



### PLANEJAMENTO DE CAPACIDADE, MODELAGEM E AVALIAÇÃO DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

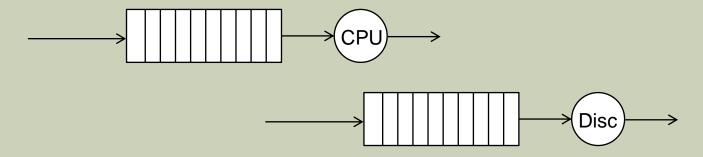
### MODELAGEM MATEMÁTICA DE SISTEMAS COPUTACIONAIS

Professor: Luis Enrique Zárate

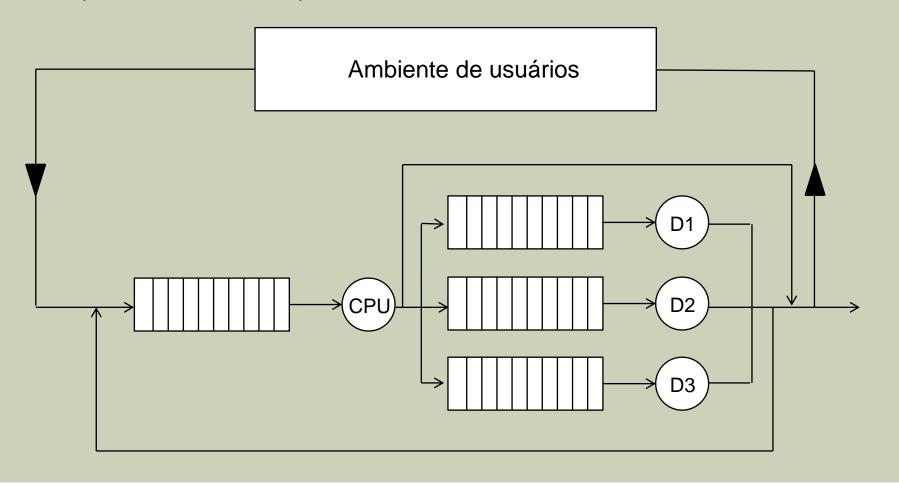
A modelagem matemática de Sistema Computacionais baseado na teoria das filas é fundamentada no fato de um sistema computacional, ao possuir velocidade finita, produz fila de espera.

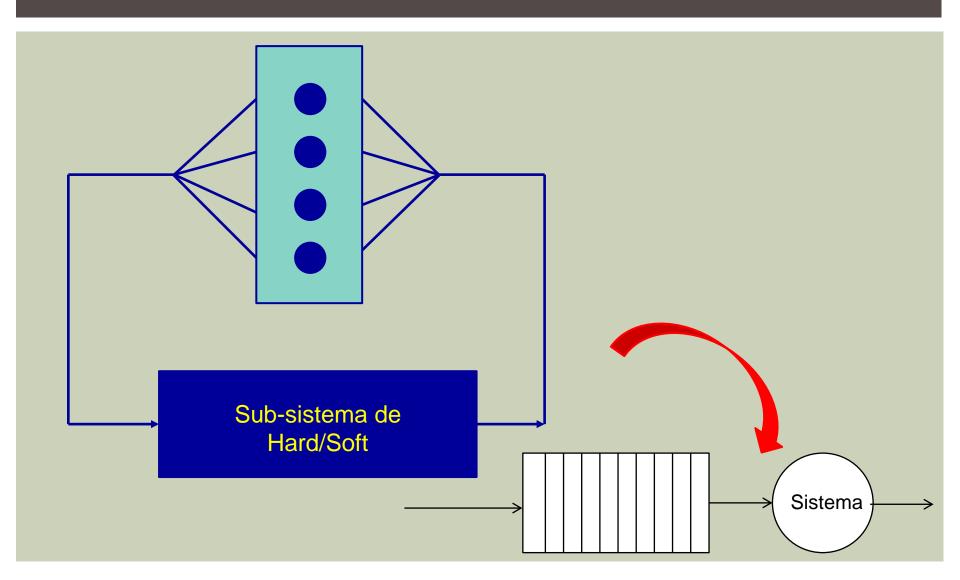
A modelagem baseado na teoria das filas, pode modelar qualquer sistema em distintos níveis de abstração.

**Dispositivos isolados:** 

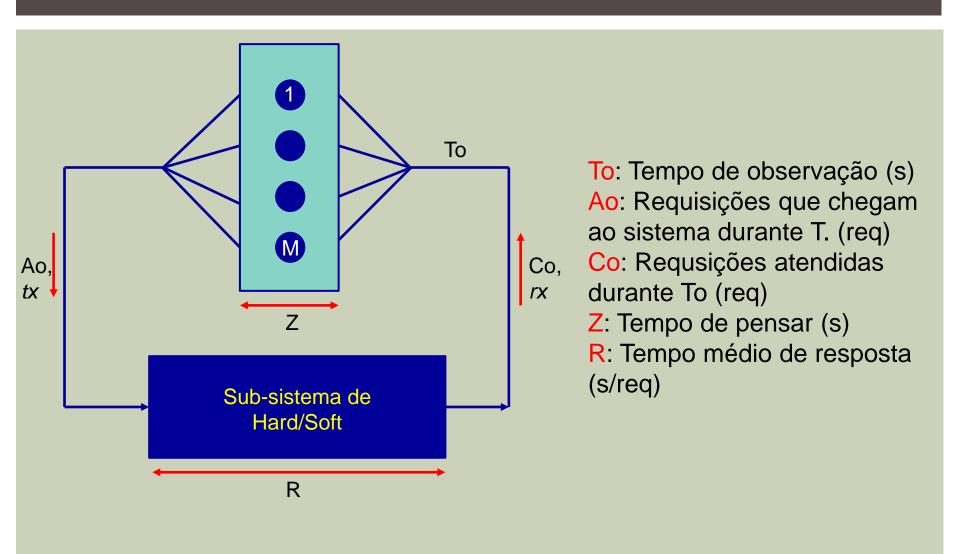


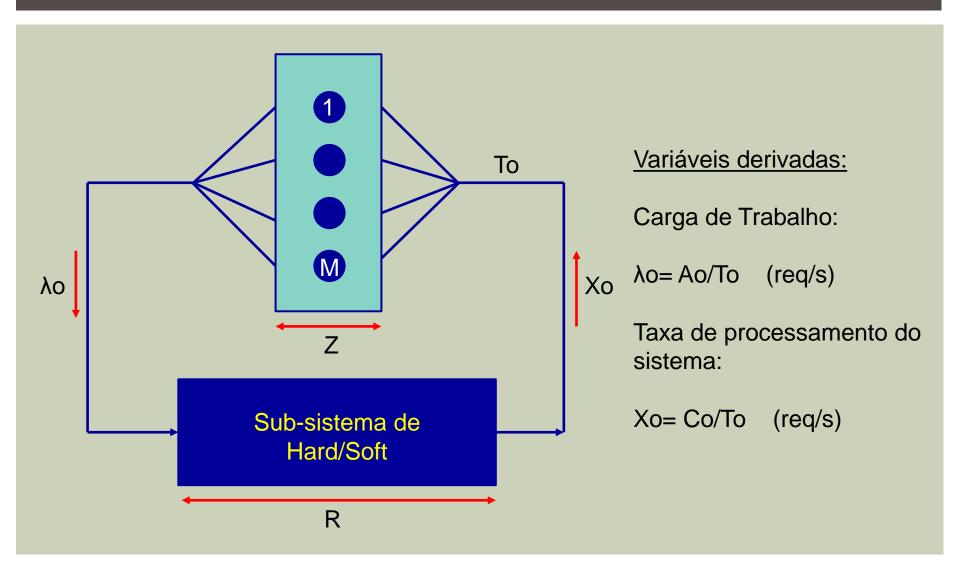
Exemplo de Sistema Computacional:

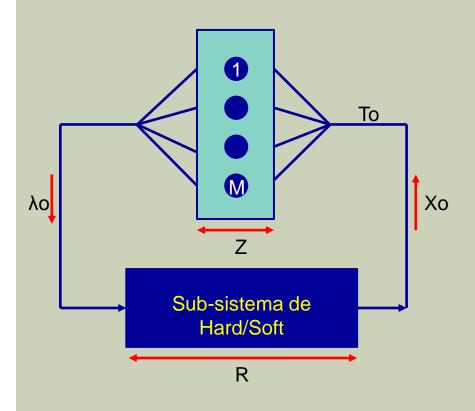




### Modelagem de sistemas interativos







#### Modelando o Sistema:

Considerando tempos médios,

$$Z + tx + R + rx = M/Xo$$

Tipícamente: Z ≈ 7 s.

Como  $rx \approx 0$ ,  $rx \approx 0$ , quando comparando com Z

$$Z + R = M/Xo$$

Tempo médio de resposta:

$$R = M/Xo - Z$$

#### Exemplo de aplicação:

Um sistema interativo foi observado durante 1 hora. Durante esse período de tempo foram atendidas 7200 requisições de usuários. Par aum tempo de pensar de 7 s. Calcular o tempo médio de resposta (R) se o número de terminais é de 40.

To = 
$$3600 \text{ s}$$
; Co =  $7200 \text{ req.}$ ; M =  $40$ ; Z =  $7\text{s}$ 

$$Xo = Co/To = 7200/3600 = 2 \text{ req./s}$$

Tempo médio de resposta:

$$R = M/Xo - Z = (40/2) - 7 = 13 \text{ s/req}$$

#### Melhorando o Desempenho do Sistema:

Para diminuir R↓ é possível:

$$R \downarrow = M \downarrow / Xo \uparrow - Z$$

a) <u>Diminuíndo o número de terminais (M):</u>

Supondo que 10 terminais são desligados:

M = 30 terminais

To = 
$$3600 \text{ s}$$
; Co =  $7200 \text{ req.}$ ; Z =  $7\text{s}$ 

Tempo médio de resposta:

$$R = M/Xo - Z = (30/2) - 7 = 8 \text{ s/req}.$$

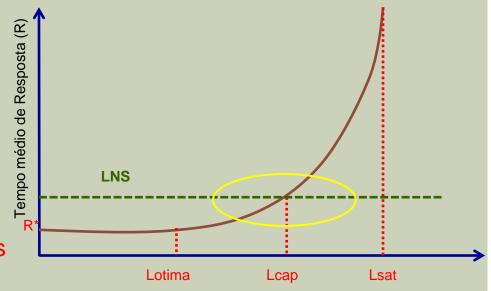
b) Propondo um tempo de resposta limite (LNS):

Para R = 3 s/req

To = 3600 s; Co = 7200 req.; Z = 7s

Número de terminais:

M = Xo (R + Z) = 2 (3 + 7) = 20 terminais



c) Número minimo de teriinas:

 $M_min = ?$ 

d) <u>Aumentando a taxa de processamento:</u>

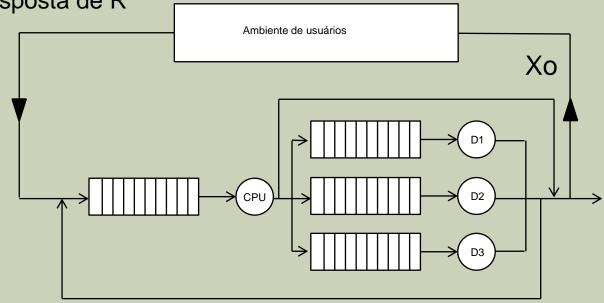
Sendo:

$$Xo = M / (R + Z)$$

Para um Tempo médio de resposta de R

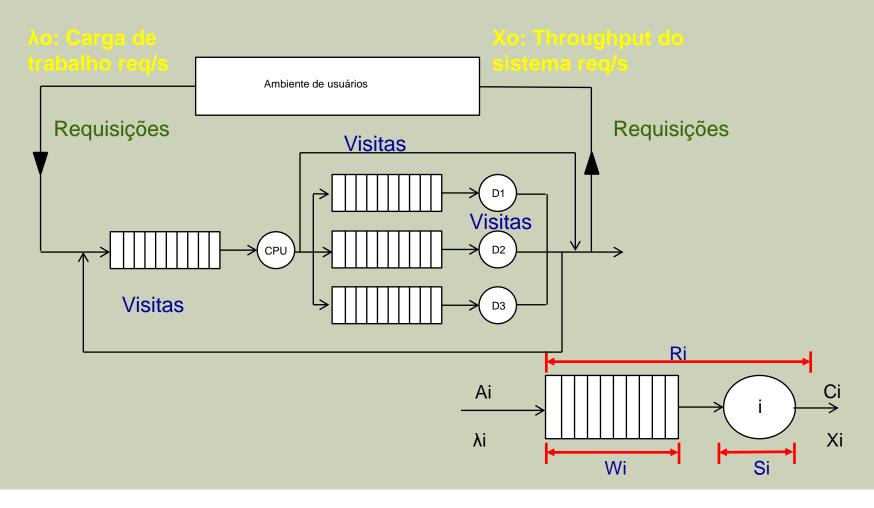
= 3 s/req.

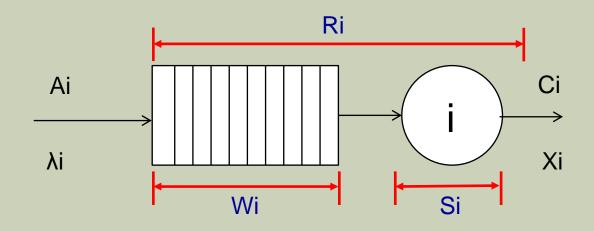
$$Xo = 40 / (3 + 7) = 4 \text{ req/s}$$





### Requisições e visitas por requisição:





### Variáveis Básicas:

To: Tempo de observação [s.]

Ai: Visitas que chegam ao dispositivo "i", durante To [visitas]

Bi: Tempo de ocupação do dispositivo "i", durante To [s.]

Ci: Visitas atendidas pelo dispositivo "i", durante To [visitas]

Exemplo: 
$$To = 10 \text{ s. Bi} = 8 \text{ s. Ai} = Ci = 10 \text{ v.}$$

#### Variáveis Derivadas:

$$\lambda$$
i: Carga de trabalho do dispositivo "i", [v/s]  
 $\lambda$ i = Ai/To = 10/10 = 1 v/s

Xi: Taxa de processamento do dispositivo "i", [v/s]  
Xi = Ci/To = 
$$10/10 = 1$$
 v/s

Exemplo: 
$$To = 10 \text{ s. Bi} = 8 \text{ s. Ai} = Ci = 10 \text{ v.}$$

$$Udi = 1 - Ui = 1 - 0.80 = 0.20 => 20 \%$$

$$Si = Bi/Ci = 8/10 = 0.8 s/v$$

Ri: Tempo médio de resposta do dispositivo "i" [s/v]

$$Ri = Si/(1-Ui) = 0.8/(1-0.8) = 4 s/v$$

Wi: Tempo médio de espera do dispositivo "i" [s/v]

$$Wi = Ri - Si = 4.0 - 0.8 = 3.2 \text{ s/v}$$

#### Lei da Utilização:

#### Como:

$$Ui = \frac{Bi}{To}$$

$$Ui = \frac{Bi}{To} x \frac{Ci}{Ci}$$

$$Ui = Si \times Xi$$

# Ai Ci Xi Xi

#### Teorema da Utilização:

Pela hipótese do Equilíbrio de Fluxo:

$$Ai = Ci$$

Então:

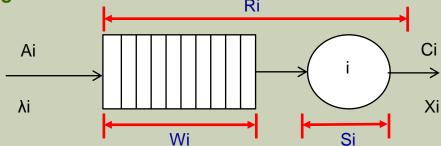
$$\frac{Ai}{To}x\frac{Ci}{To}$$

Por tanto:  $\lambda i = Xi$ 

Logo:  $Ui = Si \times \lambda i$ 

#### Aumentando a carga de trabalho em 10 % (mês 1)

$$\lambda i_novo = 1,10 \times \lambda i_atual$$
  
 $\lambda i_novo = 1,10 \times 1,0 = 1,10 \text{ v/s}$ 



$$Udi = 1 - Ui = 1 - 0.880 = 0.12 => 12 \%$$

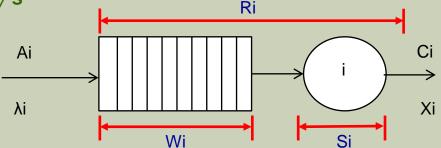
$$Ri = Si/(1-Ui) = 0.8/(1-0.88) = 6.7 s/v$$

$$Wi = Ri - Si = 6.7 - 0.8 = 5.9 \text{ s/v}$$

#### Aumentando a carga mais trabalho em 10 % (mês 2)

$$\lambda i_novo = 1,10 \times \lambda i_novo = 1$$

$$\lambda i_novo = 1,10 \times 1,10 = 1,21 \text{ v/s}$$



$$Ui_novo = 0.88 \times 1.10 = 0.97 => 97 \%$$

$$Udi = 1 - Ui = 1 - 0.97 = 0.03 => 3 \%$$

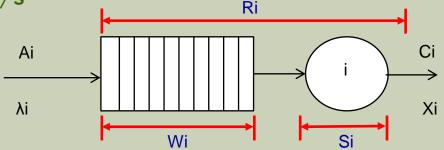
$$Ri = Si/(1-Ui) = 0.8/(1-0.97) = 26.7 \text{ s/v}$$

$$Wi = Ri - Si = 26,7 - 0,8 = 25,9 \text{ s/v}$$

#### Aumentando a carga mais trabalho em 10 % (mês 3)

$$\lambda i_novo = 1,10 \times \lambda i_novo = 1$$

$$\lambda i_novo = 1,10 \times 1,21 = 1,31 \text{ v/s}$$



$$Udi = 1 - Ui = 1 - 1,06 => 0,0 \%$$

$$Ri = Si/(1-Ui) = 0.8/(1-0.99) = 80 s/v$$

$$Wi = Ri - Si = 80 - 0.8 = 79.2 \text{ s/v}$$

$$\lambda i_atual = 1.0 \text{ v/s} -> 60 \text{ v/min}$$

$$Ri_atual = 4 s/v$$

$$Ui_atual = 0.80$$

#### Mês 1:

$$\lambda i_1 = 1,10 \text{ v/s} -> 66 \text{ v/min}$$

$$Ri_1 = 6.7 \text{ s/v}$$

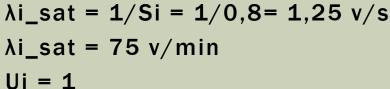
$$Ui_1 = 0.88$$

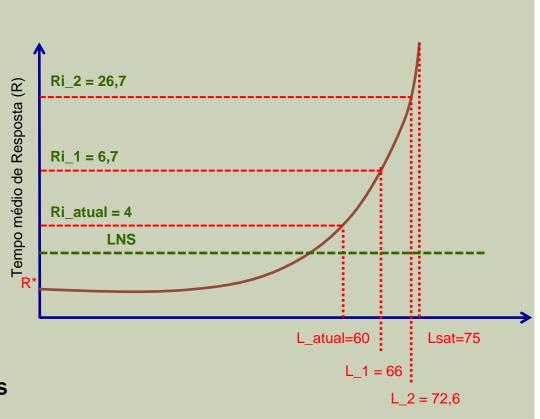
#### Mês 2:

$$\lambda i_2 = 1.21 \text{ v/s} -> 72.6 \text{ v/min}$$

$$Ri_2 = 26.7 \text{ s/v}$$

$$Ui_2 = 0.97$$





#### **Troca de Dispositivo:**

Considere a carga futura de 1,32 v/s e um tempo limite (LNS) para o tempo de resposta de 5 s/v. Determine as características do novo dispositivo

$$\lambda i = 1,32 \text{ v/s}; \text{Ri} = 5 \text{ s/v}; \text{Si} = ?$$

$$Si_novo = 5/(1+5x1,32) = 0.65 s/v$$

#### Como:

$$Ri = Si/(1-Ui)$$

е

$$Ui = Si \times \lambda i$$

#### Substituíndo:

Ri = Si/(1-Ui) = Si/(1-Si 
$$x \lambda i$$
)

#### Então:

$$Si = Ri/(1+Rix \lambda i)$$

Como: Si ant = 
$$0.80 \text{ s/v}$$

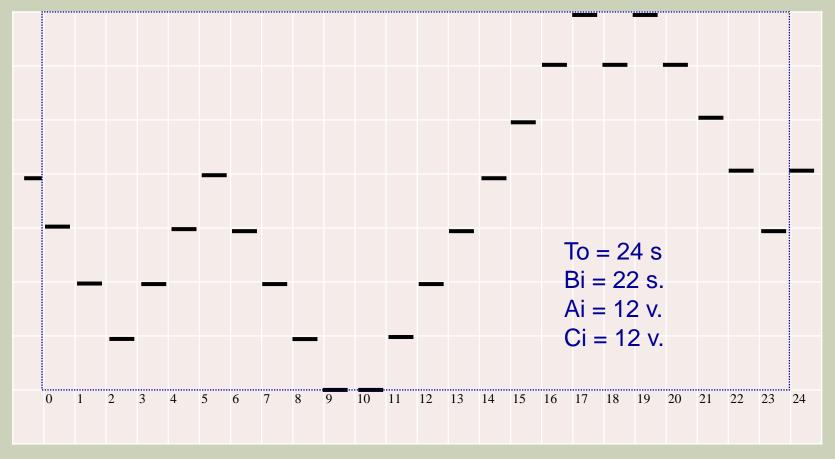
Fator=
$$0.80/0.65 = 1.23$$
 (23% mais rápido)

#### Exemplo:

Se processador atual faz 10^6 somas/s -> 1230,000 somas/s

Se disco gasta 10 ms/IO -> 7,7 ms/IO

#### Diagramas de Sequenciamento:



Exemplo: 
$$To = 24 \text{ s. Bi} = 22 \text{ s. Ai} = Ci = 12 \text{ v.}$$

#### Variáveis Derivadas:

λi: Carga de trabalho do dispositivo "i", [v/s]

$$\lambda i = Ai/To = 12/24 = 0.5 \text{ v/s}$$

Xi: Taxa de processamento do dispositivo "i", [v/s]

$$Xi = Ci/To = 12/24 = 0.5 \text{ v/s}$$

Ui: Utilização do dispositivo "i", adim, [%]

$$Ui = Bi/To = 22/24 = 0.92 => 92 \%$$

Exemplo: 
$$To = 24 \text{ s. Bi} = 22 \text{ s. Ai} = Ci = 12 \text{ v.}$$

Udi: Disponibilidade do dispositivo "i", adim, [%]

$$Udi = 1 - Ui = 1 - 0.92 = 0.08 => 8 \%$$

Si: Tempo médio de serviço do dispositivo "i", [s/v]

$$Si = Bi/Ci = 22/12 = 1,83 s/v$$

Ri: Tempo médio de resposta do dispositivo "i" [s/v]

$$Ri = Si/(1-Ui) = 1.83/(1-0.92) = 22.92 s/v$$

Wi: Tempo médio de espera do dispositivo "i" [s/v]

$$Wi = Ri - Si = 22,92 - 1,83 = 21,09 \text{ s/v}$$

#### 1) Otimizando Código em 25 %

Mantendo a carga constante:  $\lambda i = 0.5 \text{ v/s}$ 

Ui\_novo = Si x 
$$\lambda$$
i\_novo (Teorema da Utilização)  
Ui\_novo = 1,37 x 0,50 = 0,68 => 68 %

$$Ri_novo = Si_novo/(1-Ui) = 1,37/(1-0,68) = 4,28 s/v$$

$$Wi = Ri - Si = 4,28 - 1,37 = 2,91 \text{ s/v}$$

#### 2) Aumentando a Carga de trabalho em 10%

#### Exemplo de Modelo:

To = 1 hora

Co = 7200 req. Concluídas

Ucpu = 60% Como: 
$$D_i = \frac{Ui}{XO}$$
 e  $X_O = \frac{CO}{TO}$ 

$$Ud1 = 50\%$$

$$Ud2 = 80\%$$

$$Ud3 = 90\%$$

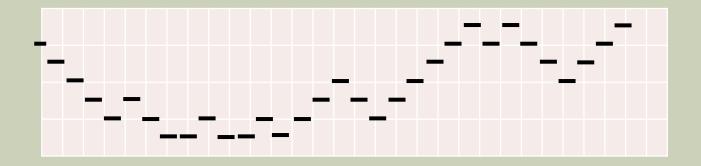
$$Xo = Co/To = 2 req./s$$

$$R = \frac{0,60/2}{1 - 0,60} + \frac{0,50/2}{1 - 0,50} + \frac{0,80/2}{1 - 0,80} + \frac{0,90/2}{1 - 0,90}$$

$$R = \frac{0,30}{0,40} + \frac{0,25}{0,50} + \frac{0,40}{0,20} + \frac{0,45}{0,10}$$

$$R = 0.75 + 0.50 + 2.00 + 4.50$$

$$R = 7,75 \text{ s/req.}$$



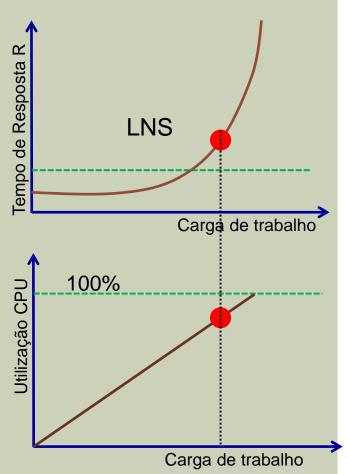
#### Aumentando a carga em 10%

$$R = \frac{0,30}{1 - 0,66} + \frac{0,25}{1 - 0,55} + \frac{0,40}{1 - 0,88} + \frac{0,45}{1 - 0,99}$$

$$R = \frac{0,30}{0,34} + \frac{0,25}{0,45} + \frac{0,40}{0,12} + \frac{0,45}{0,01}$$

$$R = 0.88 + 0.55 + 3.33 + 45.0$$

$$R = 49,76 \text{ s/req}.$$

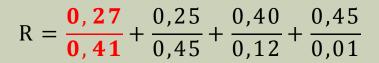


#### Melhorando o código em 10%

$$R = \frac{0,30}{1 - 0,66} + \frac{0,25}{1 - 0,55} + \frac{0,40}{1 - 0,88} + \frac{0,45}{1 - 0,99}$$

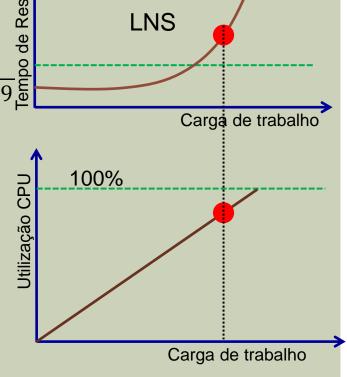
$$R = \frac{0,30}{1 - 0,66} + \frac{0,25}{1 - 0,55} + \frac{0,40}{1 - 0,88} + \frac{0,45}{1 - 0,99}$$

$$R = \frac{0,30 * 0,90}{1 - 0,66 * 0,90} + \frac{0,25}{1 - 0,55} + \frac{0,40}{1 - 0,88} + \frac{0,45}{1 - 0,99}$$



$$R = 0,66 + 0,55 + 3,33 + 45,0$$

$$R = 49,54 \text{ s/req}.$$



#### **Aumentando o storage**

$$R = \frac{0,27}{0,41} + \frac{0,25}{0,45} + \frac{0,40}{0,12} + \frac{0,45}{1 - 0,99}$$

$$R = \frac{0,27}{0,41} + \frac{0,25}{0,45} + \frac{0,40}{0,12} + \frac{0,45}{1 - 0,50} + \frac{0,45}{1 - 0,50}$$

$$R = 0.66 + 0.55 + 3.33 + 0.90 + 0.90$$

$$R = 6.34 \text{ s/req.}$$

#### Exercício:

To = 1 hora

Co = 5400 req. Concluídas

Ucpu = 40% Como: 
$$D_i = \frac{Ui}{Xo}$$
 e  $X_o = \frac{Co}{To}$ 

Ud1 = 80%

Ud2 = 80%

Ud3 = 90%

$$Xo = Co/To = 1.5 req./s$$

$$R = \frac{0,40/2}{1 - 0,40} + \frac{0,80/2}{1 - 0,80} + \frac{0,80/2}{1 - 0,80} + \frac{0,90/2}{1 - 0,90}$$

$$R = \frac{0,20}{0.60} + \frac{0,40}{0.20} + \frac{0,40}{0.20} + \frac{0,45}{0.10}$$

$$R = 0.33 + 2.00 + 2.00 + 4.50$$

$$R = 8,83 \text{ s/req.}$$

#### Aumentando a carga em 10%

$$R = \frac{0,40/2}{1 - 0,44} + \frac{0,80/2}{1 - 0,88} + \frac{0,80/2}{1 - 0,88} + \frac{0,90/2}{1 - 0,99}$$

$$R = \frac{0,20}{0,56} + \frac{0,40}{0,12} + \frac{0,40}{0,12} + \frac{0,45}{0,01}$$

$$R = 0,36 + 3,33 + 3,33 + 45,0$$

$$R = 52,02 \text{ s/req.}$$

#### Aumentando o storage

$$R = \frac{0,20}{0,56} + \frac{0,40}{0,12} + \frac{0,40}{0,12} + \frac{0,45}{1 - 0,50} + \frac{0,45}{1 - 0,50}$$

$$R = 0.36 + 3.33 + 3.33 + 0.90 + 0.90$$

$$R = 8.82 \text{ s/req.}$$

#### Tempo mínimo que pode ser alcançado:

$$R = 0.20 + 0.40 + 0.40 + 0.45 + 0.45$$

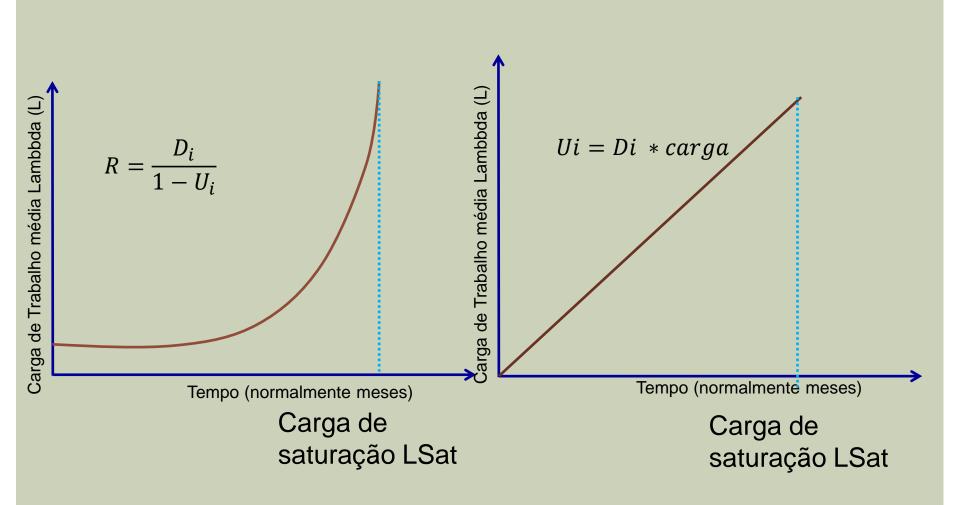
$$R = 1,90 \text{ s/req.}$$

Cómo determinar o tempo para a início da fase de Superutilização?

Tendo o modelo para a Carga de Trabalho:

$$L(t)=a+b*t$$

E, o modelo para o sistema computacional (considerando uma abstração do modelo):



### ETAPA 9: PREVISÃO DO INCÍCIO DA FASE DE SUPER-UTILIZAÇÃO

Na saturação do sistema:  $Usat = 1, R \rightarrow \infty$ 

$$Ui = Di * li$$

$$Usat = Di * lsat$$

$$lsat = \frac{1}{Di}$$

Exemplo: Di= 10 ms/req. Lsat=1/Di = 100 req./s

# ETAPA 9: PREVISÃO DO INCÍCIO DA FASE DE SUPER-UTILIZAÇÃO

Quando a saturação vai ocorrer: ?

Tendo o modelo para a Carga de Trabalho:

$$L(t)=a+b*t$$

$$t_sat = (Lsat - a)/b$$

O tempo até a saturação ocorre em:

Tempo = t\_sat - Histórico de meses

# ETAPA 9: PREVISÃO DO INCÍCIO DA FASE DE SUPER-UTILIZAÇÃO

#### **Exemplo:**

Para Di= 200 ms/req. Lsat=1/Di = 5 req./s

Tendo o modelo mensal para a Carga de Trabalho, em req/min:

$$L(t) = 50 + 25 * t$$

Como Lsat= 5 req./s = 300 req/min

$$t_sat = (Lsat - a)/b = (300 - 50)/25 = 10 meses$$

O tempo até a saturação ocorre em:

Tempo = t\_sat - Histórico de meses (Hist = 5 meses)

Tempo = 10 - 5 = 5 meses para chegar na saturação

### ETAPA 9: PREVISÃO DO INCÍCIO DA FASE DE SUPER-UTILIZAÇÃO

Na saturação do sistema,  $Usat = 1, R \rightarrow \infty$ 

$$Ui = Di * li$$
  
 $Usat = Di * lsat$ 

$$lsat = \frac{1}{Di}$$

Tendo o modelo para aCarga de Trabalho:

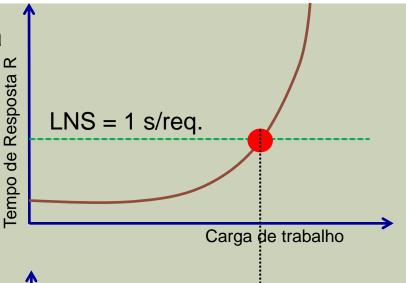
$$L(t)=a+b*t$$

E, o modelo do sistema computacional:

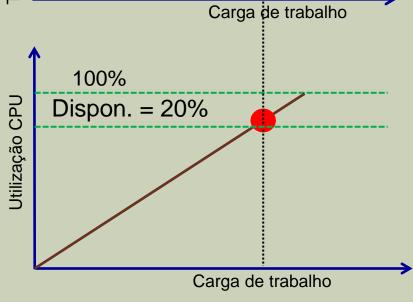
$$R = \frac{D_{CPU}}{1 - U_{CPU}} + \frac{D_{D1}}{1 - U_{D1}} + \frac{D_{D2}}{1 - U_{D2}} + \frac{D_{D3}}{1 - U_{D3}} + \dots + \frac{D_{D$$

# ETAPA 10: PROPOSTA DE NOVA CONFIGURAÇÃO

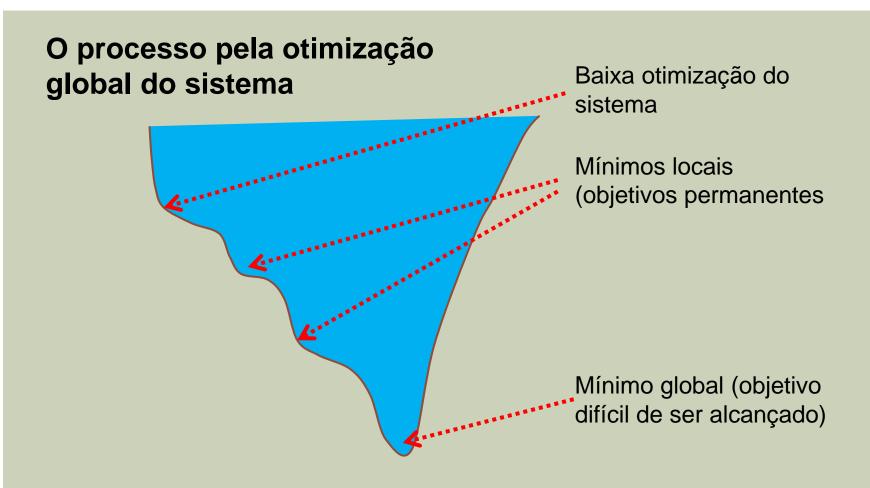
Considerando os modelos de carga de trabalho e do sistema computacional, o próximo passo é ajustar uma configuração que atenda requisitos como:



- -Vida útil do sistema
- -Disponibilidade do Sistema
- -Limites de QoS

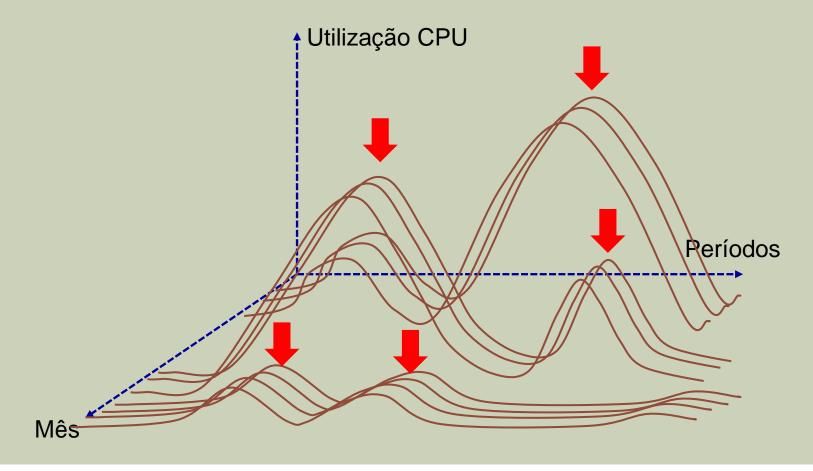


### ETAPA 6: META OTIMIZAÇÃO



## ETAPA 2: IDENTIFICAÇÃO DO HORÁRIO DE PICO DO SISTEMA COMPUTACIONAL

4) Identificar os horários de Pico:



## ETAPA 2: IDENTIFICAÇÃO DO HORÁRIO DE PICO DO SISTEMA COMPUTACIONAL

1) Variáveis a serem Monitoradas:

λ: Carga de Trabalho do sistema [req/s]

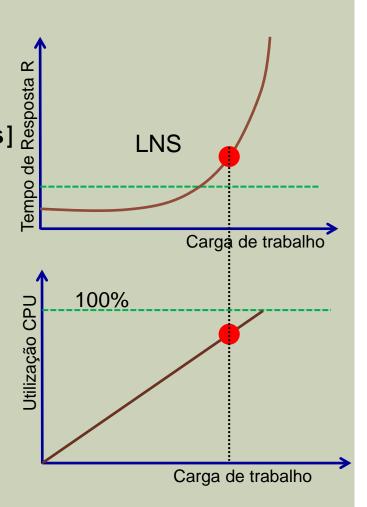
R: Tempo médio de resposta [s/req]

U: Utilização do sistema [%]

Du: Disponibilidade do sistema [%]

M: consumo médio de memória [%]

Pg: nível médio de paginação [%]



### CICLO DE VIDA DE UM SISTEMA COMPUTACIONAL

■ Fase Super-Utilização:

