PUC Minas

LICAP
Laboratório de Inteligência Computacional Aplicada

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

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Equipe MAD

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Definições preliminares:

1) Carga de Trabalho (L)

Corresponde ao conjunto de requisições que chegam ao sistema durante um intervalo de tempo. Unidades: [req./s], [req./min]...

2) Parâmetro de Controle (PtrC)

Corresponde a uma variável mensurável utilizada para avaliar o desempenho do sistema computacional. Tipicamente essas variáveis são:

Tempo médio de resposta R, [s/req.]

Utilização do Processador U, [%]

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Definições preliminares:

3) Nível de Serviço (NS)

Corresponde a um par numérico envolvendo o valor da variável de controle para um valor específico da carga de trabalho. NS = (L,R)

4) Limite do nível de serviço (LNS ou SLA)

Corresponde a uma valor subjetivo da variável de controle normalmente estipulado pela gerencia relacionado com o nível de satisfação do ambiente de usuários. É chamado ta´mbém de SLA.

Ex. Tempo de resposta LNS_R = 800 ms/req.

Utilização do processador LNS_U = 80%

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Definições preliminares:

5) Ponto de Crítico do Sistema (Pc)

Corresponde ao ponto de interseção da curva de desempenho com a reta do LNS. Pc = (Lc,LNS)

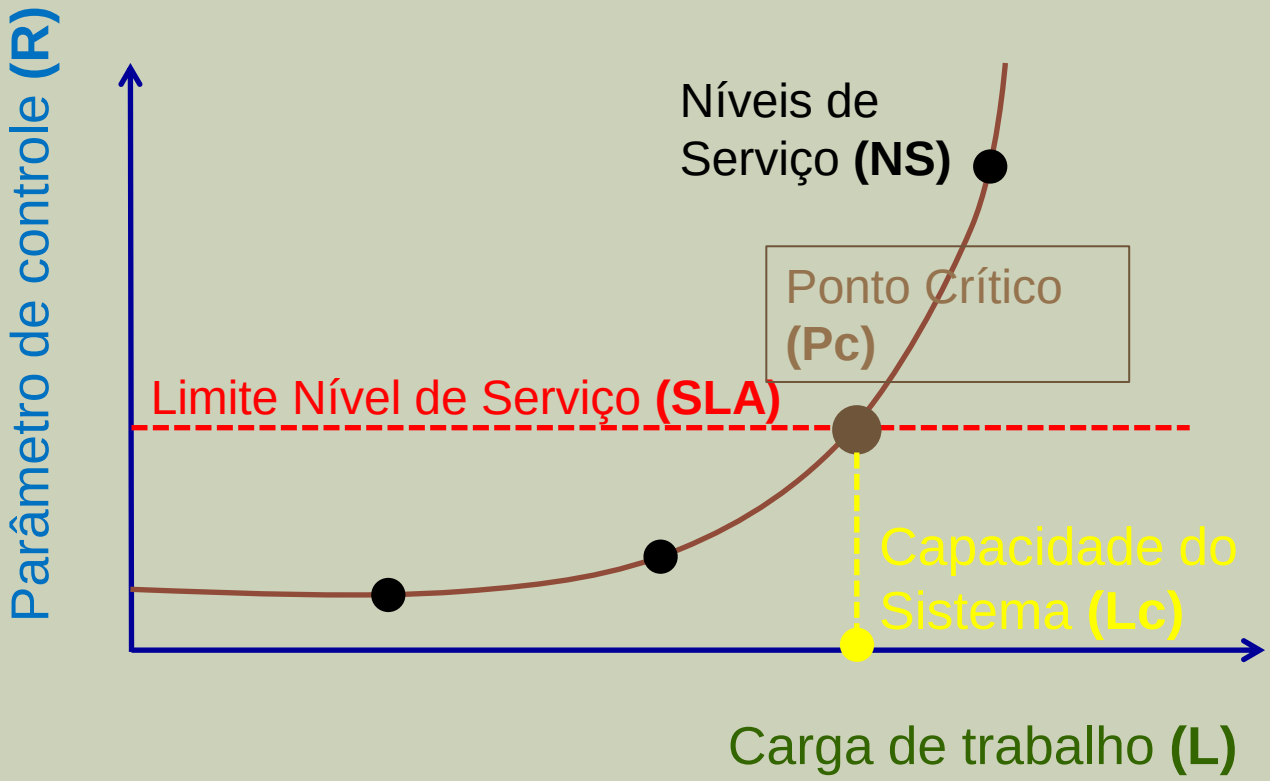
6) Capacidade de um Sistema Computacional (Lc)

Corresponde ao valor de carga (Lc) que leva ao sistema ao ponto crítico (Pc) ou seja ao limite do LNS.

7) Vida útil do Sistema Computacional

Corresponde a todo o tempo durante o qual o sistema opera, de forma permanente, abaixo do limite do LNS.

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS



CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Definições preliminares:

8) Região de Comportamento Constante (R1)

Nesta região, variações na carga de trabalho não causam variações no valor do Tempo de Resposta (R). Tipicamente esta região é caracterizada por níveis de utilização de 5%.

9) Região de Comportamento Proporcional (R2)

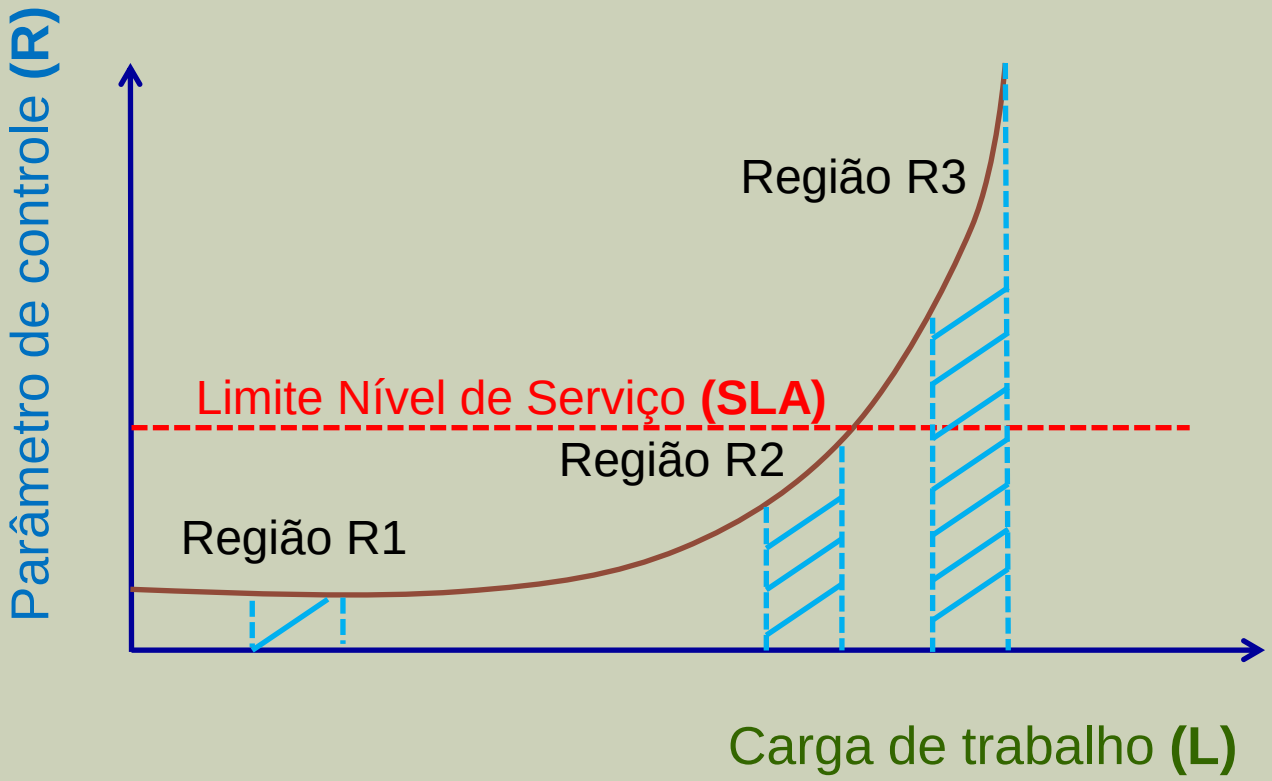
Nesta região, a proporção de variação na carga de trabalho causam variações proporcionais no valor do Tempo de Resposta (R). Tipicamente esta região é caracterizada por níveis de utilização de 30 a 40%.

10) Região de Comportamento Não-Linea (R3)

Nesta região, pequenas variações na carga de trabalho causam grandes variações no valor do Tempo de Resposta (R). Tipicamente esta região é caracterizada por níveis de utilização acima de 80%.

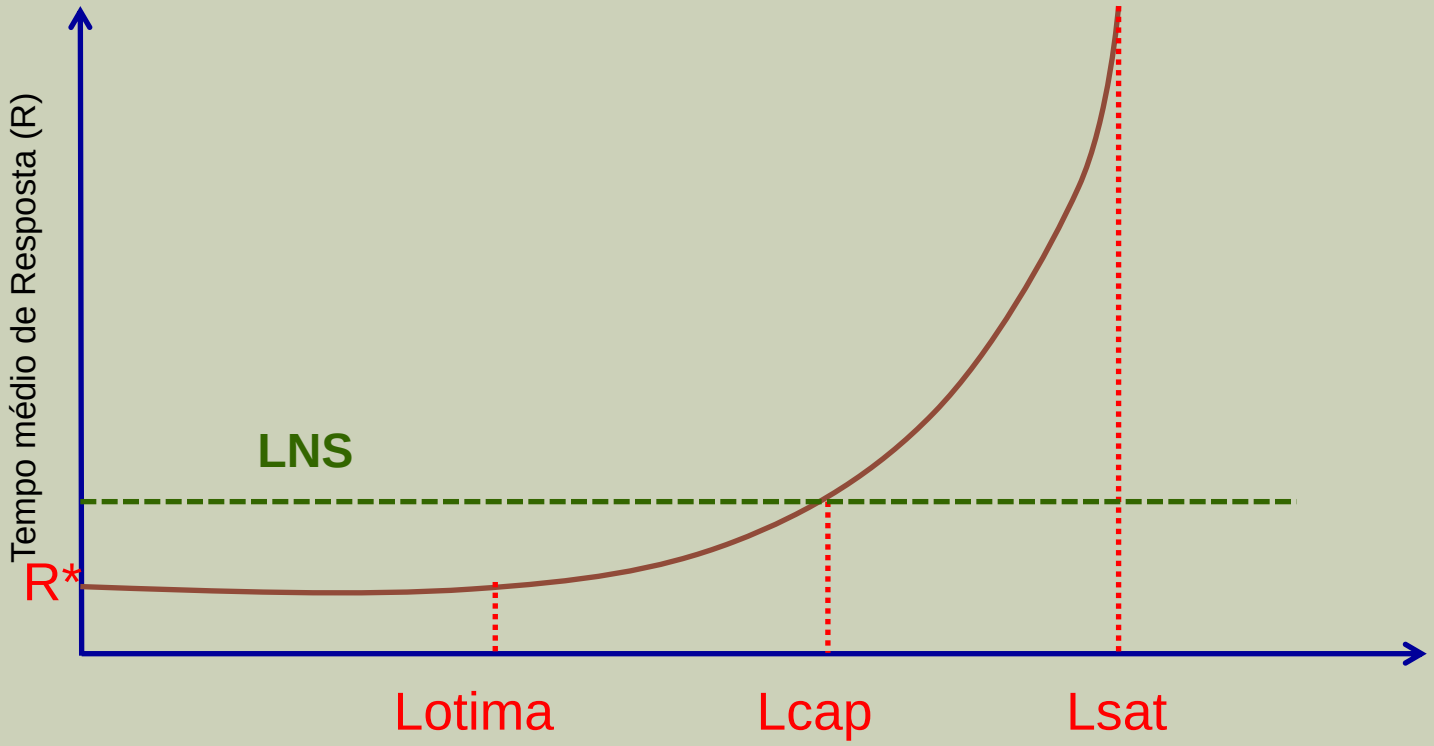
CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Regiões Típicas da Curva de Desempenho:



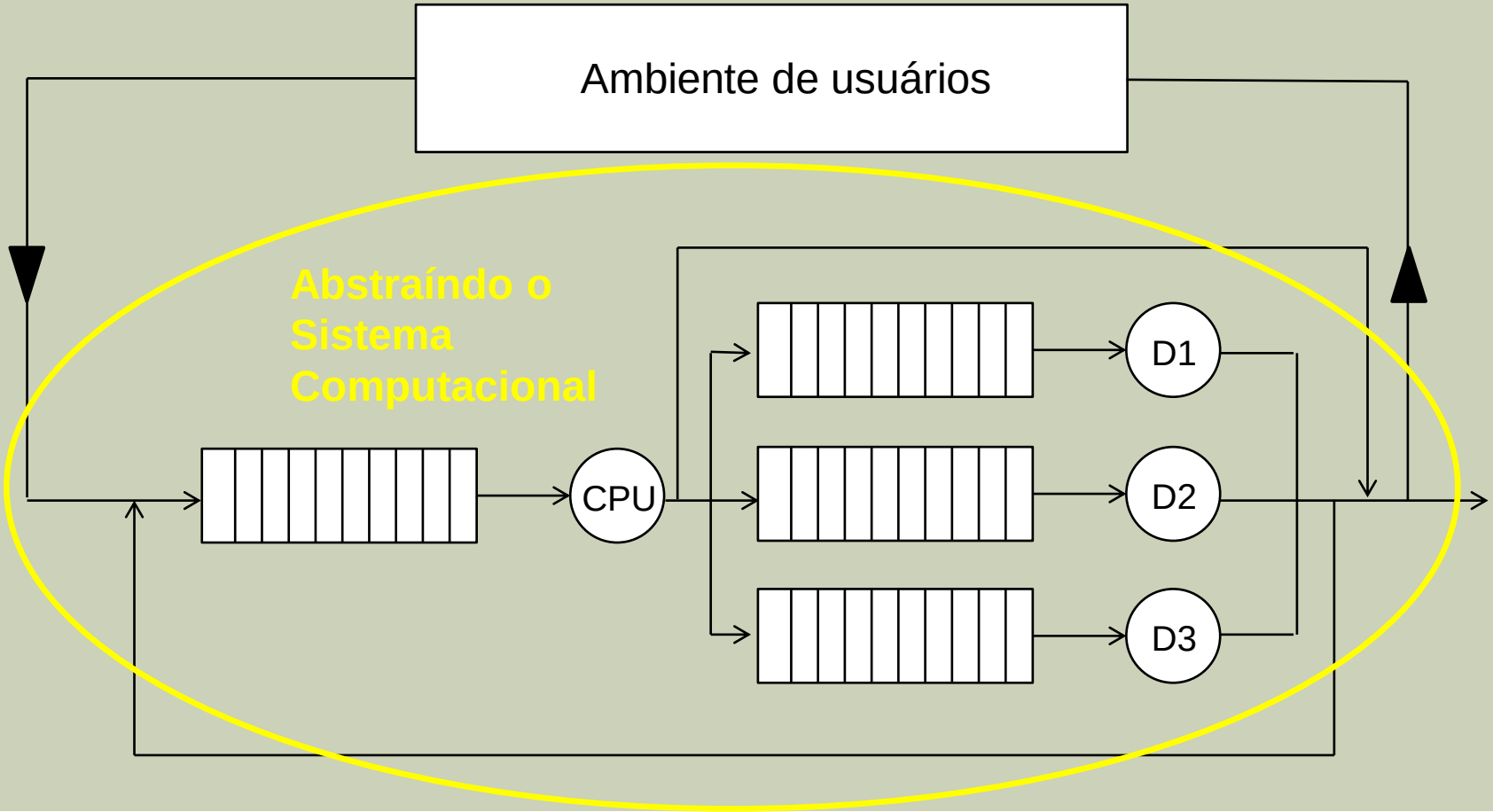
CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Parâmetros de Referencia da Curva de Desempenho:



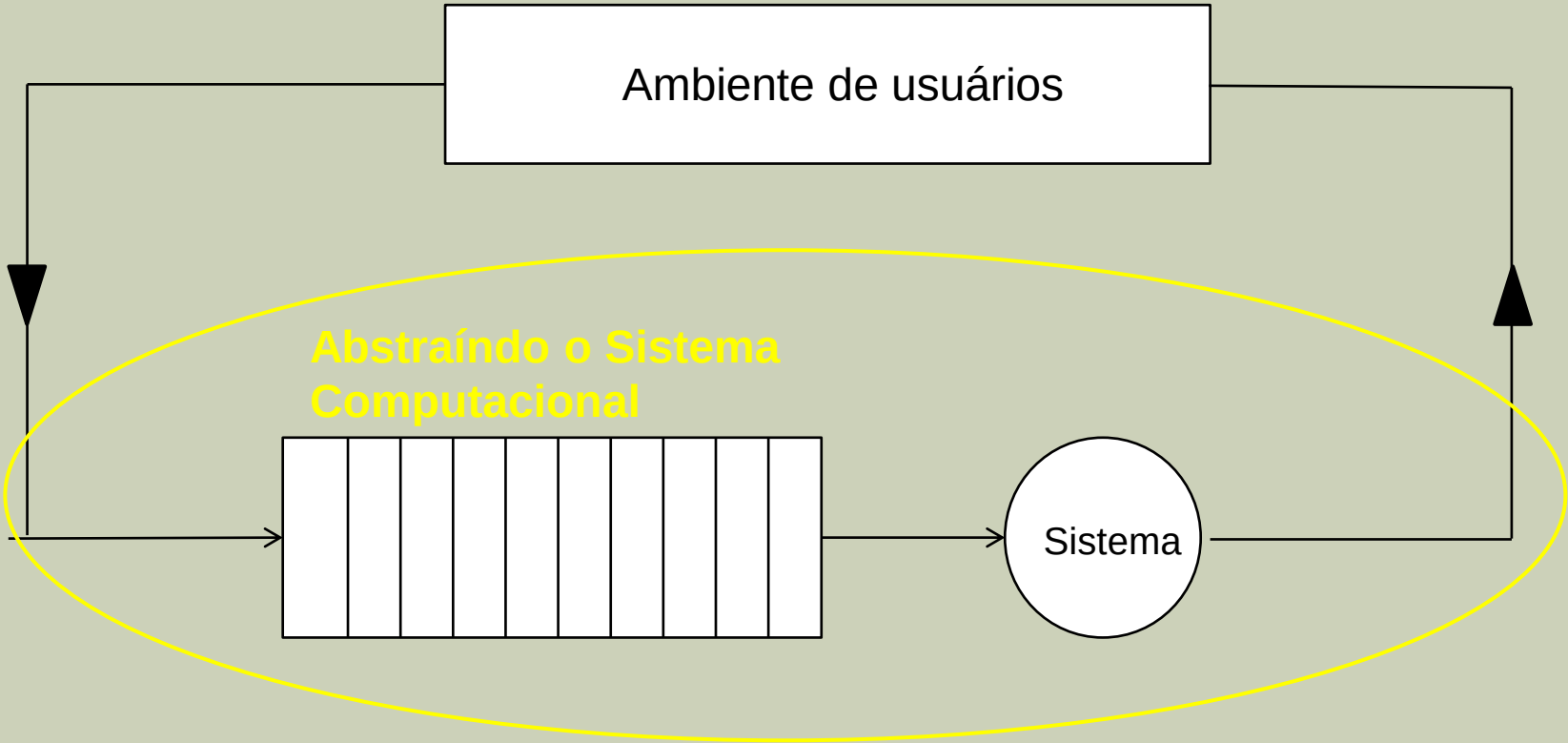
CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Modelo de Referência:



CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Modelo abstraído:



CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Modelo considerado:

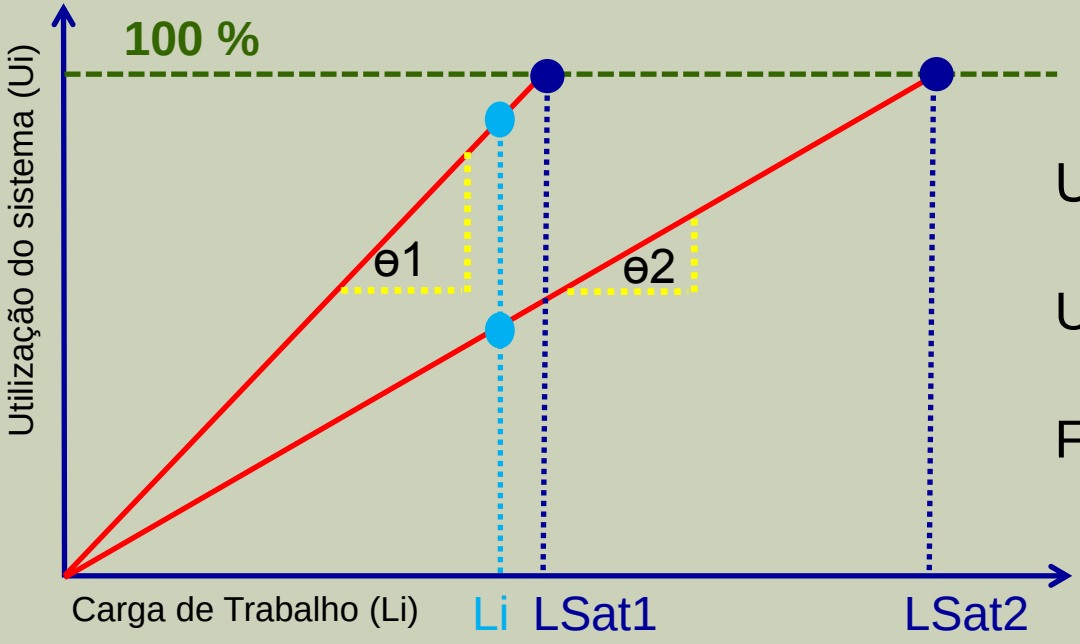
Considerando o modelo clássico da Teoria das Filas para um dispositivo:

$$R = \frac{D_i}{1 - U_i}$$

R: Tempo médio de resposta por requisição [s/req.]
Ui: Utilização do dispositivo “i”
Di: Tempo médio total gasto por uma requisição no dispositivo “i”, sem considerar tempo de espera.

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Relação Carga (Lambda) vs. Utilização (Ui):



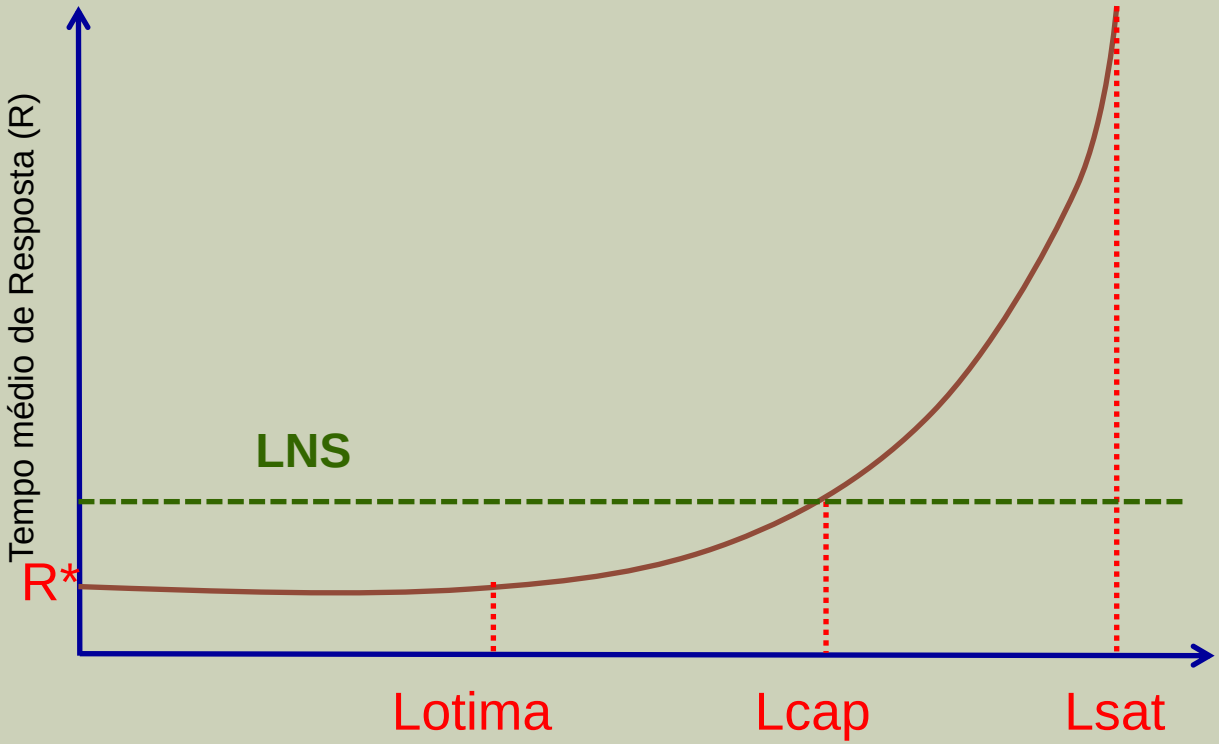
$Ui = Tg(\theta) * Li$

$Ui = Di * Li$

$\text{Fator Veloc.} = Tg(\theta 1)/Tg(\theta 2)$

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Determinando o Ponto de Início da Curva (R*):



Quando:
Li -> 0
Ui -> 0

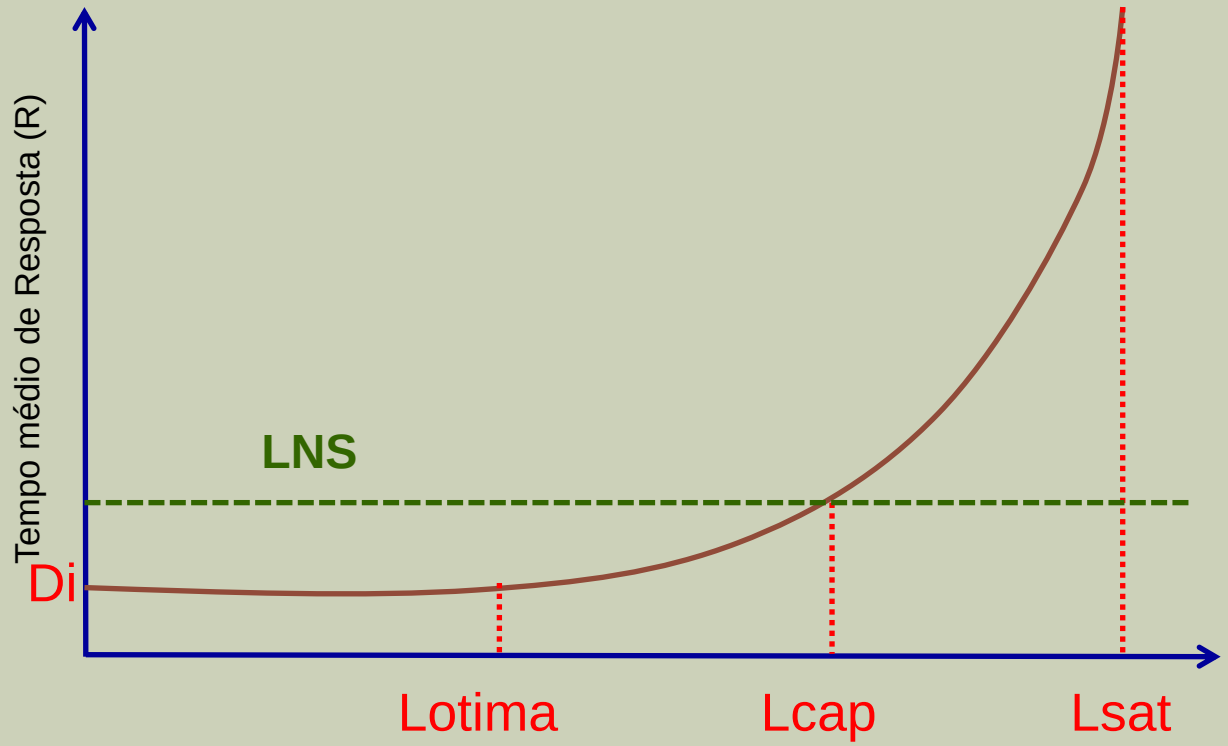
Dado que:

$$R = \frac{D_i}{1 - U_i}$$

Logo:
R* = Di

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Determinando a Carga de Saturação (Lsat):



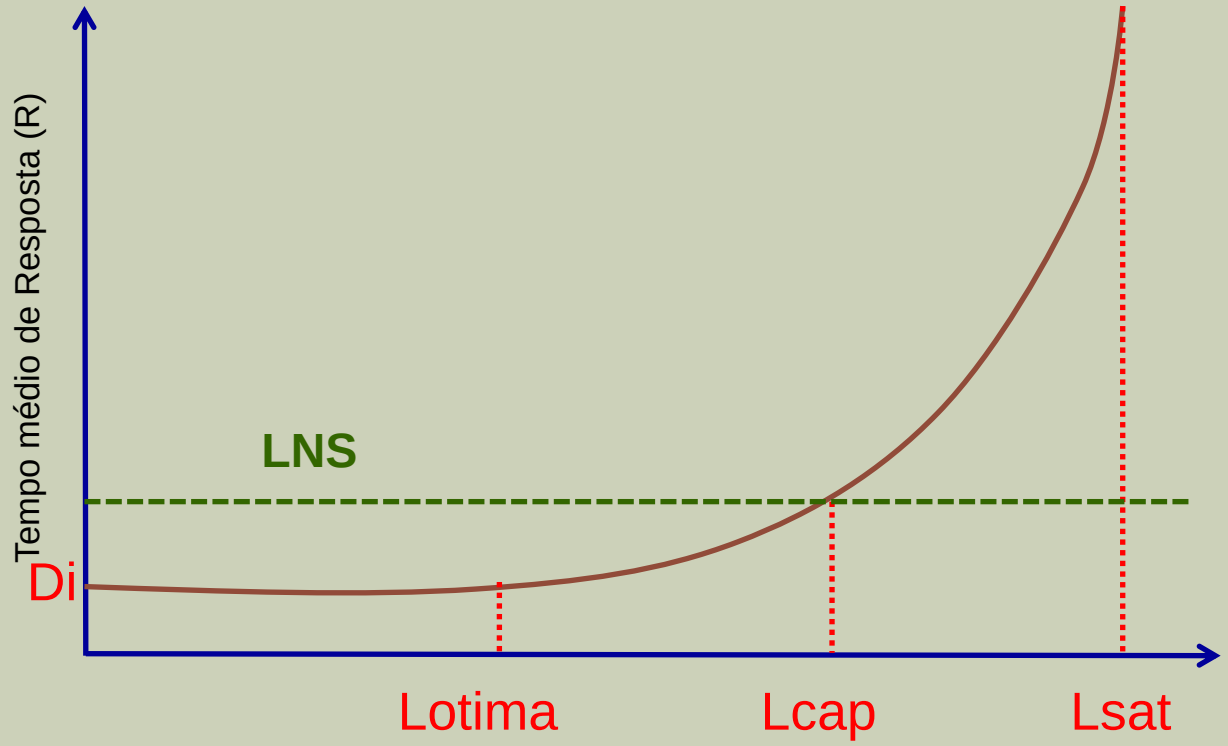
Quando:
Li -> Lsat
Ui -> 1

Dado que:
Ui = Di * Li
1 = Di * Lsat

Logo:
Lsat = 1/Di

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Determinando a Capacidade do Sistema (Lc):



Quando:
Li -> Lcap
Ri -> LNS

Dado que:

$$R = \frac{D_i}{1 - U_i}$$

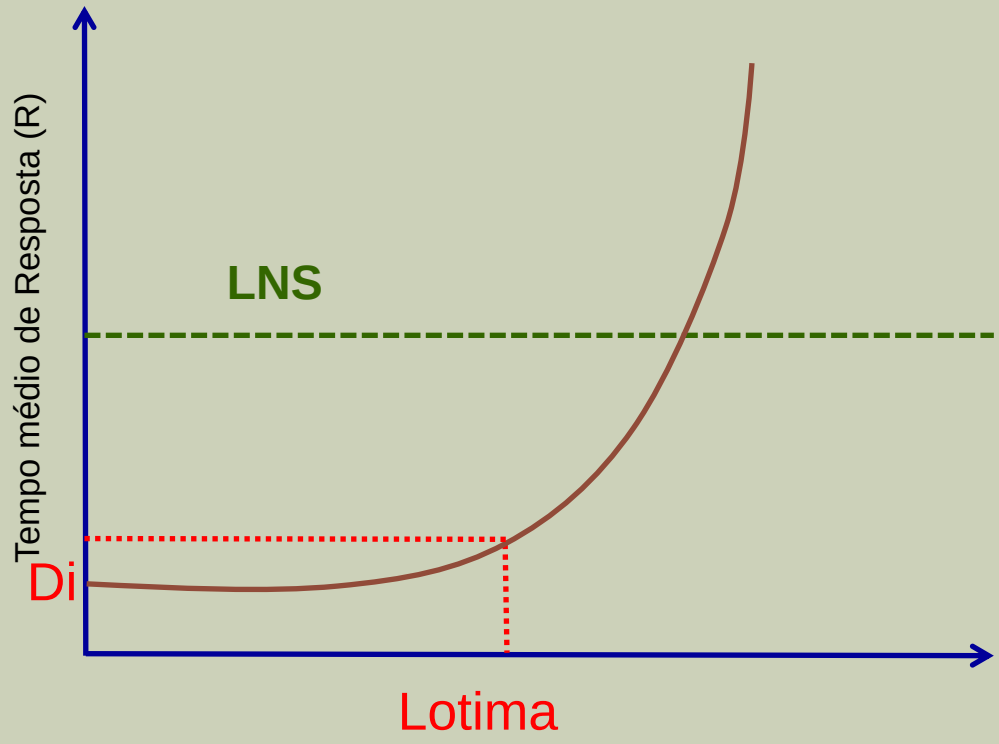
Então:

$$LNS = \frac{D_i}{1 - Di * Lcap}$$

$$\boldsymbol{Lcap = \frac{LNS - D_i}{LNS * Di}}$$

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Determinando a Carga Ótima do Sistema (Lotima):



Quando:
Li -> Lotima
Ri -> 1,05*Di

Dado que:

$$R = \frac{D_i}{1 - U_i}$$

Então:

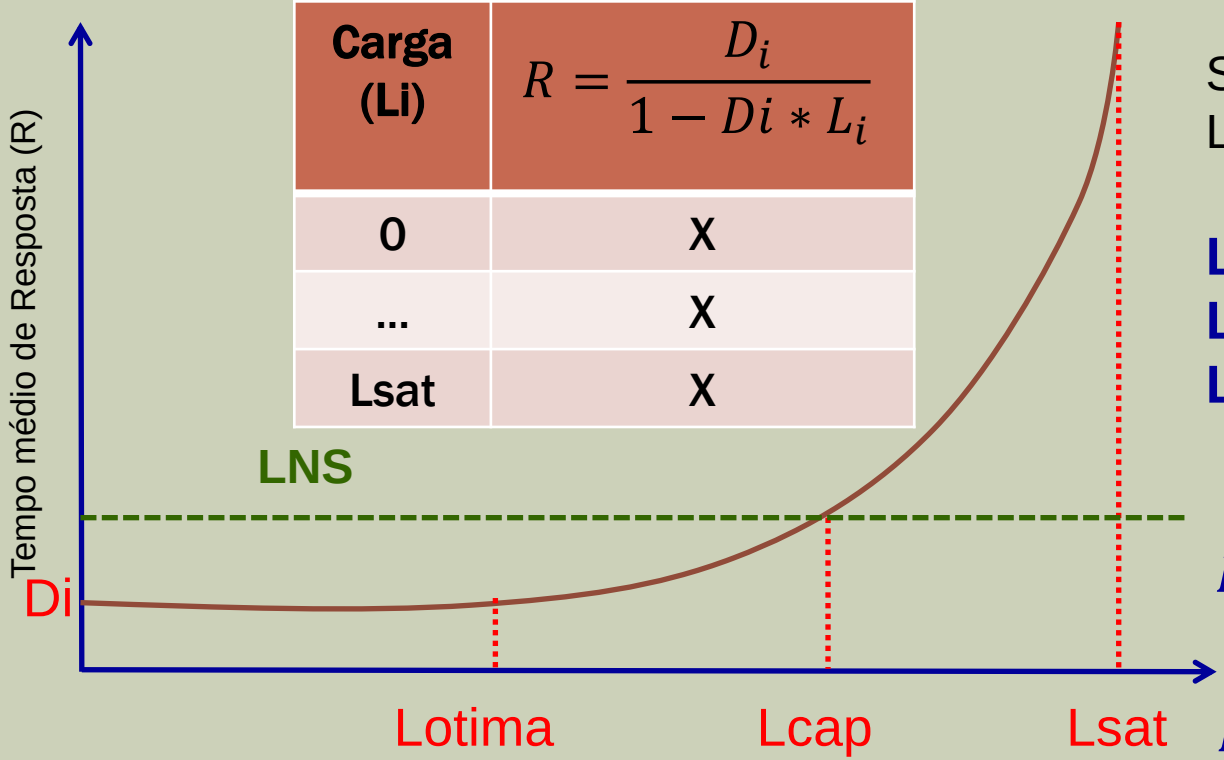
$$1{,}05 * Di = \frac{D_i}{1 - Di * Lotima}$$

$$\boldsymbol{Lotima = \frac{0{,}05}{1{,}05 * Di}}$$

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Exemplo de construção da Curva Típica de Desempenho:

Carga (Li)	$R = \frac{D_i}{1 - Di * L_i}$
0	X
...	X
Lsat	X



Supondo Di= 500 ms/req.
LNS = 800 ms/req.

Lsat = 1/Di
Lsat = 1/0,5 = 2 req/s.
Lsat = 120 req/min

$$Lcap = \frac{LNS - D_i}{LNS * Di}$$

$$Lcap = \frac{0,8 - 0,5}{0,8 * 0,5}$$

$$Lcap = 0,75\ req/s$$

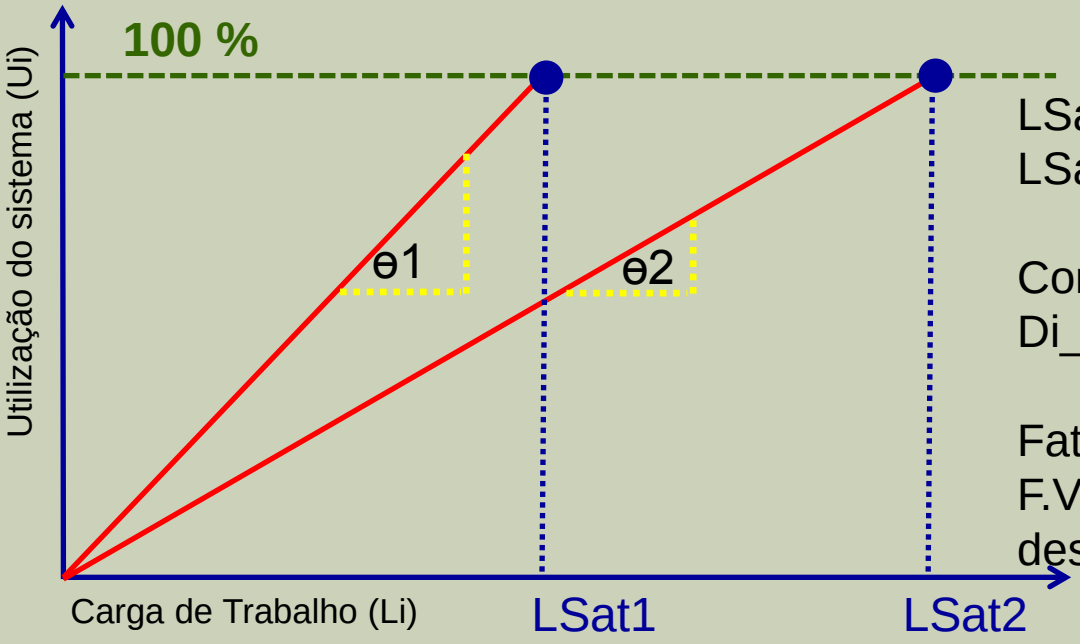
$$Lcap = 45\ req/min$$

$$Lotima = \frac{0,05}{1,05 * 0,5} = 0,095 req/s$$

$$Lotima = 5,7\ req/min$$

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Caso 1: Determinar o novo dispositivo para LSat2=150 req/min. (2,5 req/s). Considere Di_atual=500 ms/req



LSat1=1/Di_atual = 1/0,5 = 2 req/s
LSat1 = 120 req/min

Como: LSat=1/Di
Di_novo= 1/LSat2 = 1/2,5 = 0,4 req/s

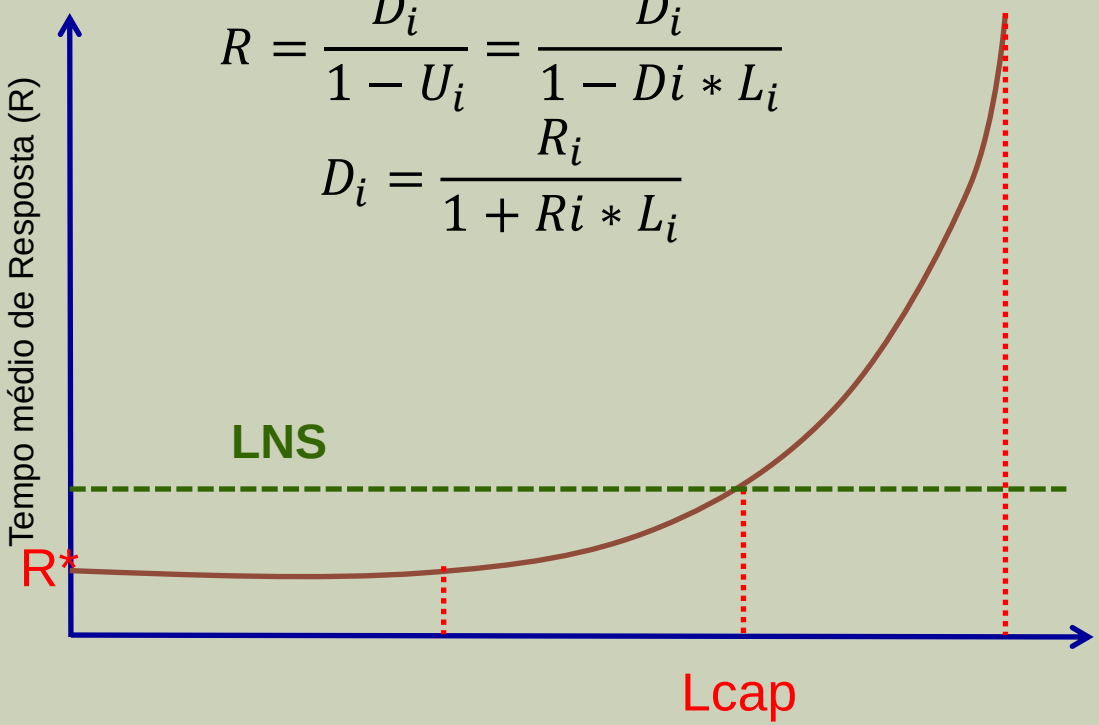
Fator Veloc. = Di_atual/Di_novo
F.V. = 0,5/0,4 = 1,25 (25% + desempenho)

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Caso 2: Determinar o novo dispositivo para R_SLA=2 s/req. L_Cap=100 r/min (1,6 r/s). Considere Di_atual=500 ms/req

$$R = \frac{D_i}{1 - U_i} = \frac{D_i}{1 - Di * L_i}$$

$$D_i = \frac{R_i}{1 + Ri * L_i}$$



Di = R/(1+Ri*Li)

Di_novo = 2/(1+2*1,6)

Di_novo = 0,47 s/req

F.V. = Di_atual/Di_novo
F.V. = 0,5/0,47 = 1,06 (6%)

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Caso 3: Para uma disponibilidade de 20%, qual seria o LNS aceitável?

Ui = 80%
Di = 500 ms/req.

$$Ui = Di * Li$$

$$Li = \frac{0,8}{0,5} = 1,6\ req/s$$

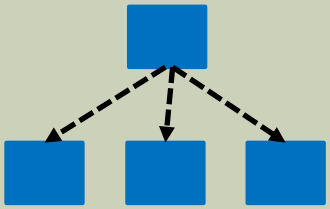
$$Lcap = Li$$

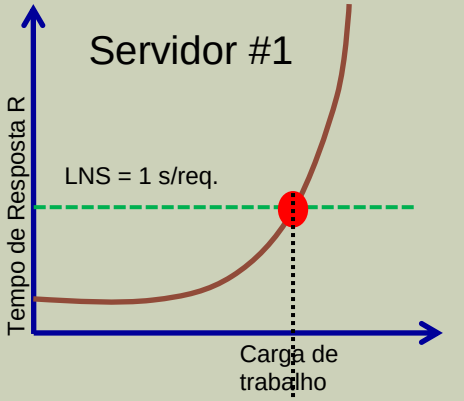
$$Ri = \frac{Di}{1 - Di * Li} = \frac{0,5}{1 - 0,5 * 1,6} = \frac{0,5}{0,2} = 2,5\ req/s$$

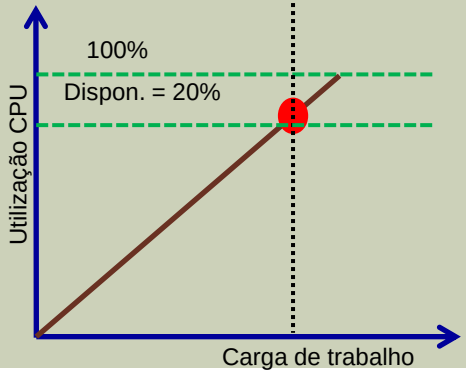
$$Ri = LNS = SLA = 2,5\ req/s$$

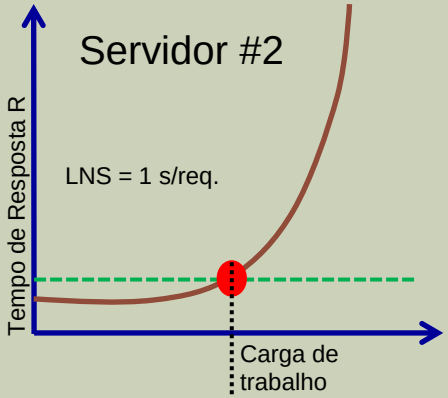
CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

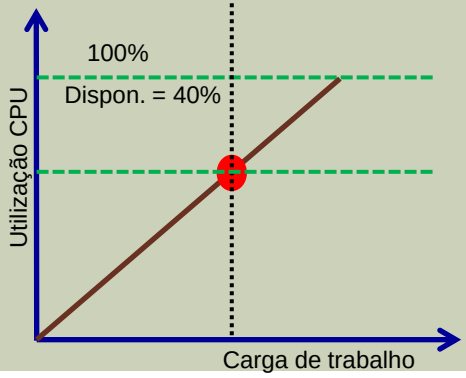
Exemplo de balanceamento de carga para cluster de servidores.

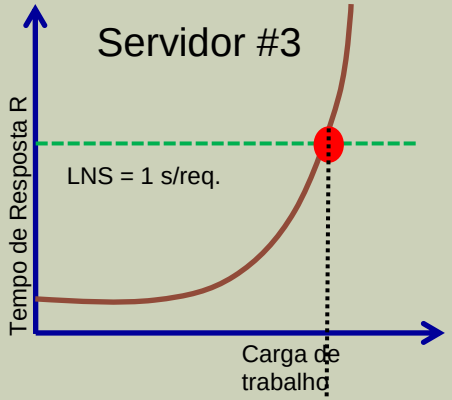


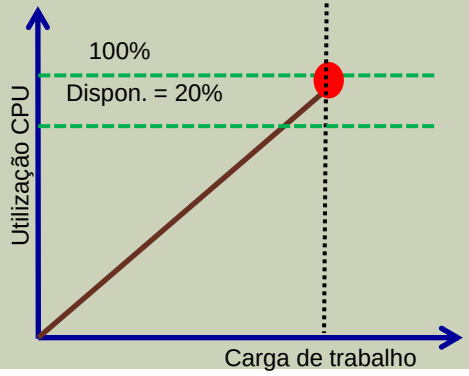










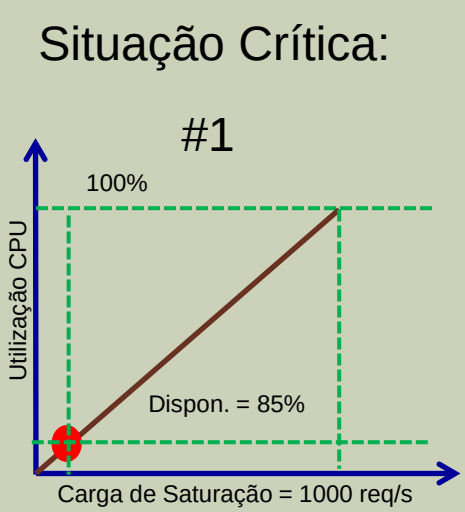


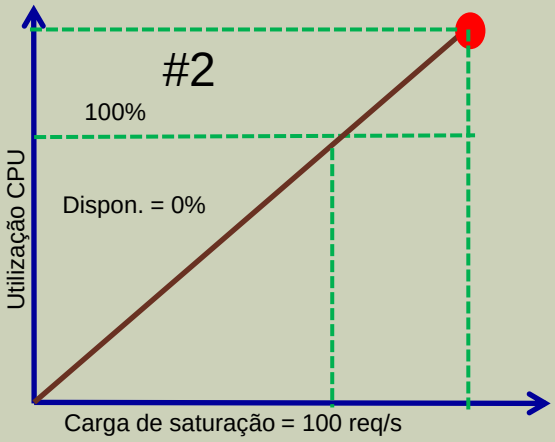
CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

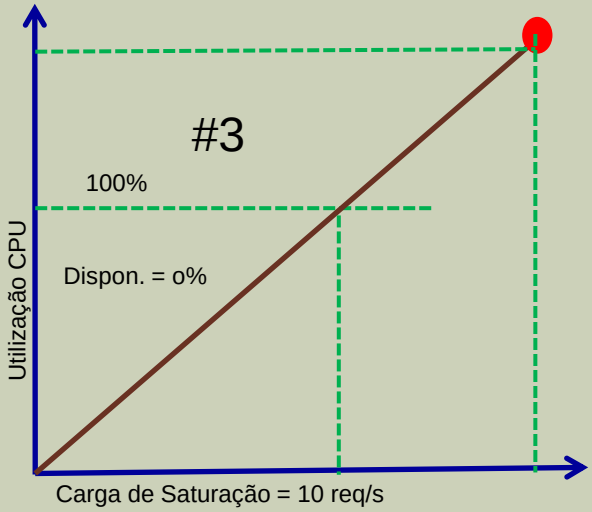
Exercício 1:

Considere três servidores com os seguintes tempos médios de serviço por processo: D1=1 ms; D2=10 ms; D3=100 ms. Considere também um único tipo de requisição. Mostre graficamente a relação para cada servidor de: (carga *vs*. Utilização). Utilize como valor de referência, para construir os gráficos, a carga de saturação. Se a carga atual para cada servidor for 150 req/s, qual ou quais servidores encontram-se disponíveis? E qual é o valor das disponibilidades.

Situação Crítica:



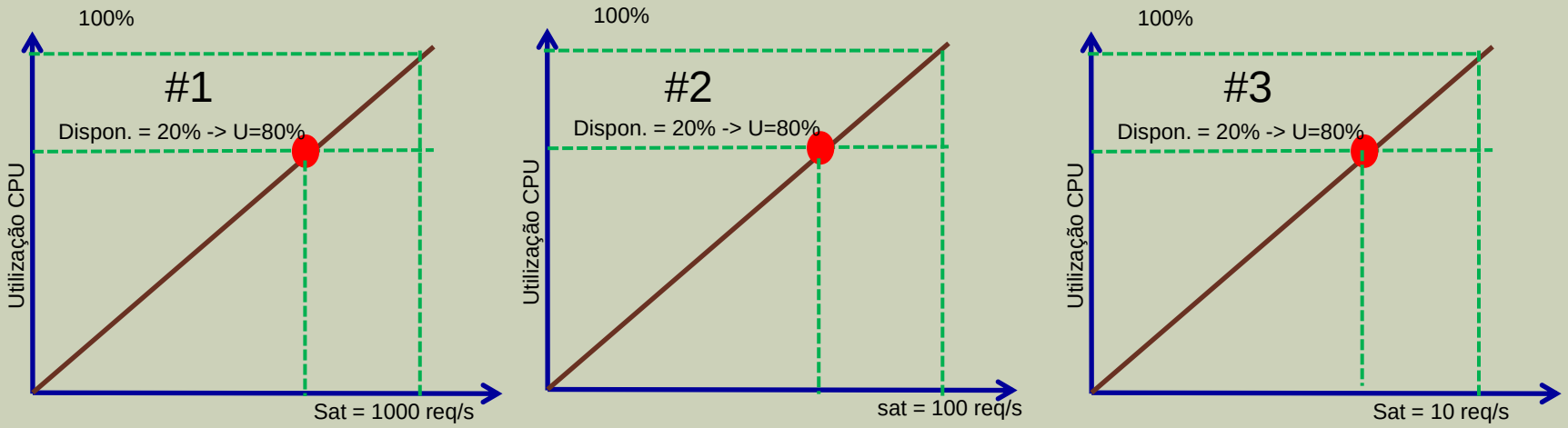




CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Exercício 1 (Continuação):

Considere a situação crítica do cenário anterior, Calcular a carga de trabalho máxima que deve ser enviada a cada servidor para garantir minimamente 20 % de disponibilidade.



Ui=Di*Li L1=0,80/0,001=800 r/s L2=0,80/0,01=80 r/s L3=0,80/0,1=8 r/s

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Exercício 2:

Considere a estrutura de servidores de aplicação mostrada na figura anterior onde cada servidor pode ser modelado através da teoria das filas. Assuma que cada um dos servidores processa requisições distintas, mas com tempos de processamento muito homogêneos. Os dados abaixo mostram os tempos médios de serviço, e o valor de carga atual para cada servidor.

Serv. Aplic #1	Di=100 ms/req	Carga atual=300 req/min
Serv. Aplic #2	Di=120 ms/req	Carga atual=280 req/min
Serv. Aplic #3	Di=80 ms/req	Carga atual=350 req/min

Considerando a carga atual de cada servidor. Qual servidor encontra-se mais disponível para aceitar uma nova requisição?**.**

Se o limite do nível de serviço é de 20% de disponibilidade para cada servidor. Qual servidor possue a maior capacidade de processamento?.

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Exercício 2 (continuação):

Serv. Aplic #1	Di=100 ms/req	Carga atual=300 req/min
Serv. Aplic #2	Di=120 ms/req	Carga atual=280 req/min
Serv. Aplic #3	Di=80 ms/req	Carga atual=350 req/min

Considerando a carga atual de cada servidor. Qual servidor encontra-se mais disponível para resever uma nova requisição?**.**

Ui=Di*Li R = Di/(1-Ui)

U1=0,1 * (300/60)= 0,50 Disp. = 50%
R1=0,1/0,5=0,2 s/req

U2=0,120 * (280/60)= 0,56 Disp. = 44%
R2=0,12/0,44=0,28 s/req.

U3=0,08 * (350/60)= 0,467 Disp. = 53,3%
R3=0,08/0,533=0,15 s/req

O Servidor #3 está mais disponível. Porém atenção deve ser dada ao desempenho do servidor, pois pode tratar-se de um servidor heterogêneo com baixa performance isoladamente.

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Exercício 2 (continuação):

Serv. Aplic #1	Di=100 ms/req	Carga atual=300 req/min
Serv. Aplic #2	Di=120 ms/req	Carga atual=280 req/min
Serv. Aplic #3	Di=80 ms/req	Carga atual=350 req/min

Considerando a carga atual de cada servidor. Qual servidor encontra-se mais disponível para resever uma nova requisição?**.**

Se o limite do nível de serviço é de 20% de disponibilidade para cada servidor. Qual servidor possue a maior capacidade de processamento?.

CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Exercício 3:

Considere que o modelo de previsão da carga de trabalho de um sistema computacional (em req/s) é dado pela equação: .L=5*t+10. O modelo de carga foi construído levando em conta os últimos 5 meses. Considerando que o tempo médio de serviço da aplicação é de 5 ms/req, e o LNS = 8 ms/req. Calcule os meses de vida útil que restam ao sistema e a utilização do sistema prevista para o próximo mês. Responda qual das opções a seguir é a correta:

- O sistema está dentro da sua vida útil e a utilização está abaixo de 50%
- O sistema está dentro da sua vida útil porém a utilização está acima do 50%
- O sistema já não se encontra mais na sua vida útil porém sua utilização atual é de 20%
- O sistema já não se encontra mais na sua vida útil e sua utilização atual é de 100%