

---

# Teoria de Filas - Resumo e Exercícios

---

Pedroso

24 de setembro de 2014

## 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  $\lambda$ . Os elementos são atendidos por  $M$  servidores com uma taxa de atendimento dada por  $\mu$ .

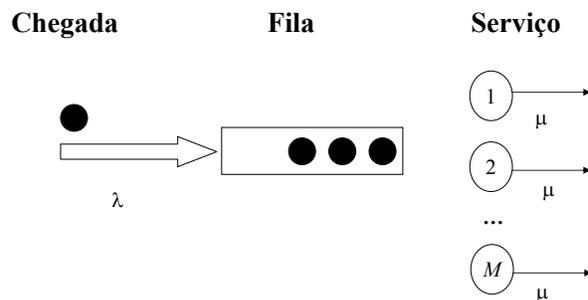


Figura 1: Seqüência de chamadas do sistema RPC

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:

$$NF = TF \cdot \lambda \quad (1)$$

$$NS = TS \cdot \lambda \quad (2)$$

## 2.2 Ocupação do Sistema

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

$$\rho = \frac{\lambda}{M \cdot \mu} \quad (3)$$

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  $1 - \rho$ .

## 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  $\infty$ )
- P Tamanho da população (valor default  $\infty$ )
- 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 1:*  $D/M/n$  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

$$\begin{aligned} NF &= \frac{\lambda^2}{\mu(\mu - \lambda)} & NS &= \frac{\lambda}{\mu - \lambda} \\ TF &= \frac{\lambda}{\mu(\mu - \lambda)} & TS &= \frac{1}{\mu - \lambda} \end{aligned}$$

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

### 5.2 Modelo M/M/c

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

### 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

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

Em anexo pode ser encontrado um gráfico com a probabilidade de perda em função de  $\rho$ .

### 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,

$$\bar{N}S = \frac{\rho^2}{2 \cdot (1 - \rho)} \cdot \left[ 1 + \frac{\sigma^2}{TA^2} \right] + \rho, \quad \rho < 1 \quad (4)$$

onde  $\rho$  é a ocupação do sistema, dado pela razão entre a taxa de chegada  $\lambda$  e a taxa de atendimento  $\mu$ ,  $\rho = \lambda/\mu$ . O desvio padrão do tempo de serviço é representado por  $\sigma$  e  $TA$  indica o tempo médio de serviço. O tempo total de atendimento pode ser calculado utilizando-se as leis operacionais de Little [Jain 1991].

Segundo [Jain 1991], 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 [Jain 1991].

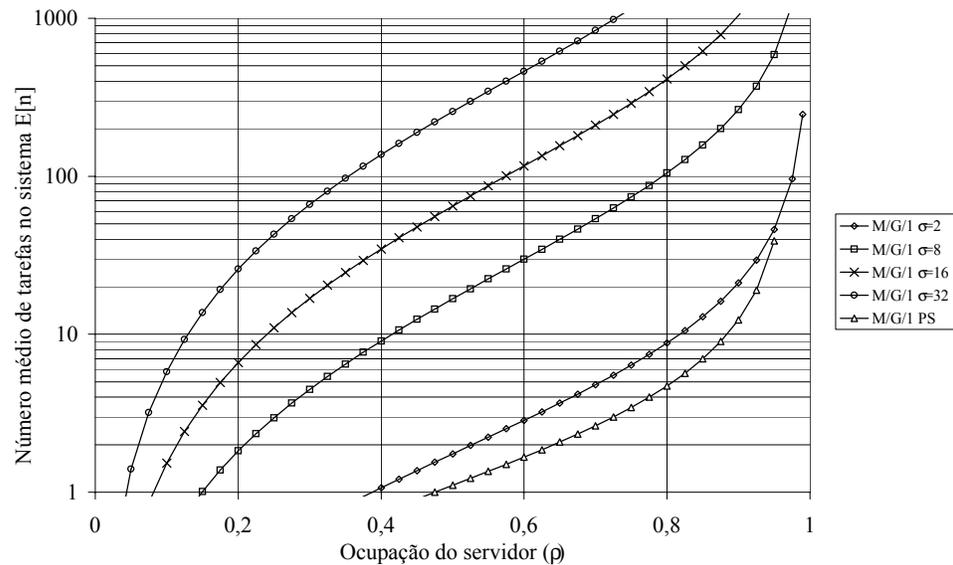


Figura 2: Número de tarefas no sistema com desvio padrão de  $\sigma = 0, 2, 8, 16, 32$  no sistema  $M/G/1$  e no sistema  $M/G/1 PS$

## 6 Exercícios

*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:

2,1	1,5	3,4	8,0	1,0	0,2	2,5	2,6	0,8	3,3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Tempo de Atendimento:

4,1	3,5	1,4	0,1	1,0	0,2	0,5	5,6	0,2	0,3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

*Exercício 2:* Suponha um sistema VoIP que utiliza o padrão G.711. Neste padrão, a voz é codificada em pacotes de tamanho fixo, tipicamente com 160 bytes, transmitidos a intervalos fixos de tempo, tipicamente de 20 milissegundos.

1. Qual seria o modelo de filas a ser utilizado para o dimensionamento de um canal para este sistema?
2. Suponha que seja alterado o padrão de codificação, realizando uma compactação que produz um tamanho variável de pacotes, mas ainda transmitidos a cada 20 milissegundos. Explique como determinar o modelo de filas que pode ser utilizado para estudar tal sistema?

□

1. Qual o ritmo médio de chegadas, o ritmo médio de atendimentos e a ocupação do sistema?
2. Qual o tempo médio de espera na fila neste sistema? (apenas observe a dinâmica da fila, não é necessário assumir nenhum modelo)

□

*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 sistema ficou 10% do tempo livre. A taxa média de chegada foi de 100 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 percebido pelo usuário;
4. A probabilidade do número de tarefas no sistema ser maior ou igual a 1.

Já é conhecido o fato de que os tempos de resposta em servidores seguem distribuições de cauda pesada. Recalcule os itens acima considerando o sistema como uma fila  $M/G/1$ , aplicando 2 e 10 vezes o valor do desvio padrão utilizado na hipótese do sistema ser  $M/M/1$ . . 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:* Considere um sistema com uma fila onde os clientes chegam de acordo com o processo de Poisson com uma taxa de 25 clientes por hora. Suponha que o tempo de atendimento possui distribuição exponencial. Existem duas opções possíveis para o projeto do sistema. A primeira utiliza dois servidores, cada um atendendo clientes com um tempo médio de atendimento de 4 minutos. A segunda utiliza um único servidor com um tempo médio de atendimento 2 minutos.

Calcule qual o tempo médio de espera na fila em cada um dos casos.

□

*Exercício 7:* Um provedor de acesso à internet possui 101.722 clientes. A taxa média de chegada é de 100 ligações por hora. Sabendo-se que o tempo médio de conexão é de 9 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%? □

## Referências

[Jain 1991] Jain, R. (1991). *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation and modeling*. John Wiley & Sons.

## A Sistema M/M/c com Perda de Chamada

