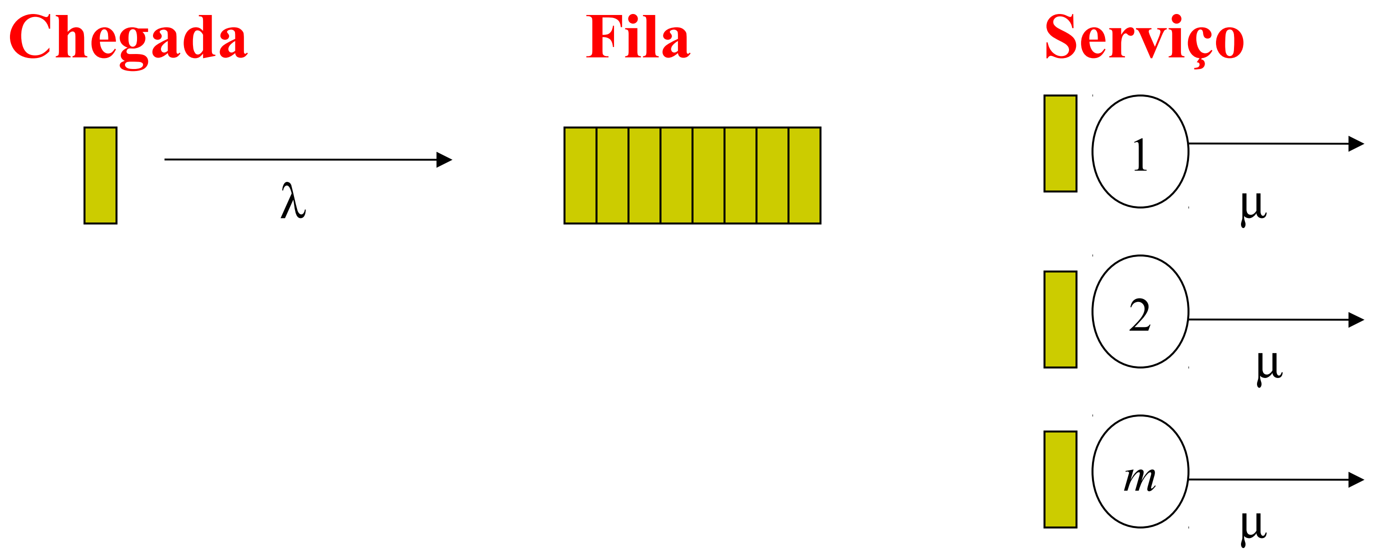
FILAS

Conceitos Fundamentais

Pós Graduação em Engenharia Elétrica - PPGEE
Prof. Carlos Marcelo Pedroso
2016

Teoria de Filas

- É uma das abordagens mais utilizadas no estudo de desempenho e dimensionamento de redes;



Teoria de Filas

- Notação básica:
  - $\lambda \Rightarrow$ taxa média de chegada
    - $\lambda=1/IC$, IC é o intervalo médio entre pacotes
  - $\mu \Rightarrow$ taxa média de atendimento
    - $\mu= 1/TA$, TA é o tempo médio de atendimento
  - TF, TS, TA: Tempo na fila, tempo no sistema e tempo de atendimento
    - TS=TS+TA

Lei de Little

Foi observado o comportamento de um certo servidor de banco de dados.

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. Considere que o tempo de resposta aparente para o usuário foi de 0,1 segundo em média.

Qual o tempo gasto na fila e no processamento da requisição? Qual o número médio de elementos na fila?

Exemplo

- Suponha um provedor de acessos a Internet que recebe uma média de 20 requisições/hora. O tempo médio de conexão é de 10 minutos. Qual a quantidade mínima de modems para atender este sistema?

Teoria de Filas - Notação

- Para o estudo de filas, convencionou-se a seguinte notação: A/S/m/B/K/SD
  - A indica a distribuição de probabilidade do intervalo entre chegadas;
  - S indica a distribuição de probabilidade do tempo de atendimento;
  - M é o número de servidores;
  - B é a capacidade do buffer;
  - K é o tamanho da população;
  - SD é a disciplina da fila (ex. Fifo).

Teoria de Filas - Notação

- As distribuições são indicadas por:
  - M – Exponencial
  - E – Erlang
  - H – Hyperexponencial
  - D – Determinística
  - G – Geral

M/M/1 com população infinita

$$NF = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

$$TF = \frac{\lambda}{\mu(\mu - \lambda)}$$

$$NS = \frac{\lambda}{\mu - \lambda}$$

$$TS = \frac{1}{\mu - \lambda}$$

$$P_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n$$

P_n *é a probabilidade de existirem n clientes no sistema*

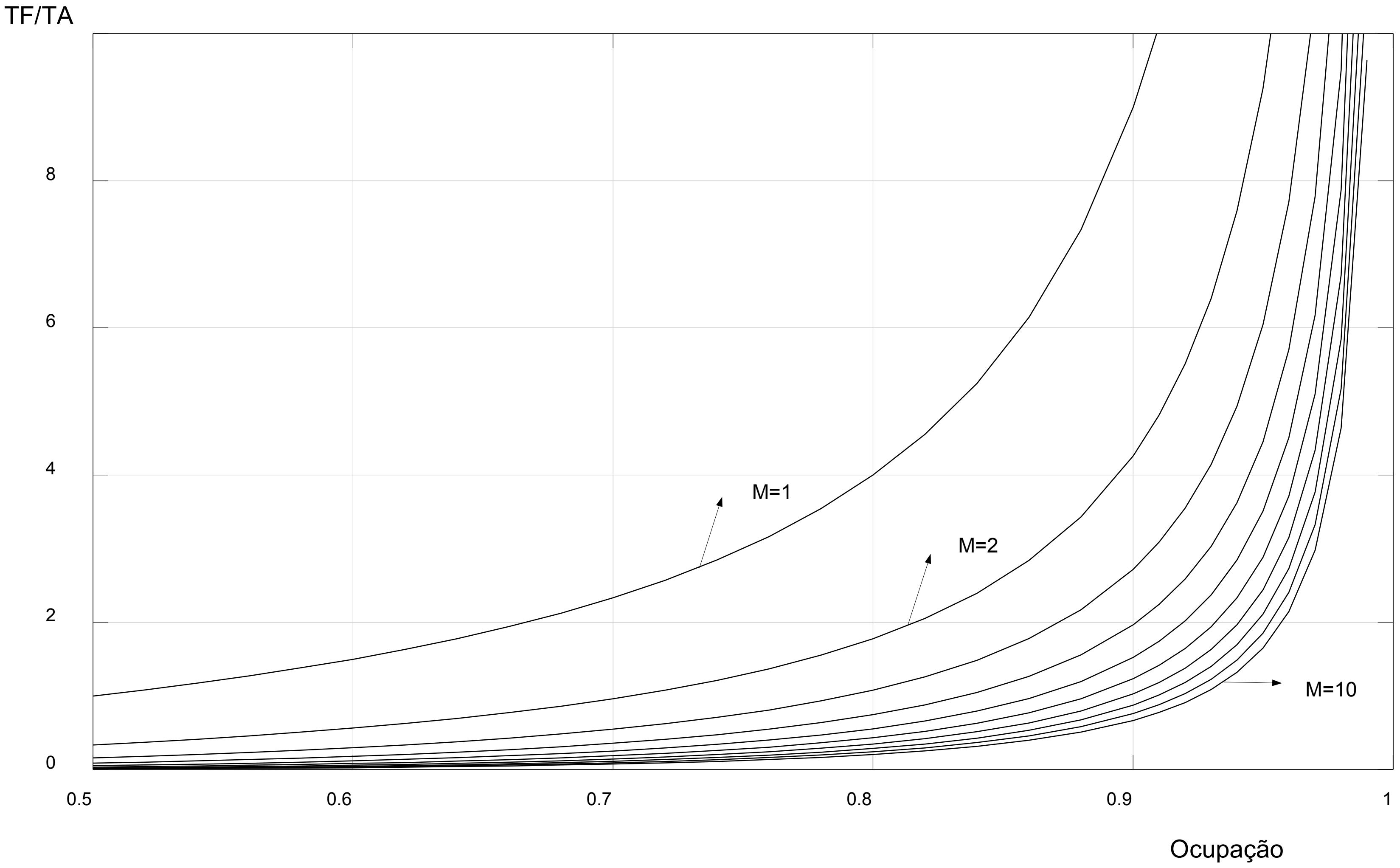
M/M/c com população infinita

- As fórmulas para o modelo M/M/c são relativamente complexas;

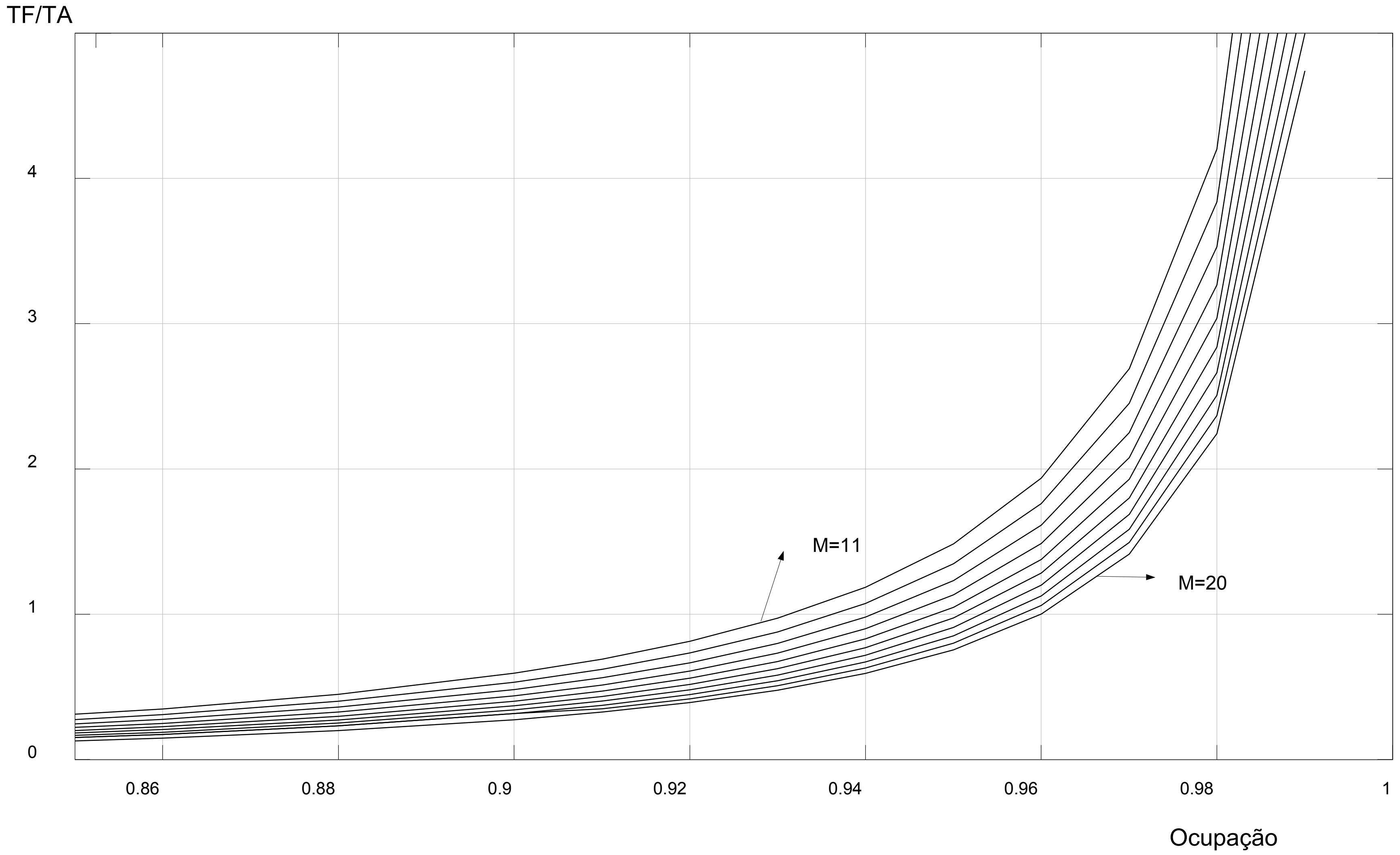
$$TF = TA.\frac{1-A}{(1-\rho)(1-\rho A)}, \quad A = \frac{\displaystyle\sum_{i=0}^{M-1} \frac{(M\rho)^i}{i!}}{\displaystyle\sum_{i=0}^{M} \frac{(M\rho)^i}{i!}}$$

- Normalmente, utiliza-se a abordagem gráfica para resolução de problemas.

M/M/c com população infinita



M/M/c com população infinita



Exemplo

Suponha um provedor de acessos a Internet que recebe uma média de 20 requisições/hora. O tempo médio de conexão é de 10 minutos. Qual a quantidade de modems necessária para oferecer uma boa qualidade aos usuários?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:

A) utilização média dos discos rígidos;

B) probabilidade do sistema esta ocioso;

C) número médio de requisições de acesso no sistema e o número médio de requisições esperando na fila;

D) O tempo médio de resposta.

Exemplo

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.

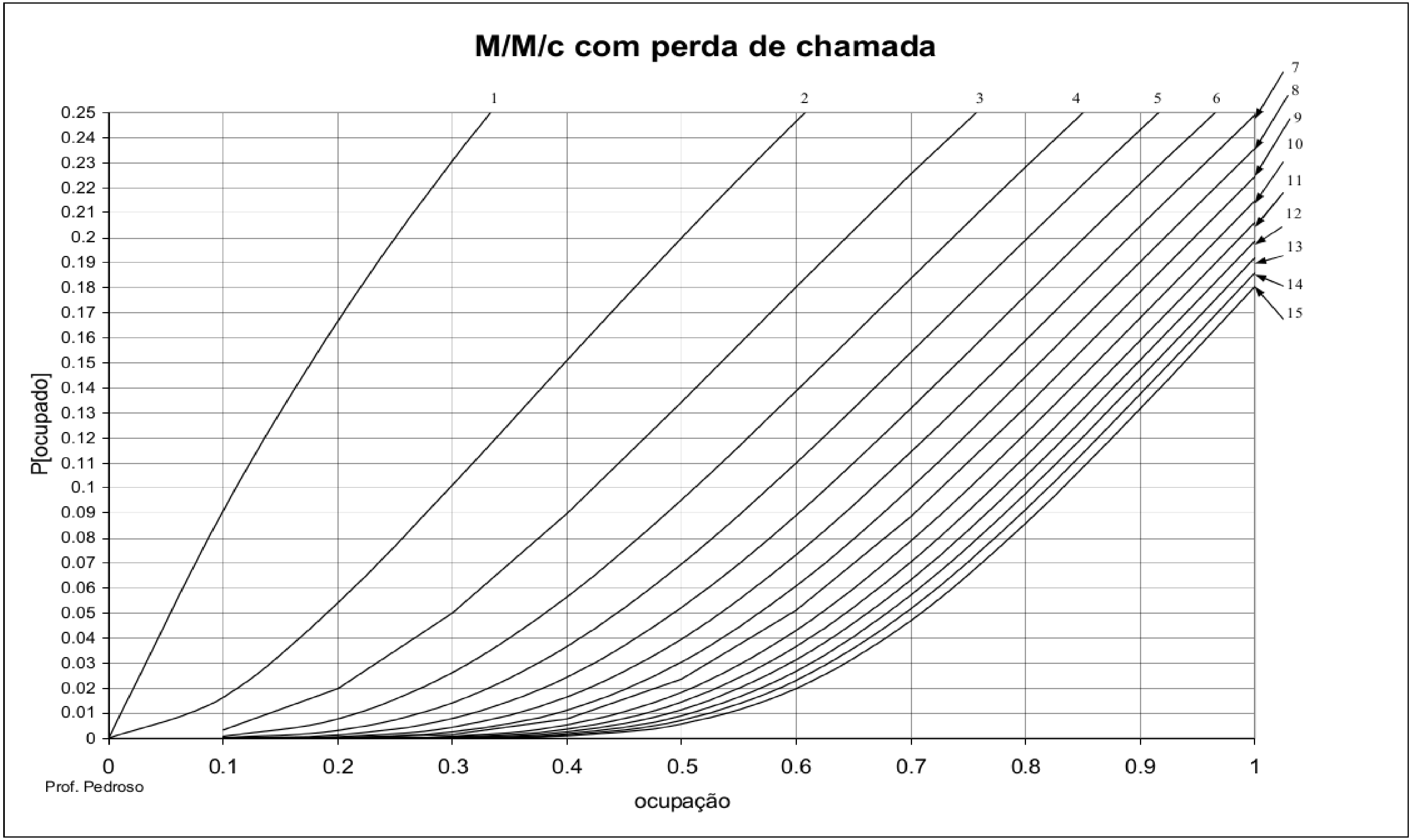
Calcule qual o tempo médio de espera na fila em cada um dos casos.

M/M/c com perda de chamada

- Para o sistema M/M/c com perda de chamada temos:

$$P[\,ocupado\,]=\frac{(M\rho)^{M}/M!}{\displaystyle\sum_{i=0}^{M}\frac{(M\rho)^{i}}{i!}}$$

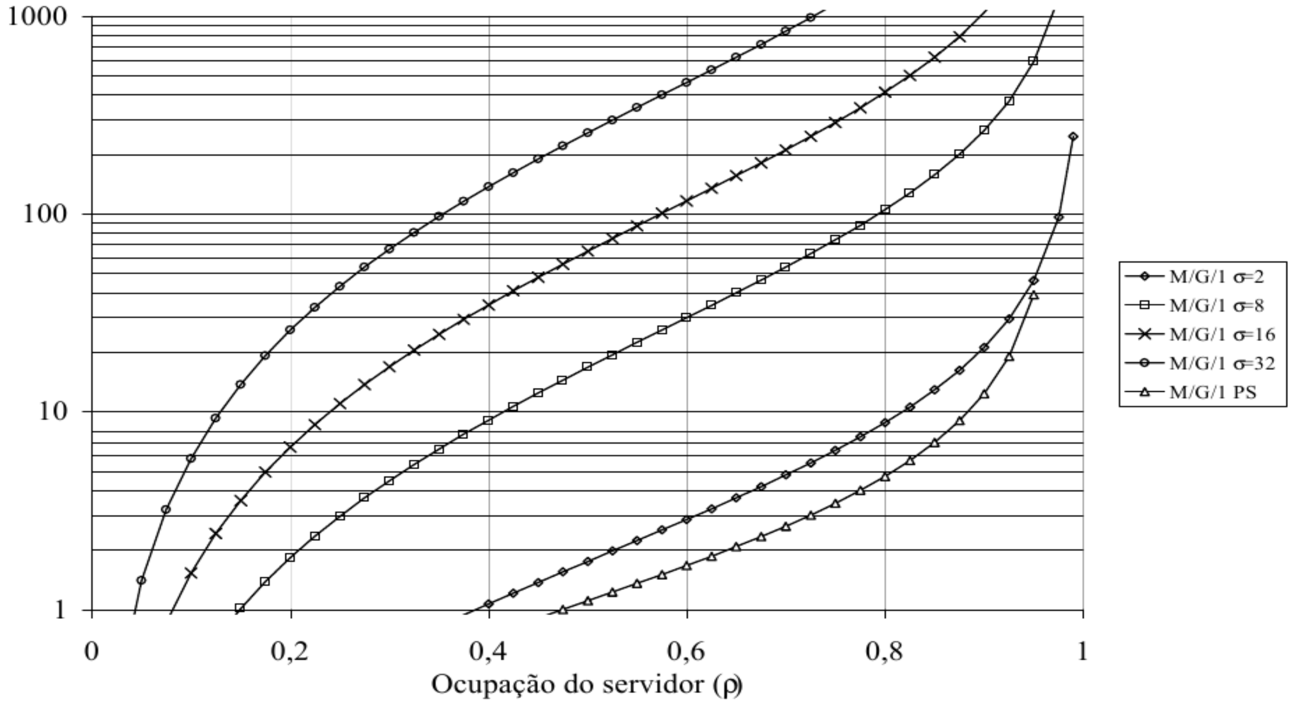
M/M/c com perda



Exemplo

- Suponha um provedor de acessos a Internet que recebe uma média de 20 requisições/hora. O tempo médio de conexão é de 10 minutos. Qual a quantidade de modems para obter um desempenho adequado.
  - Um tempo adequado é o tempo limite que um cliente espera pela conexão para fazer um pagamento com um cartão, por exemplo.

M/G/1



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

Outros modelos

A literatura apresenta uma série de modelos prontos:

- Consultar:
  - "The Art of Computer Systems Performance Analysis", Raj Jain.
  - "Discrete Event System Simulation", Jerry Banks.