

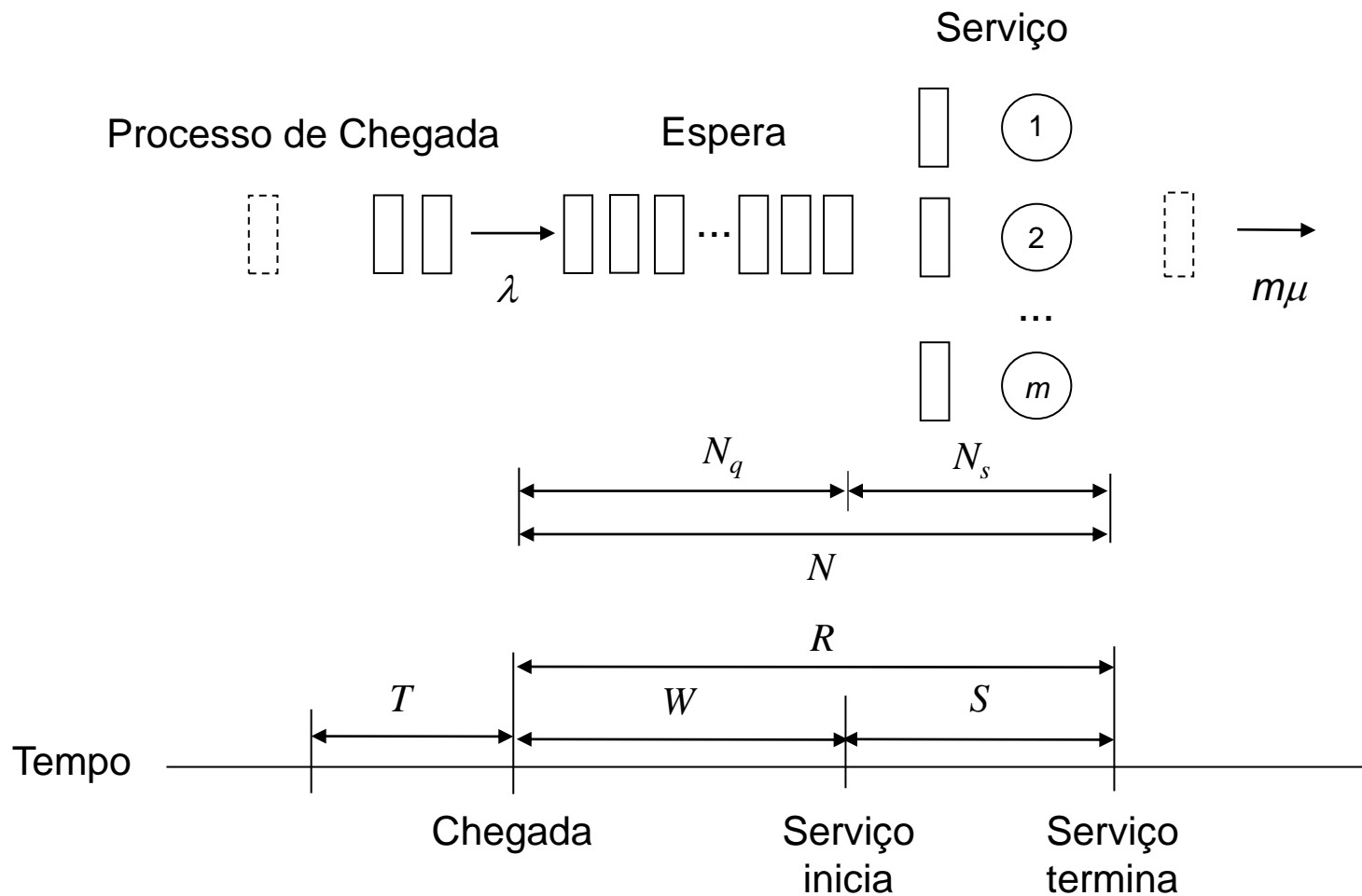
Avaliação de Desempenho

Filas MMm

Fila M/M/m

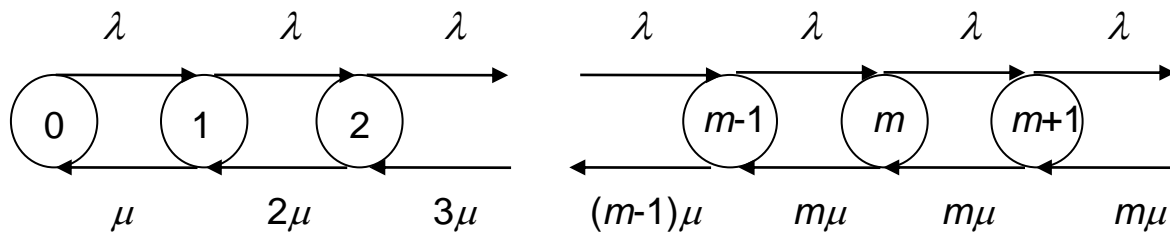
- Tempo entre chegadas é distribuído exponencialmente
- Há m servidores
- Tempo de serviço tem distribuição exponencial
- Não há limitação do buffer ou do tamanho da população
- Disciplina de serviço é FCFS

Variáveis



Número de Elementos no Sistema

- Diagrama de transição para sistema MMm



- Estado estacionário (tempo de observação muito grande)

Parâmetros da fila M/M/m

PARÂMETROS	
Taxa de chegada	λ
Tempo médio entre chegadas	$1/\lambda$
Número de servidores	m
Taxa de serviço em cada servidor	μ
Tempo médio de serviço ($E[S]$)	$1/\mu$
Intensidade de tráfego (ρ) = Utilização média de cada servidor	$\lambda/m\mu$
Condição de estabilidade	$\lambda < m\mu$ ou $\rho < 1$

Probabilidades da fila M/M/m

PROBABILIDADES	
Probabilidade de não haver nenhuma tarefa no sistema	$P[N = 0] = P_0 = \left[1 + \frac{(m\rho)^m}{m! (1 - \rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!} \right]^{-1}$
Probabilidade de haver n tarefas no sistema	$P[N = n] = \frac{(m\rho)^n}{n!} P_0, \quad n < m$ $P[N = n] = \frac{\rho^n m^m}{m!} P_0, \quad n \geq m$
Probabilidade de haver m ou mais tarefas no sistema (ε)	$\varepsilon = \frac{P_0(\rho^m m^m)}{m! (1 - \rho)}$ $\varepsilon = P[N \geq m]$

Probabilidades da fila M/M/m

PROBABILIDADES	
Função de distribuição acumulada do tempo de espera	$F_w(w) = 1 - (\varepsilon \cdot e^{-m\mu(1-\rho)w})$
Função de distribuição acumulada do tempo de resposta	$F_R(r) = 1 - e^{-\mu r} - \frac{\varepsilon}{1 - m + m\rho} e^{-m\mu(1-\rho)r} \quad \rho \neq (m-1)/m$ $F_R(r) = 1 - e^{-\mu r} - \varepsilon \mu r e^{-m\mu r} \quad \rho = (m-1)/m$

Médias da fila M/M/m

MÉDIAS	
Número médio de tarefas no sistema	$E[N] = m\rho + (\varepsilon\rho) / (1-\rho)$
Número médio de tarefas na fila	$E[N_q] = (\varepsilon\rho) / (1-\rho)$
Número médio de tarefas sendo servidas	$E[N_s] = m\rho$
Tempo de resposta médio	$E[R] = \frac{1}{\mu} + \frac{\varepsilon}{m\mu(1-\rho)}$
Tempo na fila médio médio	$E[W] = \frac{\varepsilon}{m\mu(1-\rho)}$
Tempo de serviço médio	$E[S] = \frac{1}{\mu}$

Exemplo 1

- Um servidor web recebe solicitações em uma taxa de 3000 chegadas por minuto. A taxa de serviço do servidor é 58,8235 solicitações por segundo.
- Qual é o tempo de serviço médio?
- Qual é o tempo de resposta médio?
- Qual é o tamanho médio da fila?
- Qual é a probabilidade de o servidor ficar ocioso?
- Qual é a probabilidade de haver mais do que 3 tarefas no sistema?
- Qual é a probabilidade de o tempo de resposta ser menor do que 0,5 segundos?

Exemplo 1

- Um servidor web recebe solicitações em uma taxa de 3000 chegadas por minuto. A taxa de serviço do servidor é 58,8235 solicitações por segundo.

$$\lambda = 3000/60 = 50 \text{ solicitações por segundo}$$

$$\mu = 58,8235 \text{ solicitações por segundo}$$

$$\text{fila} = \text{FilaMMm}(50, 58.8235, 1)$$

- Qual é o tempo de serviço médio?

$$\text{fila.E_S} \quad E[S] = 0,017 \text{ segundos}$$

- Qual é o tempo de resposta médio?

$$\text{fila.E_R} \quad E[R] = 0,1133 \text{ segundos}$$

- Qual é o tamanho médio da fila?

$$\text{fila.E_Nq} \quad E[N_q] = 4,8167$$

Exemplo 1

- Um servidor web recebe solicitações em uma taxa de 3000 chegadas por minuto. A taxa de serviço do servidor é 58,8235 solicitações por segundo.

$$\lambda = 3000/60 = 50 \text{ solicitações por segundo}$$

$$\mu = 58,8235 \text{ solicitações por segundo}$$

$$\text{fila} = \text{FilaMMm}(50, 58.8235, 1)$$

- Qual é a probabilidade de o servidor ficar ocioso?

$$\text{fila.pmf_N}(0) \quad P[N=0] = 0,1500$$

- Qual a probabilidade de haver mais do que 2 tarefas no sistema

$$1-(\text{fila.pmf_N}(0)+\text{fila.pmf_N}(1)+\text{fila.pmf_N}(2)) \quad P[N>2] = 0,6141$$

- Qual é a probabilidade de o tempo de resposta ser menor do que 0,5 segundos?

$$\text{fila.cdf_R}(0.5) \quad P[R<0.5] = 0,9879$$

Exemplo 2

- Um sistema com um único servidor recebe 4 tarefas por segundo, com os tempos distribuídos exponencialmente. O tempo de serviço médio é de 0,208 segundos.
- Dados: $\lambda = 4$ $E[S] = 0,208$ $m = 1$
 $\mu = 1/E[S] = 1/0,208 = 4,8$ tarefas por segundo
fila1 = FilaMMm(4, 4.8, 1)
- Qual é o tempo de resposta médio?
fila1.E_R
 $E[R] = 1,2500$ segundos

Exemplo 2

- Suponha que foi comprado outro servidor para o sistema, e que metade do fluxo de entrada seja direcionado para cada servidor.
- Dados: $\lambda = 2$ $E[S] = 0,208$ $m = 1$
 $\mu = 1/E[S] = 1/0,208 = 4,8$ tarefas por segundo
 $\text{fila2} = \text{FilaMMm}(2, 4.8, 1)$
- Qual é o tempo de resposta médio?
 fila2.E_R
 $E[R] = 0,3571$ segundos

Exemplo 2

- Suponha que foi comprado outro servidor para o sistema, e que os dois servidores vão ser utilizados sejam através de uma fila única (M/M/2).
- Dados: $\lambda = 4$ $E[S] = 0,208$ $m = 2$
 $\mu = 1/E[S] = 1/0,208 = 4,8$ tarefas por segundo
 $\text{fila3} = \text{FilaMMm}(4, 4.8, 2)$
- Qual é o tempo de resposta médio?
 fila3.E_R
 $E[R] = 0,2521$ segundos

Exemplo 2

- Suponha que em vez de comprar dois servidores, foi comprado um servidor duas vezes mais rápido.
- Dados: $\lambda = 4$ $E[S] = 0,104$ $m = 1$
 $\mu = 1/E[S] = 1/0,104 = 9,6$ tarefas por segundo
 $\text{fila4} = \text{FilaMMm}(4, 9.6, 1)$
- Qual é o tempo de resposta médio?
 fila4.E_R
 $E[R] = 0,1786$ segundos

Comparação dos Modelos

- Uma fila M/M/2 é melhor do que 2 filas M/M/1
- Este resultado é sempre válido: fila única é sempre melhor do que múltiplas filas onde a taxa de serviço é igual
- A melhor solução é a última, onde a taxa de serviço foi dobrada
 - Em muitos casos pode ser a solução mais cara
 - A decisão por uma ou outra escolha deve ser feita por uma análise de custo/benefício

Avaliação de Desempenho

Filas MMm