



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**

Professor: Luis Enrique Zárate

# 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)

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 também de SLA.

Ex.      Tempo de resposta  $LNS\_R = 800 \text{ ms/req.}$   
            Utilização do processador  $LNS\_U = 80\%$

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Definições preliminares:

### 5) Ponto de Crítico do Sistema ( $P_c$ )

Corresponde ao ponto de interseção da curva de desempenho com a reta do LNS.  $P_c = (L_c, LNS)$

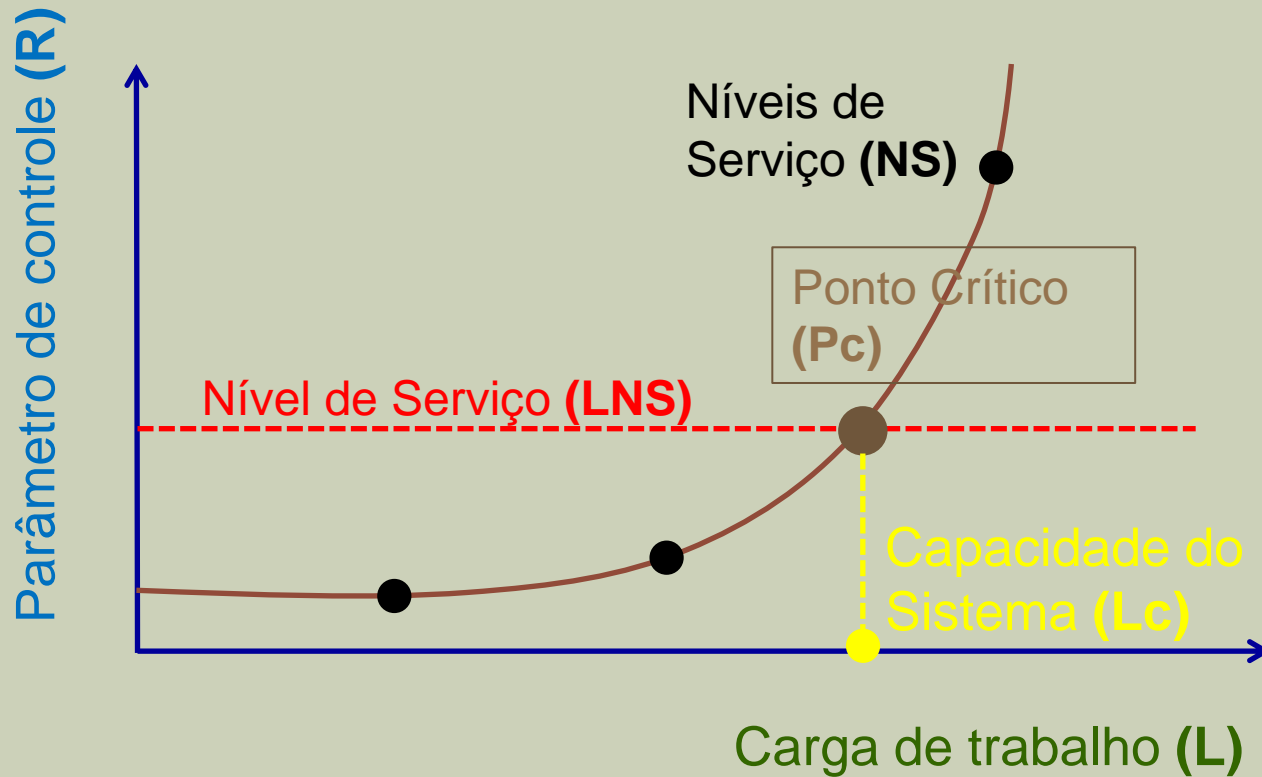
### 6) Capacidade de um Sistema Computacional ( $L_c$ )

Corresponde ao valor de carga ( $L_c$ ) que leva ao sistema ao ponto crítico ( $P_c$ ) 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)

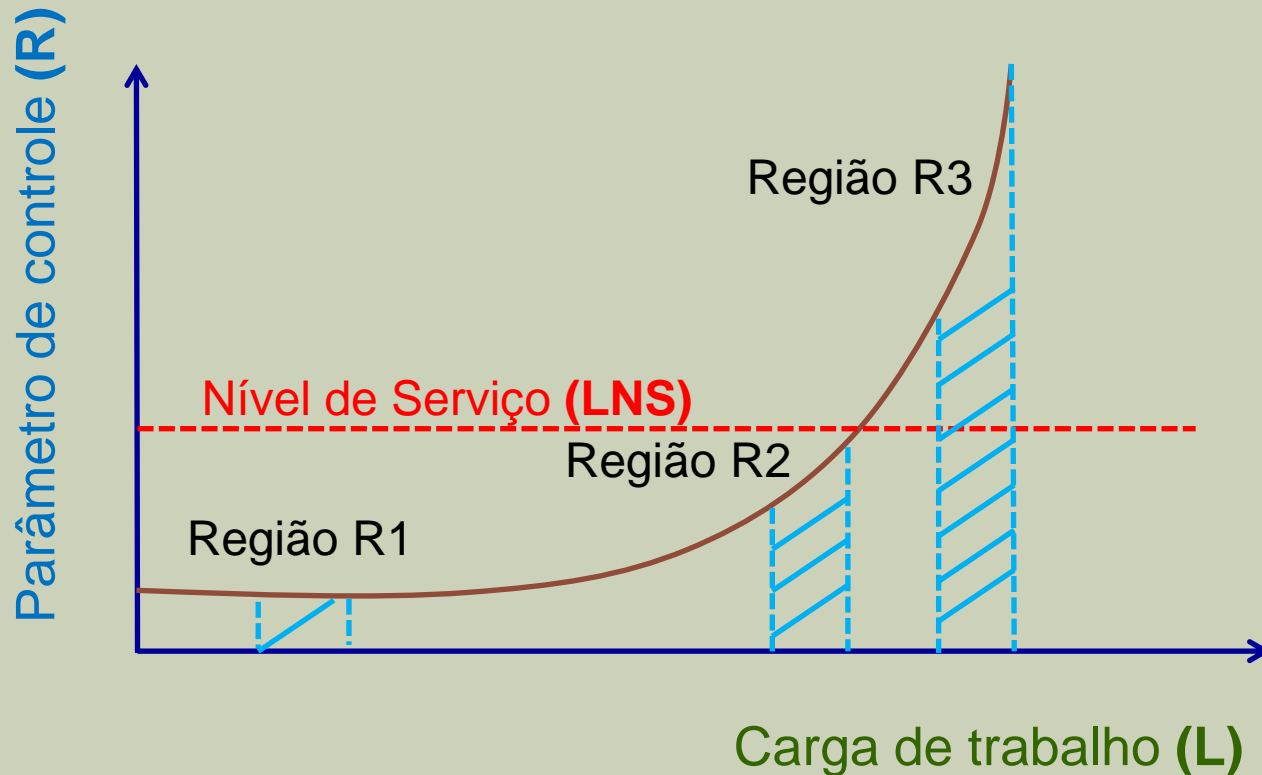
Nesta região, a proporção de variação na carga de trabalho causam variações proporcionais no valor do Tempo de Resposta (R). Tipicamente esta região é caracterizada por níveis de utilização de 30 a 40%.

### 10) Região de Comportamento Não-Linear (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%.

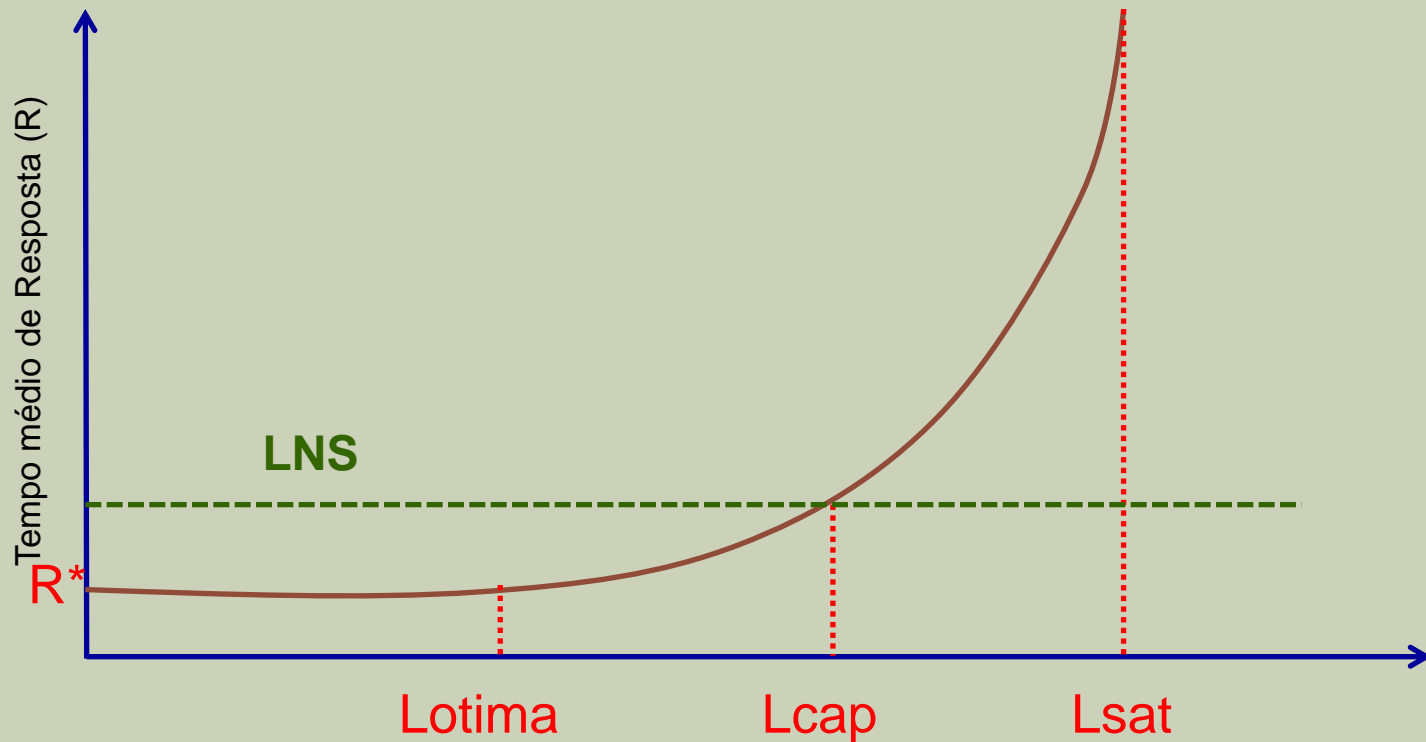
# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

Regiões Típicas da Curva de Desempenho:



# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

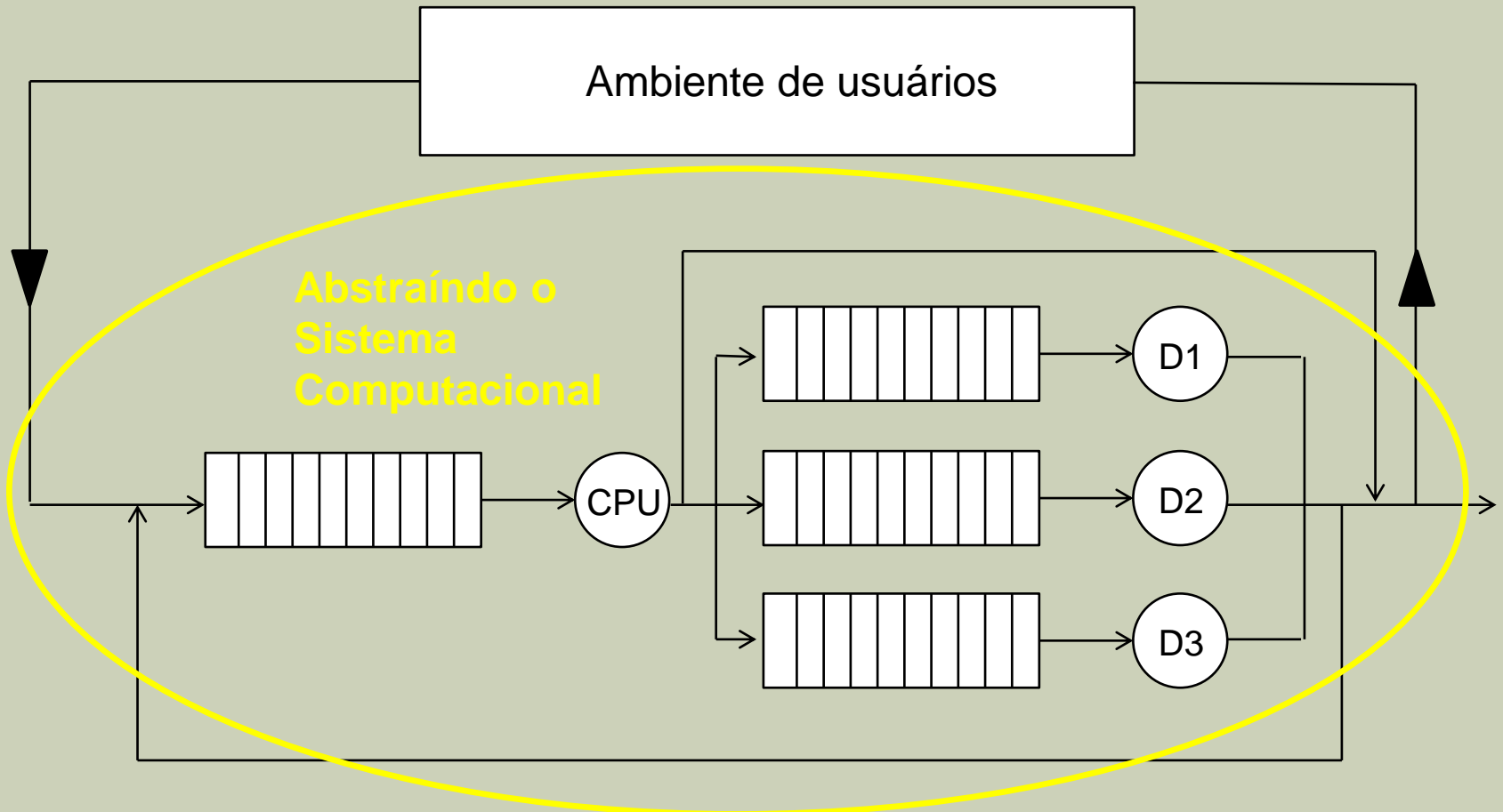
## Parâmetros de Referencia da Curva de Desempenho:





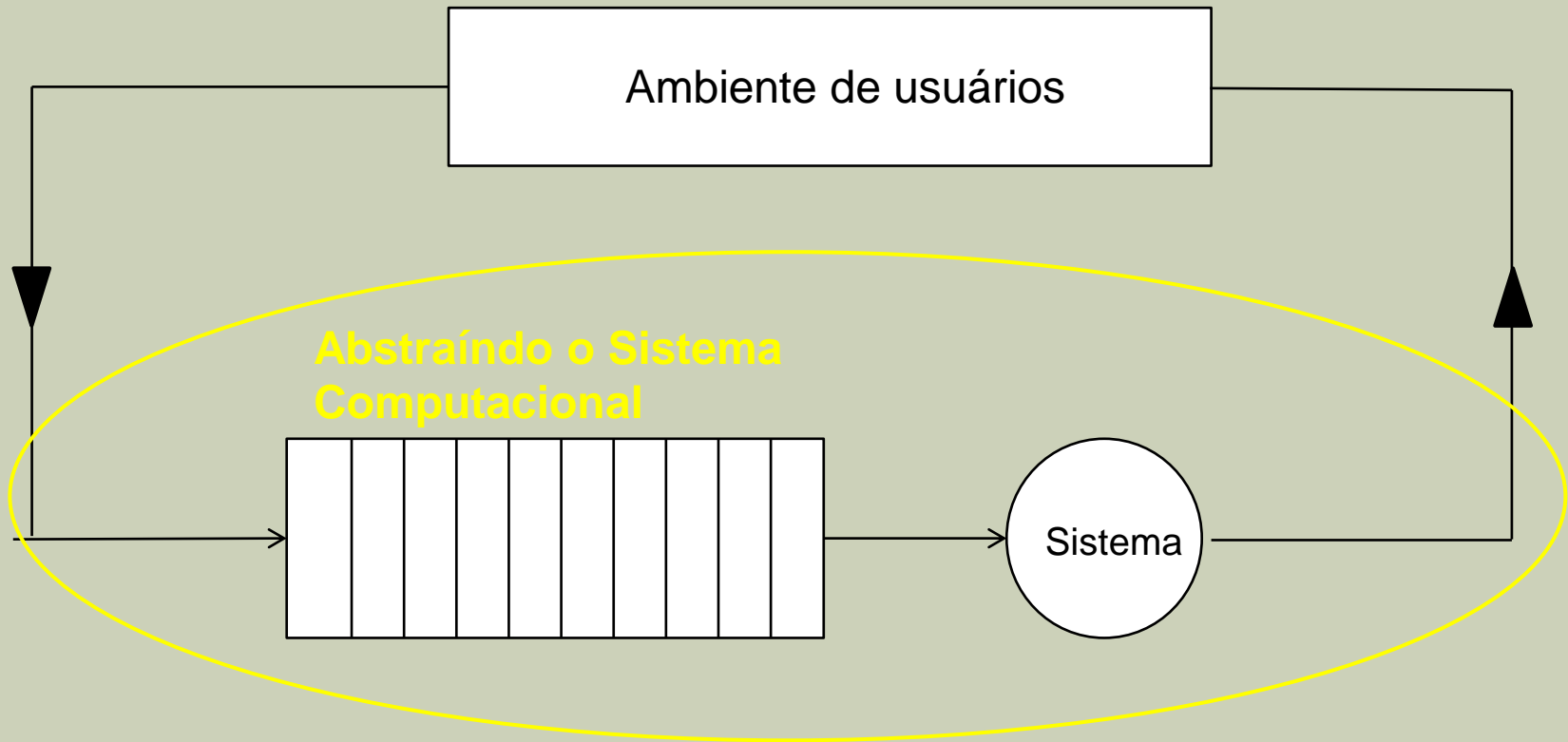
# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## 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:

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

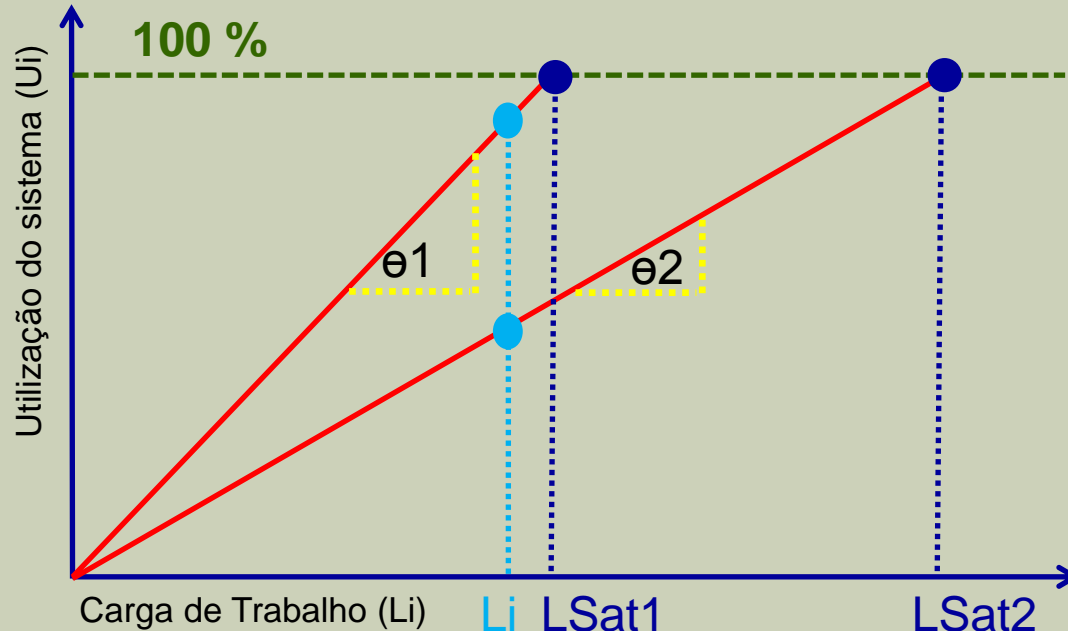
**R:** Tempo médio de resposta por requisição [s/req.]

**U<sub>i</sub>:** Utilização do dispositivo “i”

**D<sub>i</sub>:** 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):**

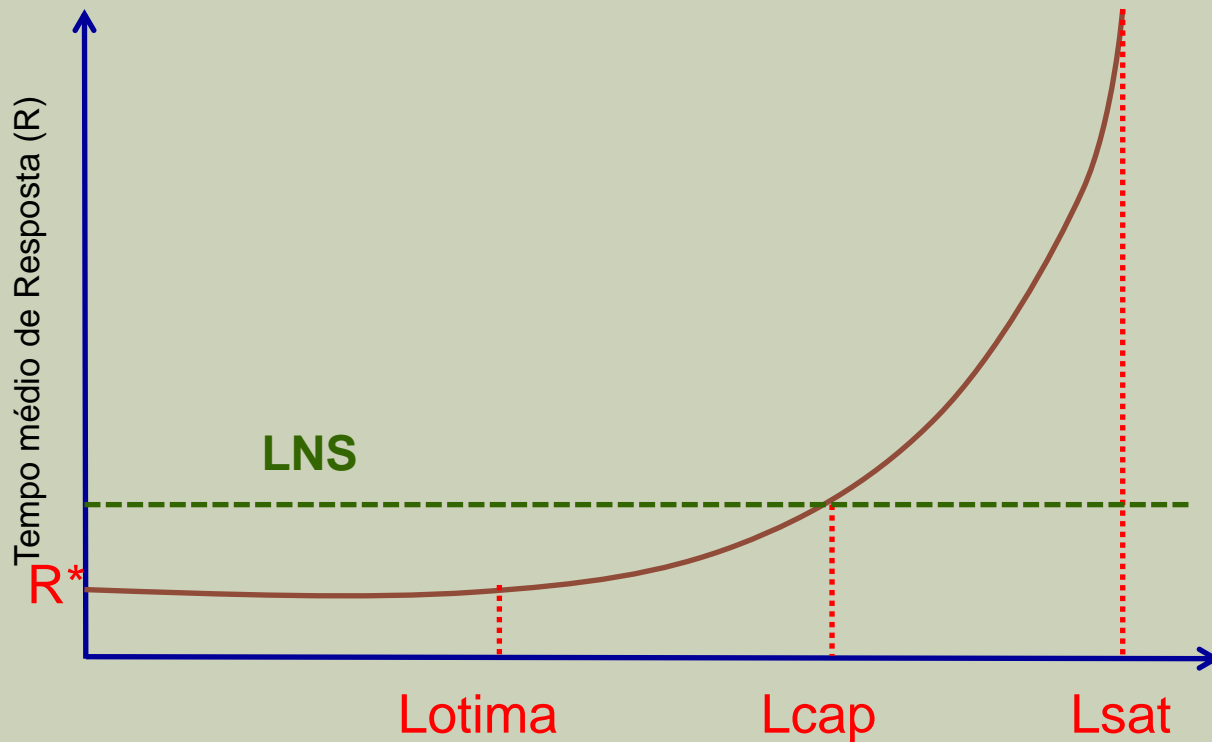


$$U_i = T_g(\theta) * L_i$$

$$U_i = D_i * L_i$$

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

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



Quando:

$L_i \rightarrow 0$

$U_i \rightarrow 0$

Dado que:

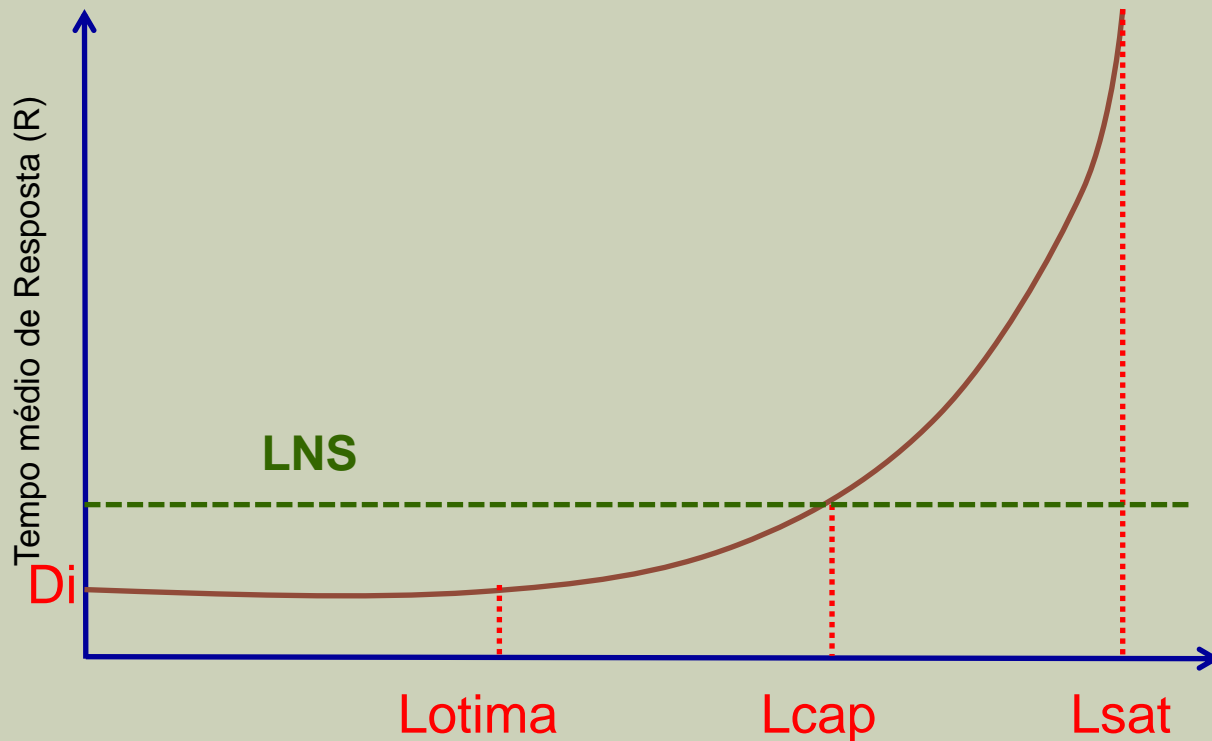
$$R = \frac{D_i}{1 - U_i}$$

Logo:

$$R^* = D_i$$

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Determinando a Carga de Saturação (Lsat):



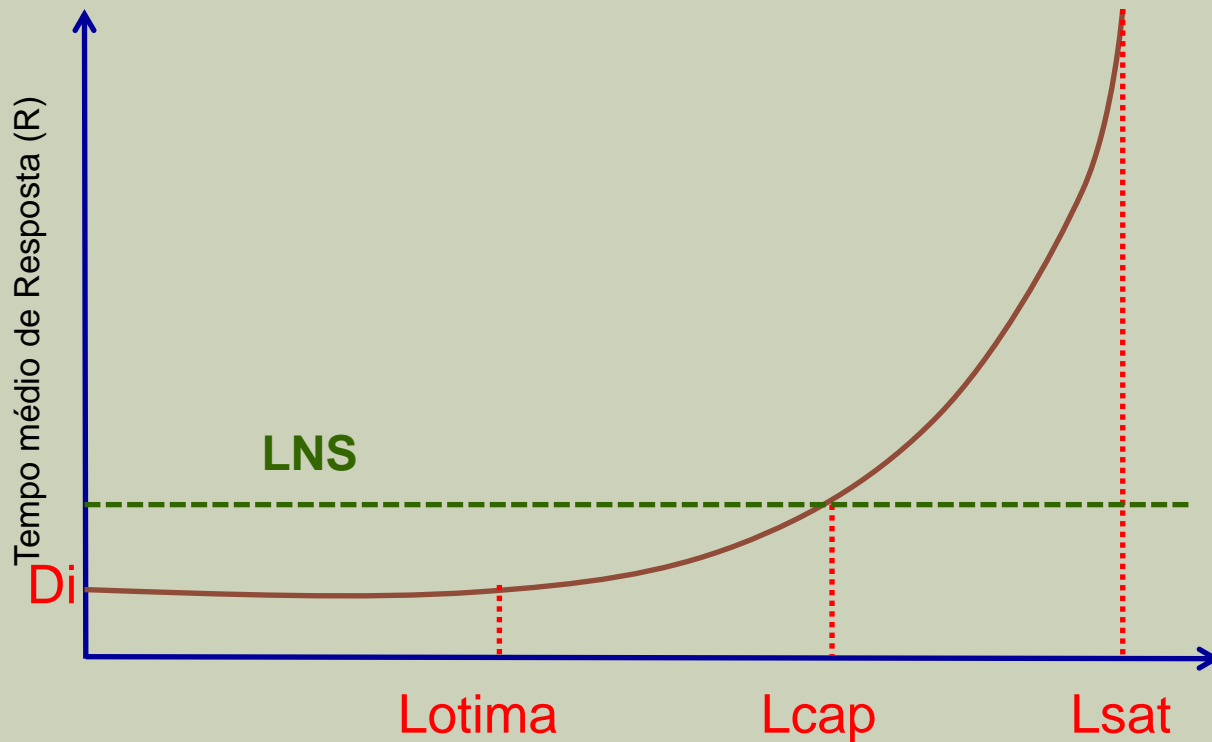
Quando:  
 $L_i \rightarrow L_{sat}$   
 $U_i \rightarrow 1$

Dado que:  
 $U_i = D_i * L_i$   
 $1 = D_i * L_{sat}$

Logo:  
 **$L_{sat} = 1/D_i$**

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Determinando a Capacidade do Sistema (Lc):



Quando:  
 $L_i \rightarrow L_{cap}$   
 $R_i \rightarrow LNS$

Dado que:  
$$R = \frac{D_i}{1 - U_i}$$

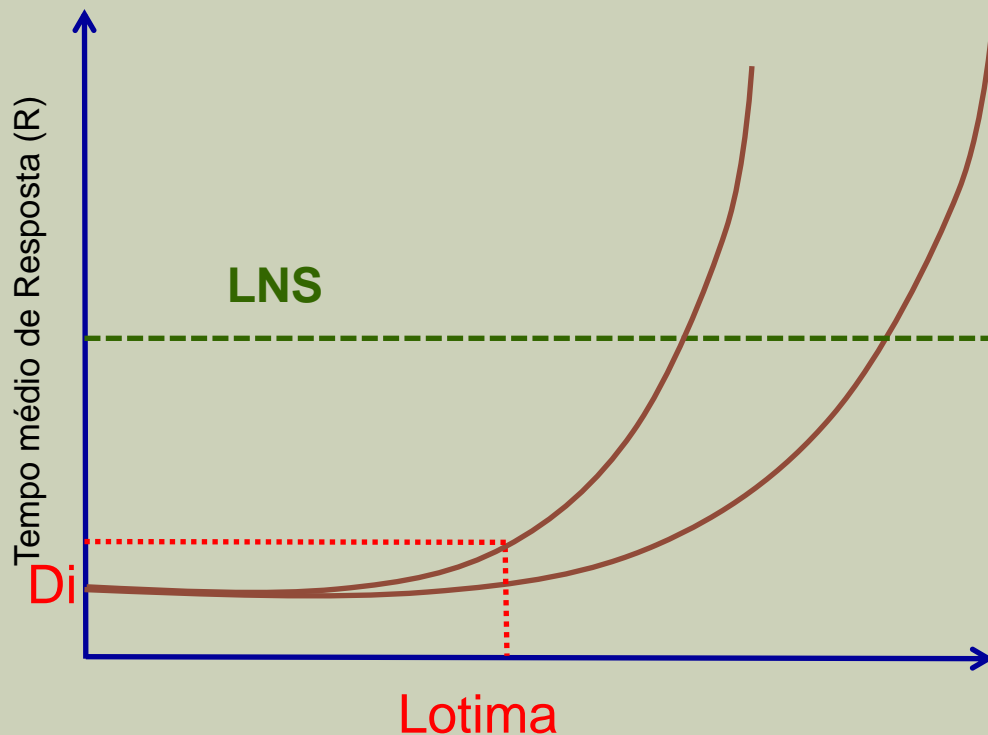
Então:

$$LNS = \frac{D_i}{1 - D_i * L_{cap}}$$

$$L_{cap} = \frac{LNS - D_i}{LNS * D_i}$$

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Determinando a Carga Ótima do Sistema (Lotima):



Quando:

$L_i \rightarrow Lotima$

$R_i \rightarrow 1,05 * D_i$

Dado que:

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

Então:

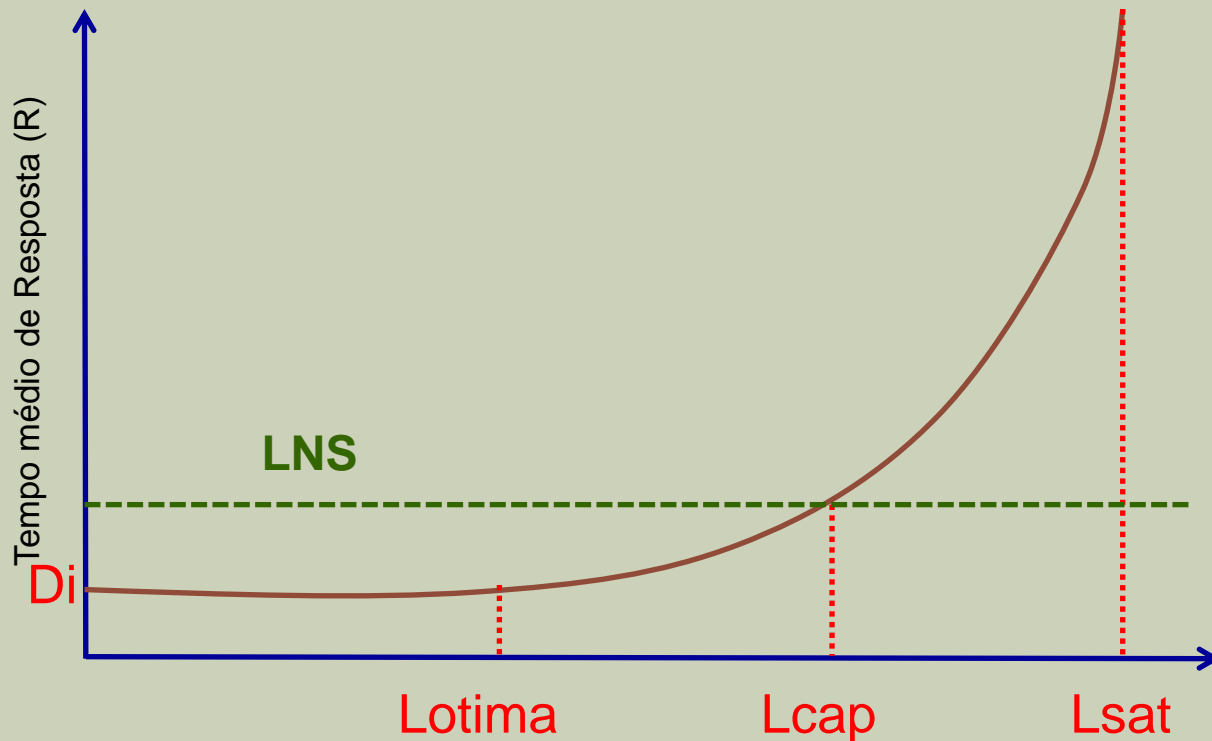
$$1,05 * D_i = \frac{D_i}{1 - D_i * Lotima}$$

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



# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Determinando a Carga de Saturação (Lsat):



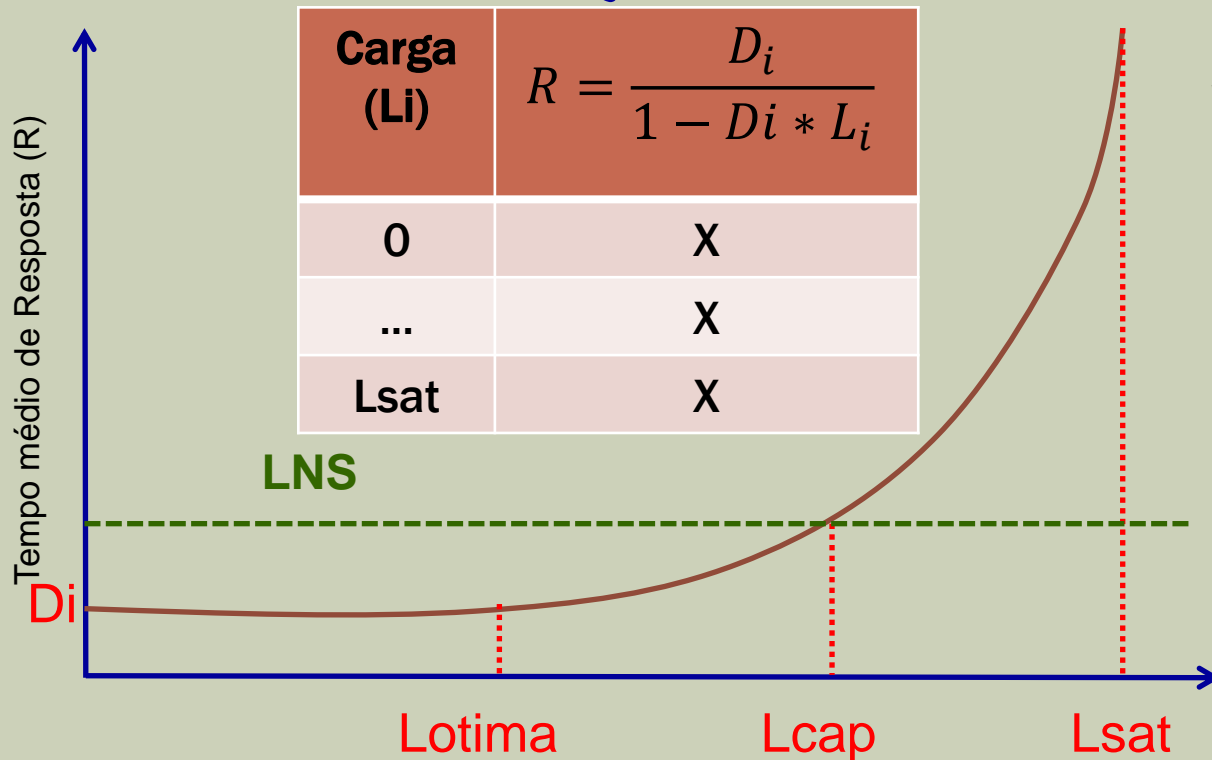
Quando:  
 $L_i \rightarrow L_{sat}$   
 $U_i \rightarrow 1$

Dado que:  
 $U_i = D_i * L_i$   
 $1 = D_i * L_{sat}$

Logo:  
 **$L_{sat} = 1/D_i$**

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Exemplo de construção da Curva Típica de Desempenho:



Supondo  $D_i = 500$  ms/req.  
 $LNS = 800$  ms/req.

$$Lsat = 1/D_i$$

$$Lsat = 1/0,5 = 2 \text{ req/s.}$$

$$Lsat = 120 \text{ req/min}$$

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

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

$$Lcap = 0,75 \text{ req/s}$$

$$Lcap = 45 \text{ req/min}$$

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

$$Lotima = 5,7 \text{ req/min}$$

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

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

$U_i = 80\%$

$D_i = 500 \text{ ms/req.}$

$$U_i = D_i * L_i$$

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

$$L_{cap} = L_i$$

$$R_i = \frac{D_i}{1 - D_i * L_i} = \frac{0,5}{1 - 0,5 * 1,6} = \frac{0,5}{0,2} = 2,5 \text{ req/s}$$

$$R_i = LNS = 2,5 \text{ req/s}$$

# CURVA TEÓRICA DE DESEMPENHO DE SISTEMAS COMPUTACIONAIS

## Exemplo de balanceamento de carga para cluster de servidores.

