a medida que a variação do tempo de serviço aumenta? □

Exercício 4: Um sistema com uma base de dados consiste de 3 discos rígidos compartilhando uma fila única. O tempo de serviço para uma requisição de E/S é de 50m segundos. As requisições de E/S chegam ao sistema a uma taxa de 30 requisições por segundo. Utilizando o modelo M/M/3, determine o seguinte:

1. A utilização média dos discos rígidos;
2. A probabilidade do sistema esta ocioso;
3. O número médio de requisições de acesso no sistema e o número médio de requisições esperando na fila;
4. O tempo médio de resposta.

□

Exercício 5: Resolva o problema anterior assumindo que cada disco rígido possua uma fila separada. □

Exercício 6: Um banco possui dois funcionários trabalhando no setor de atendimento ao público. O primeiro trabalha apenas com depósitos e o segundo, com retiradas. Sabe-se que ambos atendem uma média de 3 minutos por cliente (processo Markoviano), com um desvio padrão de 5 minutos. As chegadas são Markovianas, com média de 16 chegadas por hora para os depositantes e 14 chegadas por hora para os que vão fazer retirada. Qual seria o efeito no tempo médio no sistema (TS) se ambos os funcionários trabalhassem tanto com retiradas como com depósitos? □

Exercício 7: Suponha que um sistema computacional foi estruturado da seguinte maneira: um cluster composto de 3 servidores recebe uma requisição, processa e encaminha para um dos dois servidores de banco de dados que contem a informação desejada. O cluster foi criado porque a capacidade de processamento dos computadores que processam a requisição é pequena em relação aos servidores de banco de dados. Suponha que o tempo médio de atendimento de uma requisição por um dos computadores do cluster é de 2 segundos. O tempo médio de atendimento de uma requisição para o servidor de banco de dados é de 0,5 segundo. O sistema recebe 3600 requisições por hora. Considere que a distribuição de probabilidade que modela os tempos de atendimento é a exponencial e que o processo de chegada é um processo Markoviano. Utilizando a teoria de filas, determine:

1. Ocupação de cada servidor;
2. Tamanho médio da fila em cada um dos servidores e total;
3. Tempo médio de resposta em cada um dos servidores e total.

□

Exercício 8: Um provedor de acesso à internet possui 52.291 clientes. A taxa média de chegada é de 40 ligações por hora. Sabendo-se que a tempo médio de conexão é de 10 minutos, e que ambos seguem a distribuição exponencial, qual será a quantidade de modems necessários para que a probabilidade de perda de ligação seja menor que 3%? □

A Sistema M/M/c com Perda de Chamada