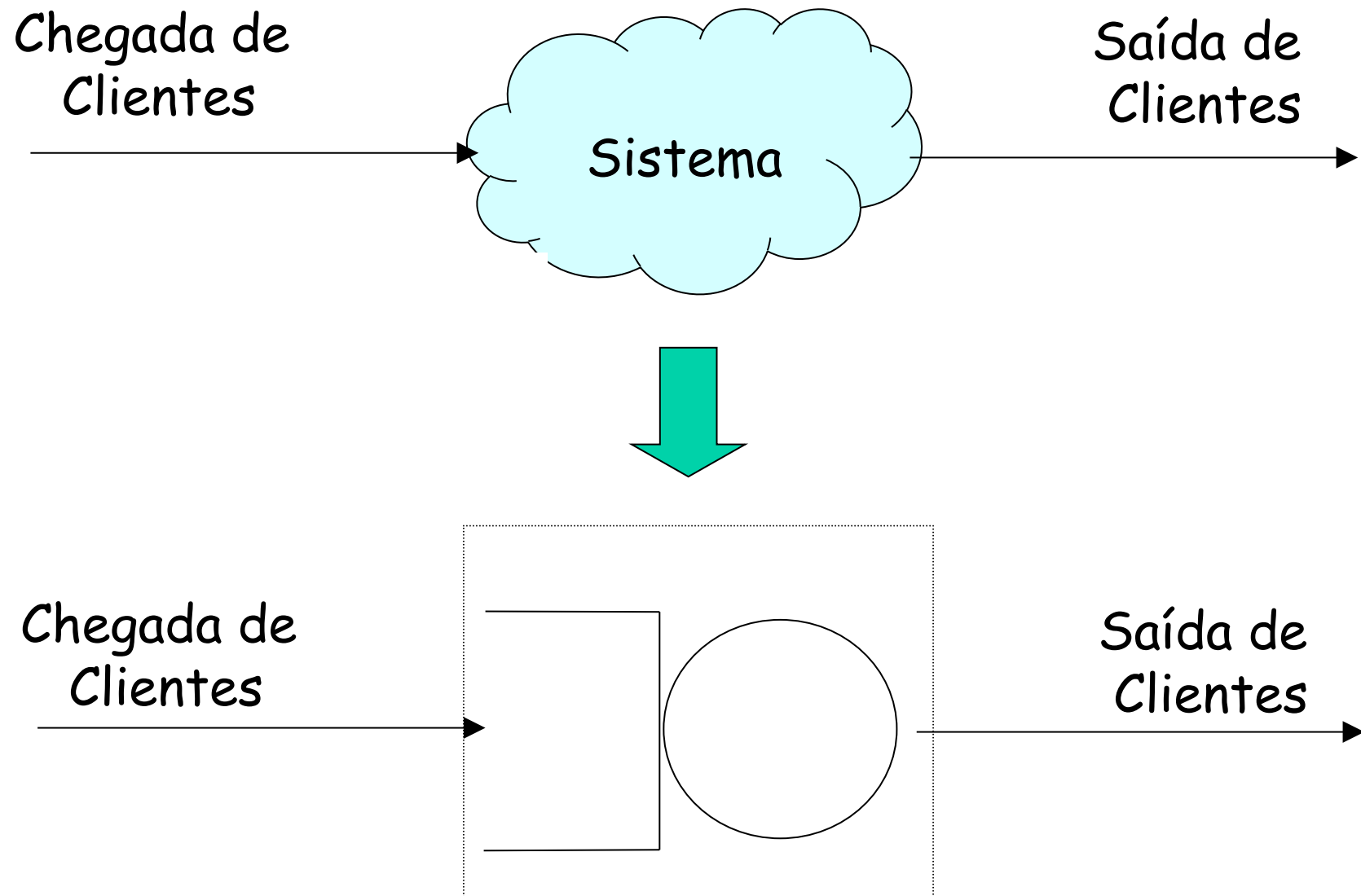


Leis Fundamentais

Profa. Jussara M. Almeida
1º Semestre de 2011

Sistema Abstrato



Quantidades Básicas

- **T**: duração da medição
(intervalo de tempo durante o qual o sistema foi observado)
- **A**: número de chegadas de *clientes* observadas em T
- **C**: número de saídas/términos de clientes observadas em T
- **B**: quantidade de tempo sistema ficou ocupado processando

Entradas dos Modelos de Filas

- Taxa de Chegadas λ

$$\lambda \equiv A/T$$

- Throughput X

$$X \equiv C/T$$

- Utilização U

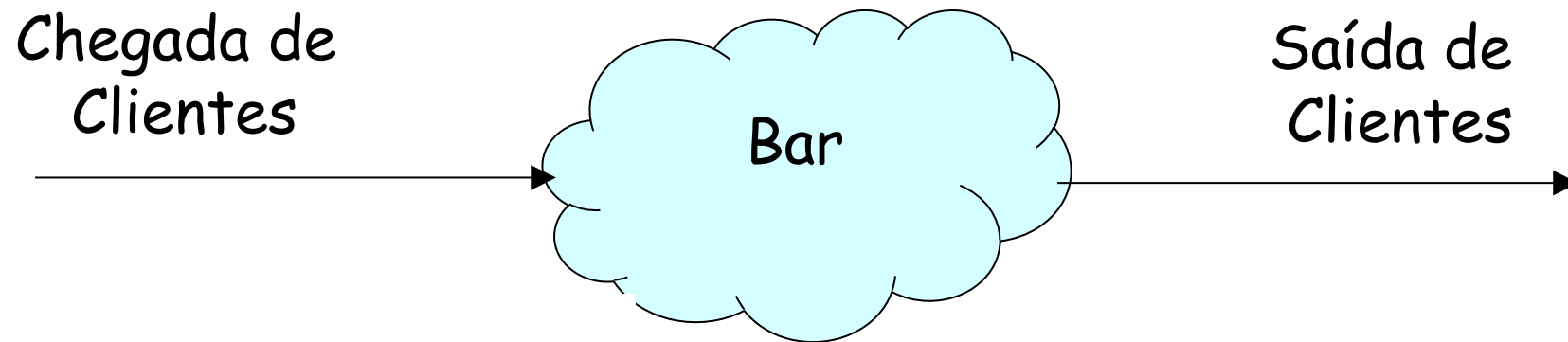
$$U \equiv B/T$$

- Tempo de Serviço Médio por Cliente S

$$S \equiv B/C$$

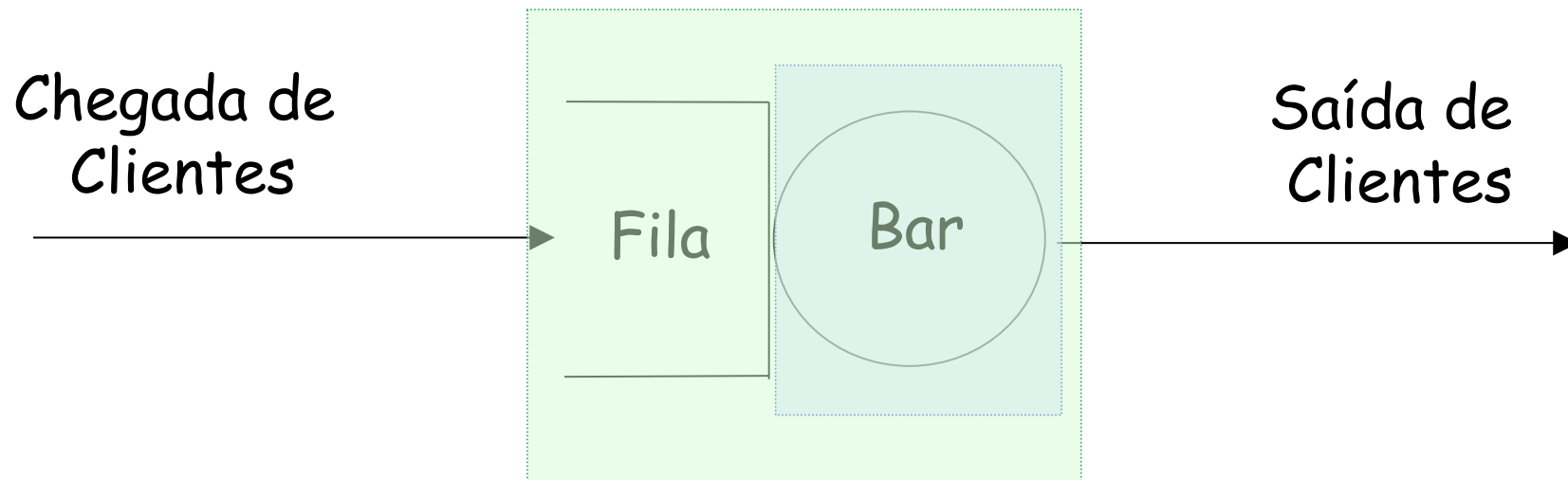
Lei de Little

- Qual o número médio de clientes no sistema?



Lei de Little

- Qual o número médio de clientes no sistema?



Qual o escopo do sistema?

Opção 1: As instalações internas do bar.

Opção 2: As instalações internas + externas (fila)

Lei de Little: Definições

- W = Tempo acumulado no sistema

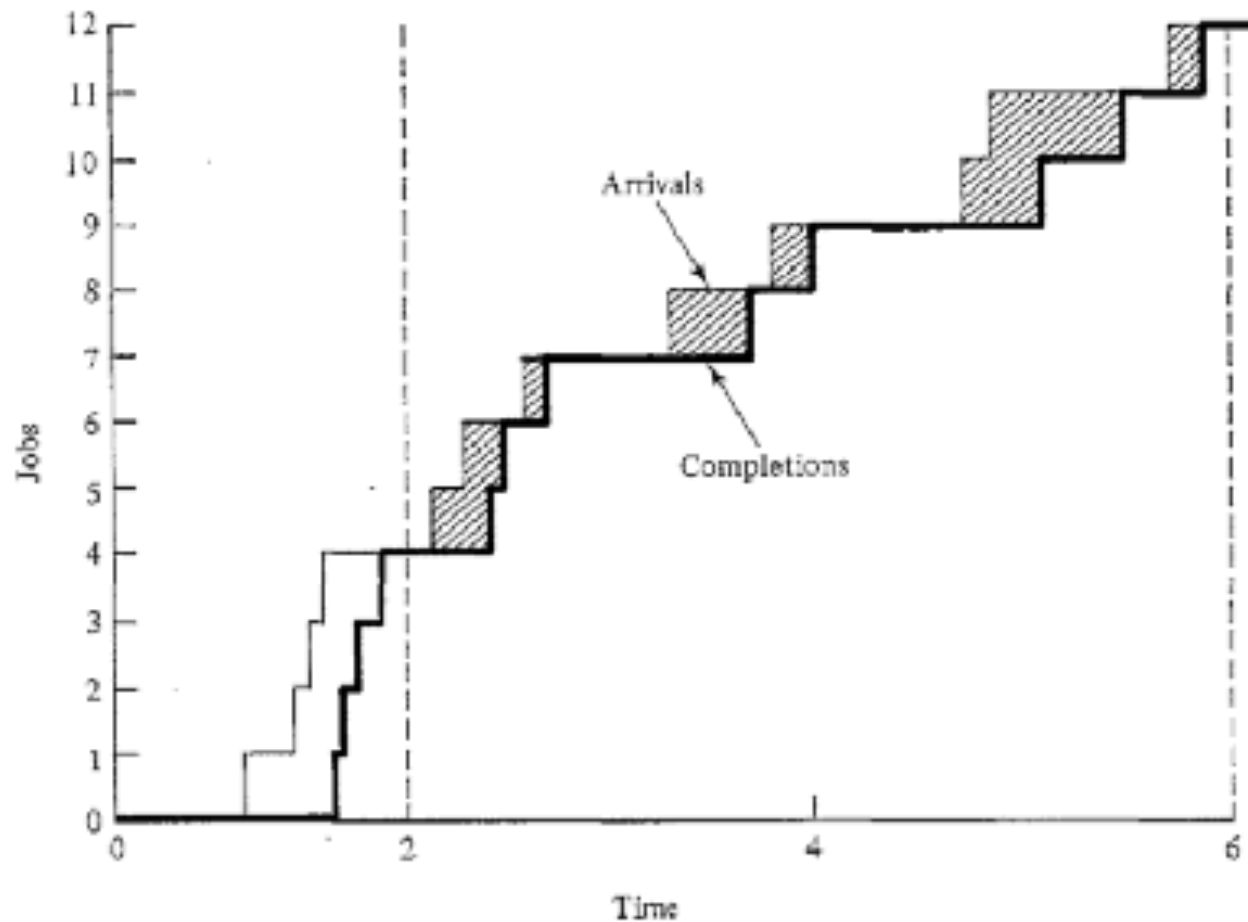


Figure 3.2 – System Arrivals and Completions

Lei de Little: Definições

- N : número médio de clientes no sistema
- Como calcular N ?

$$N \equiv W/T$$

- R : tempo médio de residência no sistema para cada cliente

$$R \equiv W/C$$

- Algebraicamente:

$$\frac{W}{T} = \frac{C}{T} \times \frac{W}{C}$$

Lei de Little

$$N = XR$$

Importância:

1. Amplamente aplicável (validação de experimentos)
2. Conhecimento de duas medidas e necessidade da terceira
3. Central para algoritmos de avaliação de redes de filas que estudaremos a seguir

Aplicação da Lei de Little

- Aplicável em vários níveis de um sistema: um recurso, um subsistema ou o sistema como um todo
- Importante: consistência
 - Definições de população de clientes (N), throughput (X) e tempo de residência (R) devem ser compatíveis
- Exemplo: aplicação de Lei de Little em diferentes níveis do servidor de arquivos DCC
 - Assuma 1 CPU, 3 discos e carga balanceada (prob. de requisição ser enviada para um disco é $1/3$)

Exemplo

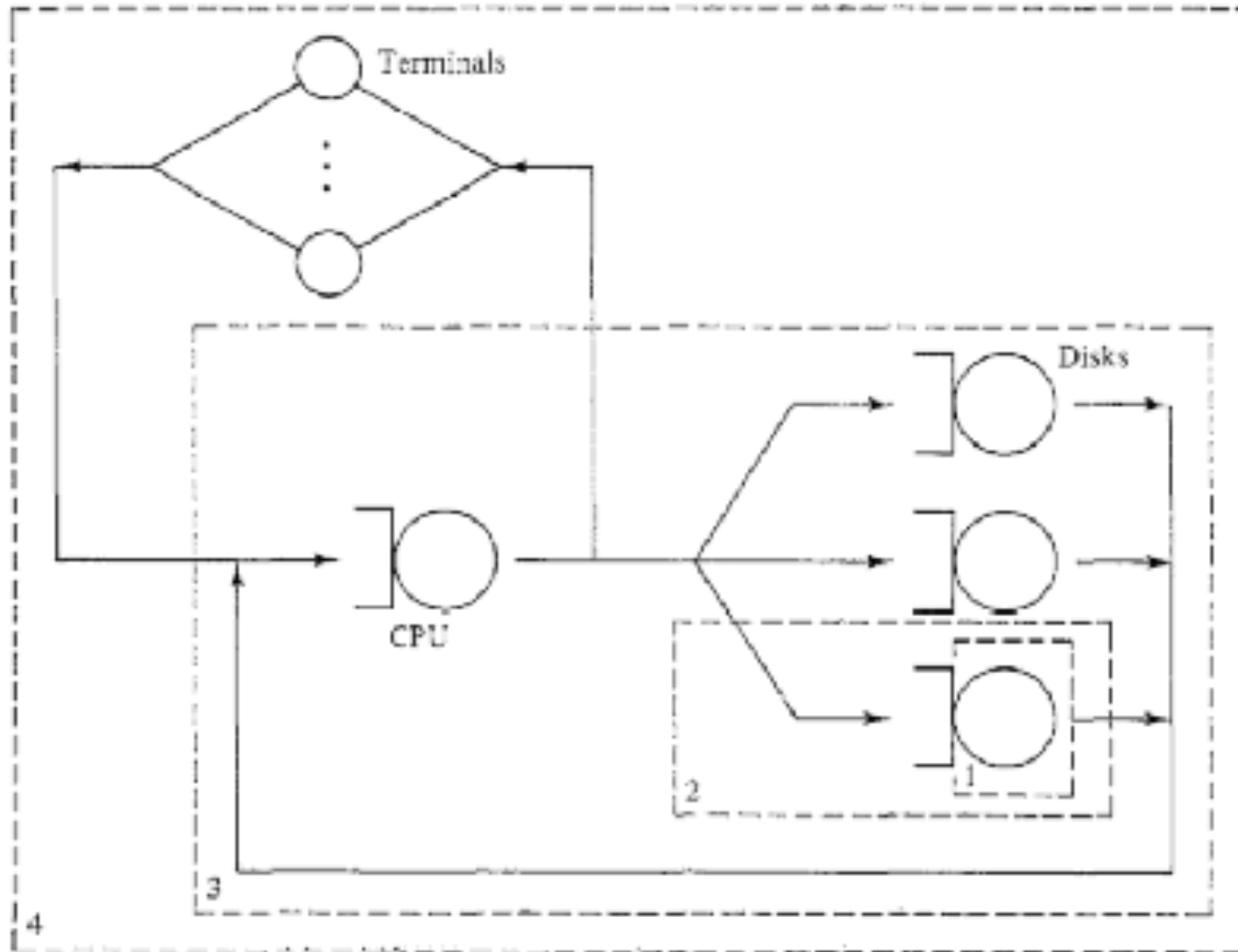


Figure 3.3 – Little's Law Applied at Four Levels

Aplicação da Lei de Little:

- Ex: Suponha que um disco sirva em média 40 requisições/seg e que uma requisição típica demande 0.0225 segundos para ser servida pelo disco. Qual a utilização do disco?

Exemplo: Servidor de Arquivos DCC

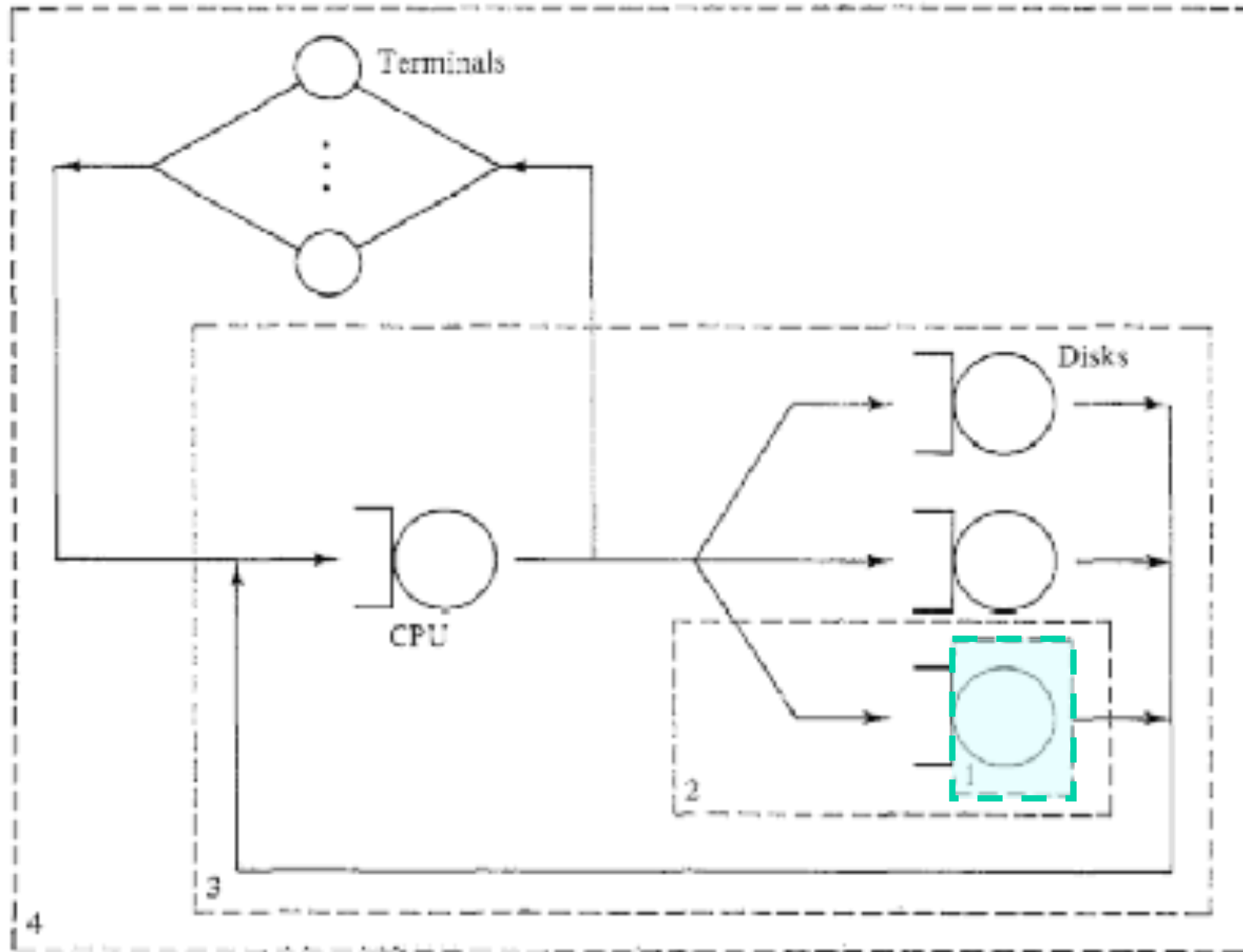


Figure 3.3 – Little's Law Applied at Four Levels

Aplicação da Lei de Little: Um Recurso sem a Fila

- Ex: Suponha que um disco sirva em média 40 requisições/seg e que uma requisição típica demande 0.0225 segundos para ser servida pelo disco. Qual a utilização do disco?
- População de clientes $N = U$ utilização do recurso (0 - 1)
- Tempo de residência $R = S$ requisito (tempo) de serviço médio por cliente (não inclui atraso na fila)

$$R = S = 0.0225 \quad X = 40$$

$$N = U = XS = 40 \times 0.0225 = 0.9$$

$$U = 90\%$$

Lei da Utilização

A utilização de um recurso é igual ao produto do throughput daquele recurso e o tempo médio de serviço no recurso

$$U \equiv XS = C/T \times B/C$$

(caso especial da Lei de Little)

Aplicação da Lei de Little:

- Ex: Suponha que para o mesmo disco do exemplo anterior, foi verificado que existem, em média 4 requisições para leitura pendentes. Qual o tempo que uma requisição permanece na fila de espera do disco? Qual o tamanho médio da fila de espera?

Exemplo: Servidor de Arquivos DCC

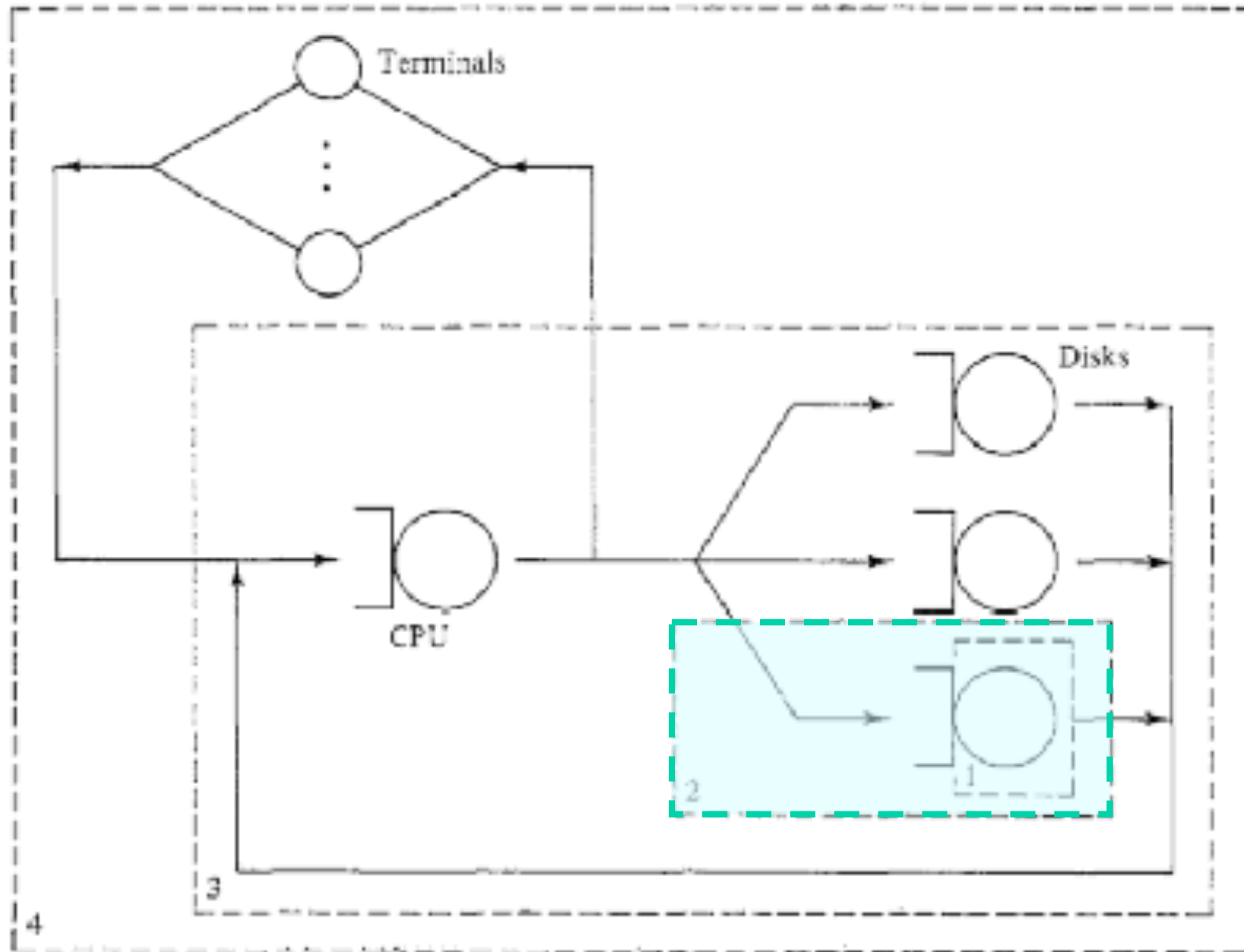


Figure 3.3 – Little's Law Applied at Four Levels

Aplicação da Lei de Little: Um Recurso com a Fila

- Ex: Suponha que para o mesmo disco do exemplo anterior, foi verificado que existem, em média 4 requisições para leitura pendentes. Qual o tempo que uma requisição permanece na fila de espera do disco? Qual o tamanho médio da fila de espera?
- N = requisições na fila e em serviço
- R = tempo médio que um cliente permanece no recurso por visita (tempo de fila + tempo de serviço)

$$N = 4 \quad X = 40$$

$$N = RX \quad R = N/X = 0.1 \text{ segundos} \quad (\text{Lei de Little})$$

$$\text{Tempo na fila: } R - 0.0225 = 0.1 - 0.0225 = 0.0775 \text{ seg}$$

$$\text{Tamanho da fila: } N - U = 4 - 0.9 = 3.1$$

Aplicação da Lei de Little:

- Ex: Suponha que o servidor de arquivos consiga processar em média 1 requisição a cada 2 segundos e que haja em média 7.5 usuários submetendo requisições simultaneamente. Qual o tempo de resposta médio observado por estes usuários?

Exemplo: Servidor de Arquivos DCC

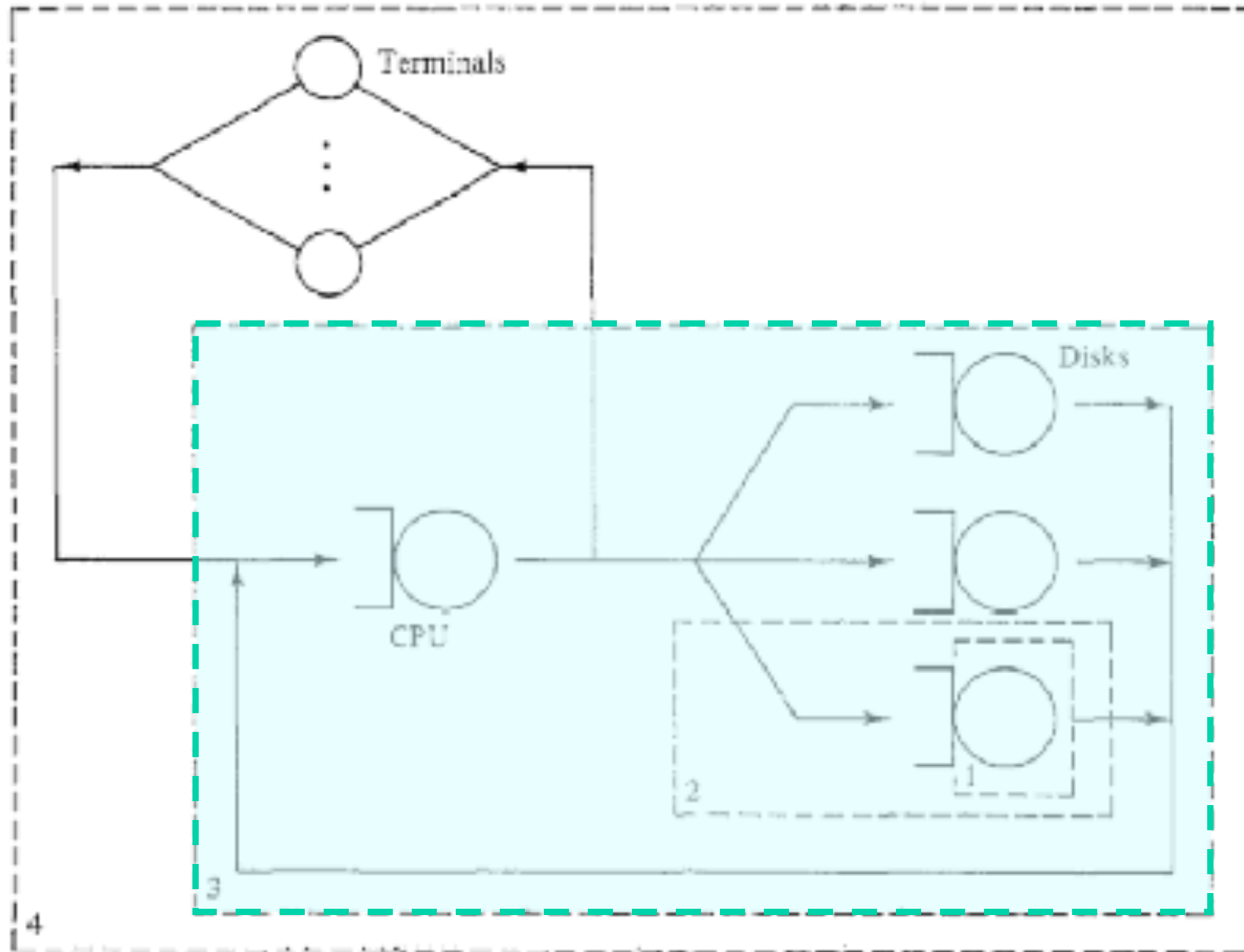


Figure 3.3 – Little's Law Applied at Four Levels

Aplicação da Lei de Little: Servidor sem Terminais

- Ex: Suponha que o servidor de arquivos consiga processar em média 1 requisição a cada 2 segundos e que haja em média 7.5 usuários submetendo requisições simultaneamente. Qual o tempo de resposta médio observado por estes usuários?
- N = interações a nível de sistema (realmente clientes)
- X = taxa de interações entre terminais e subsistema
- R = tempo de resposta

$$N = 7.5 \quad X = 1/2$$

$$N = RX \quad R = N/X = 15 \text{ segundos}$$

(Lei de Little)

Aplicação da Lei de Little: Servidor de arquivos DCC

- Ex: Suponha que 10 usuários utilizem o sistema. Estes usuários fazem processamento local por 5 segundos, em média, antes de submeterem requisições ao servidor central. O tempo de resposta médio observado por eles é de 15 segundos. Qual o throughput do sistema?

Exemplo: Sistema Timesharing

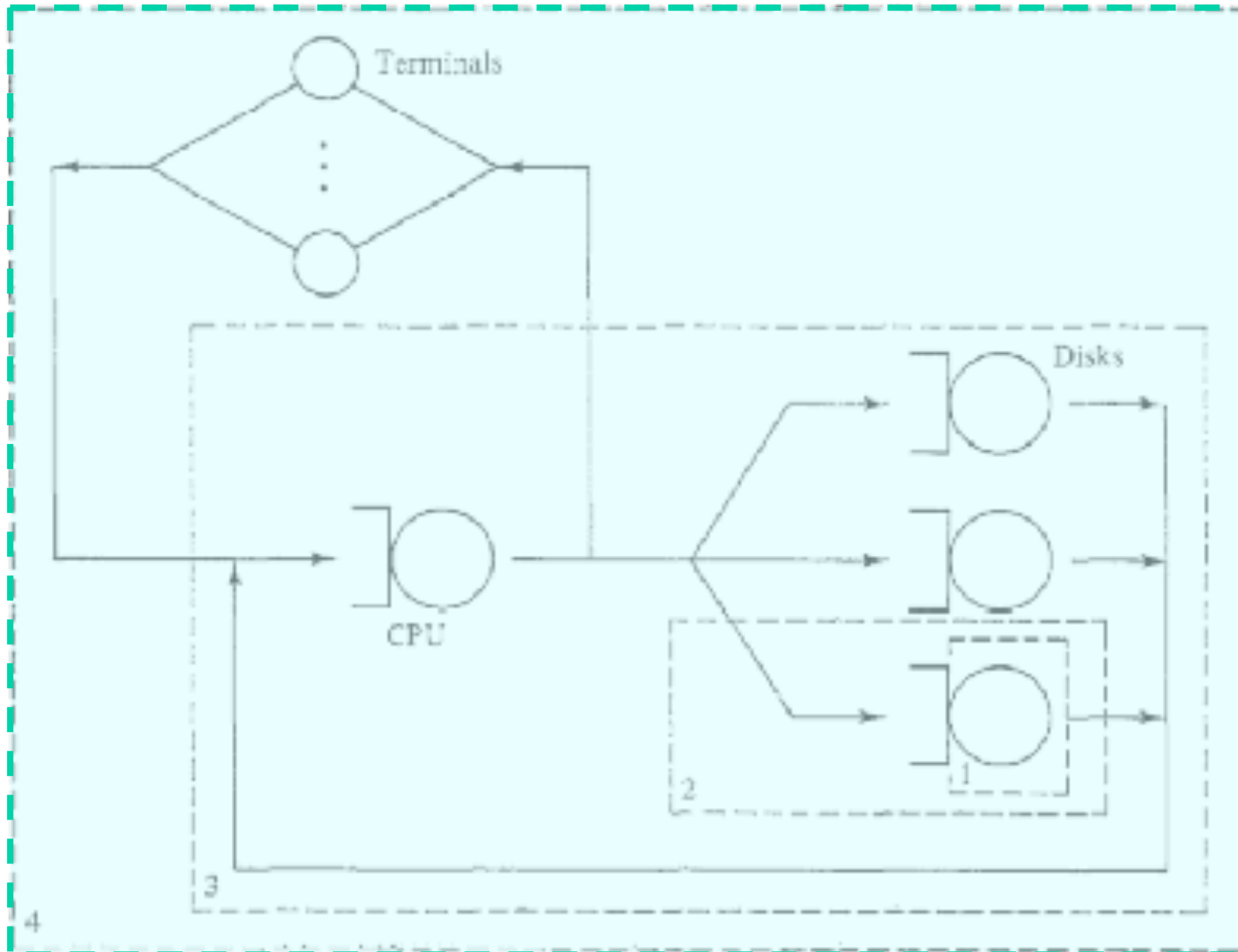


Figure 3.3 – Little's Law Applied at Four Levels

Aplicação da Lei de Little: Sistema Interativo

- Ex: Suponha que 10 usuários utilizem o sistema. Estes usuários fazem processamento local por 5 segundos, em média, antes de submeterem requisições ao servidor central. O tempo de resposta médio observado por eles é de 15 segundos. Qual o throughput do sistema?
- N = número total de usuários
- X = taxa de interações entre terminais e subsistema
- Tempo de residência = tempo de resposta (R) + think time (Z)

$$N = 10 \quad R = 15 \quad Z = 5$$

$$N = X(R+Z) \quad X = N/(R+Z) = 10/20 = 0.5 \text{ interações/seg}$$

Lei do Tempo de Resposta

O tempo de resposta de um sistema interativo é igual à razão do número de clientes no sistema pelo throughput do sistema menos o think time médio dos clientes

$$R = N/X - Z$$

Lei do Fluxo Forçado (Forced Flow Law)

Definição de cliente depende do escopo da análise
disco \Rightarrow cliente = acesso a disco
servidor \Rightarrow cliente = requisição do usuário

$$V_k = \text{número de visitas ao recurso } k \quad V_k \equiv C_k / C$$

Ex: Durante período de medição, 10 requisições foram servidas e houve 150 acessos ao disco 1:

$$V_{\text{disco}_1} = 150/10 = 15 \text{ acessos/requisição}$$

Lei do Fluxo Forçado (Forced Flow Law)

Os throughputs (fluxos de clientes) em todas as partes do sistema devem ser proporcionais uns aos outros

$$X_k = V_k X$$

Exemplo 1

Cada job em um sistema de processamento batch requer uma média de 6 acessos a um disco específico A. Através de medições, você conclui que o disco A em questão está servindo uma média de 12 acessos dos jobs batch a cada segundo.

Qual o throughput do sistema de processamento?

$$X_A = 12$$
$$V_A = 6$$

$$X_A = V_A X \Rightarrow X = X_A / V_A$$
$$X = 12/6 = 2 \text{ jobs/s}$$

Exemplo 2

Você fica sabendo que um outro disco B do mesmo sistema está servindo 18 requisições de jobs batch por segundo, considerando a mesma carga do exemplo anterior.

Quantos acessos ao disco B cada job realiza, em média?

$$X_B = 18$$

$$X_k = V_k X \Rightarrow V_k = X_k / X$$

$$X = 2$$

$$V_k = 18/2 = 9 \text{ acessos}$$

Exemplo 3

Determine o tempo médio de resposta de um sistema interativo com as seguintes características conhecidas:

25 terminais

Think time médio de 18 segundos

Cada interação faz 20 acessos ao disco, em média

Disco está ocupado em média 30% do tempo, durante medição

Tempo de serviço médio por acesso ao disco igual a 25mseg

Sistema interativo \Rightarrow carga interativa \Rightarrow modelo fechado

População de clientes: $N = 25$ e $Z = 18$

Número médio de visitas ao disco $V_{\text{disco}} = 20$

Utilização do disco $U_{\text{disco}} = 0.30$

$S_{\text{disco}} = 0.025$

Qual o valor de R ?

Solução

$$R = N / X - Z \quad (\text{Lei do Tempo de Resposta})$$

Precisamos do valor de X (thpt do sistema)?

$$X_{\text{disco}} = V_{\text{disco}} X \quad (\text{Lei do Fluxo Forçado})$$

Como calcular X_{disco} ?

$$U_{\text{disco}} = X_{\text{disco}} S_{\text{disco}} \quad (\text{Lei da Utilização})$$

$$X_{\text{disco}} = U_{\text{disco}} / S_{\text{disco}} = 0.30 / 0.025 = 12 \text{ acessos/seg}$$

$$X = X_{\text{disco}} / V_{\text{disco}} = 12 / 20 = 0.6 \text{ interações/seg}$$

$$R = N / X - Z = 25 / 0.6 - 18 = 23.7 \text{ segs}$$

Demanda por Serviço em um Dispositivo

- Os requisitos de serviço em cada dispositivo podem ser especificados de duas maneiras

Pelo número de visitas que um cliente faz ao dispositivo uma vez no sistema V_k +
tempo de serviço por visita S_k

ou

Demanda total por serviço do dispositivo D_k

$$D_k \equiv V_k S_k$$

Demanda por Serviço em um Dispositivo

Se um job faz em média 20 visitas a um disco e requer, em cada visita, um tempo médio de serviço de 25 ms.

$$V_k = 20 \quad S_k = 25 \quad D = 20 \times 25 = 500 \text{ ms}$$

A demanda média por serviço no disco é 0.5 seg

Demanda total de um job por todos os recursos do sistema

$$D = \sum_{k=1}^K D_k$$

Lei da Utilização Revisitada

$$U = XS$$

$$U_k = X_k S_k = X V_k S_k = X D_k$$

É mais fácil medir D_k do que V_k e S_k

$$D_k = B_k / C = U_k T / C$$

Exemplo 4

Suponha que um servidor com 3 discos atendendo uma comunidade fechada tenha memória limitada: pode ocorrer swapping e portanto, antes de competir pelos recursos do sistema central, uma interação deve competir por uma partição da memória.

O sistema foi observado e medido:

número médio de usuários : 23

tempo de resposta médio percebido por um usuário: 30 s

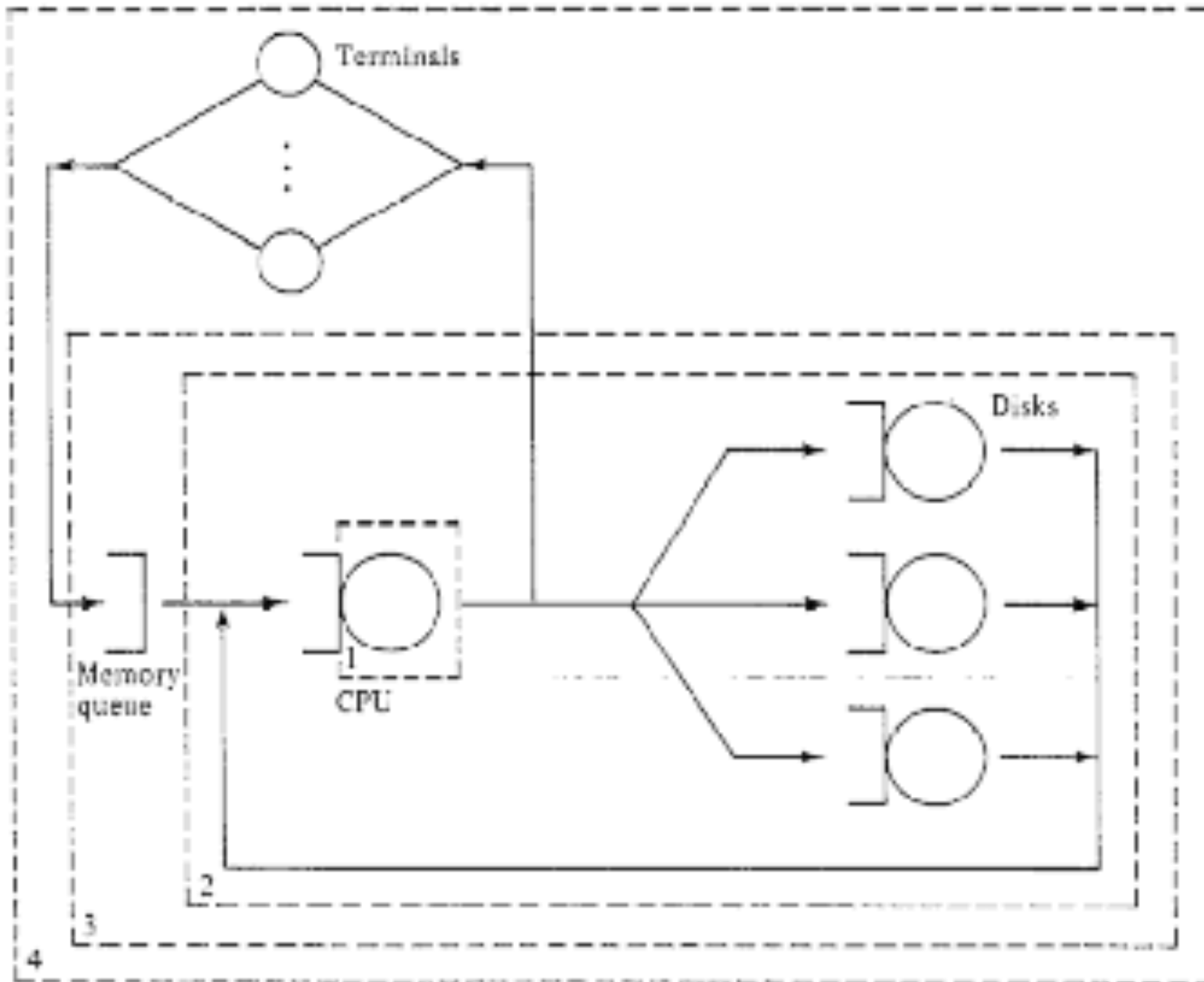
throughput do servidor : 0.45 interacoes / s

número médio de requisições ocupando memória: 1.9

demanda média por CPU para cada interação: 0.63 s

$$N = 23 \quad R = 30 \quad X = 0.45 \quad N_{in_mem} = 1.9 \quad D_{CPU} = 0.63$$

Exemplo 4: Modelo de Filas



Exemplo 4

Qual o think time médio de um usuário?

$$R = N/X - Z \quad Z = N/X - R = 23/0.45 - 30 = 21 \text{ segundos}$$

Em média, quantos usuários estão tentando obter serviço (não estão em think time)?

Aplicar Lei de Little na Caixa 3:

$$N_{\text{want_mem}} = XR = 0.45 \times 30 = 13.5 \text{ usuários}$$

Em média, quantos estão esperando na fila de memória?

$$N_{\text{mem_queue}} = N_{\text{want_mem}} - N_{\text{in_mem}} = 13.5 - 1.9 = 11.6 \text{ usuários}$$

Exemplo 4

Em média, quanto tempo passa desde a aquisição de memória até o término de uma interação?

Aplicar Lei de Little na Caixa 2:

$$N_{\text{in_mem}} = X R_{\text{in_mem}} \quad R_{\text{in_mem}} = N_{\text{in_mem}} / X = 1.9 / 0.45 = 4.2 \text{ s}$$

Qual o tempo médio gasto na fila de memória?

$$R_{\text{mem_queue}} = R - R_{\text{in_mem}} = 30 - 4.2 = 25.8 \text{ segundos}$$

Qual a utilização de CPU pela carga de timesharing?

Aplicar Lei da Utilização na Caixa 1:

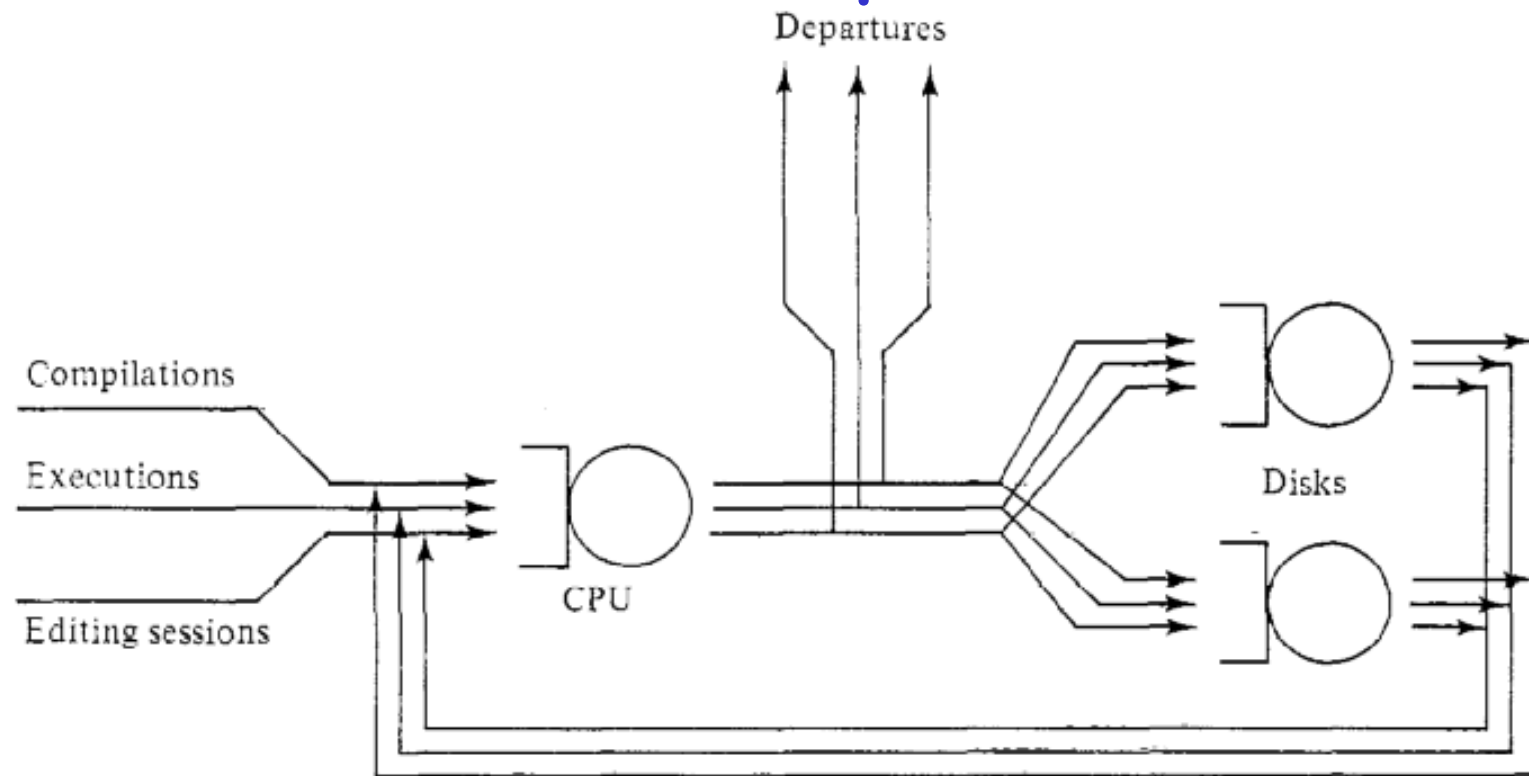
$$U_{\text{CPU}} = X D_{\text{CPU}} = 0.45 \times 0.63 = 28\%$$

Premissa do Equilíbrio de Fluxos

Durante um período de observação,
o número de chegadas no sistema é
igual ao número de saídas

$$A = C \quad \text{logo} \quad \lambda = X$$

Exemplo 5



transaction class	arrival rate trans./hr.	service demand, seconds/transaction		
		CPU	disk 1	disk 2
compilation	480	2.0	0.75	0.25
execution	120	11.9	5.0	5.7
editing session	600	0.5	0.2	0.6

Exemplo 5

Qual a utilização total de CPU?

Assumir Equilíbrio de fluxo: $X = \lambda$

Aplicar Lei da Utilização para cada classe, separadamente

$$U_{\text{comp,CPU}} = X_{\text{comp}} \times D_{\text{comp,CPU}} = 480 / 3600 \times 2 = 27\%$$

$$U_{\text{exec,CPU}} = X_{\text{exec}} \times D_{\text{exec,CPU}} = 120 / 3600 \times 11.9 = 40\%$$

$$U_{\text{edit,CPU}} = X_{\text{edit}} \times D_{\text{edit,CPU}} = 600 / 3600 \times 0.5 = 8\%$$

$$U_{\text{CPU}} = 75\%$$

Se $U_{\text{CPU}} < 100\%$, a premissa do equilíbrio de fluxo é razoável
Deve-se analisar as classes independentemente sem contabilizar as interferências (so faz sentido se $U_{\text{CPU}} < 100\%$).
A interferência aumenta o tempo de resposta.

Leis Fundamentais

Utilizando as leis fundamentais em conjunto, é possível estimar métricas de desempenho do sistema como um todo (e.g. tempo de resposta do sistema), conhecendo métricas de carga de um único dispositivo do sistema

(vide Exemplo 3)

Leis Fundamentais: Sumário

- Lei de Little

$$N = XR$$

- Lei da Utilização

$$U_k = X_k S_k = XD_k$$

- Lei do Tempo de Resposta

$$R = N/X - Z$$

- Lei do Fluxo Forçado

$$X_k = V_k X$$

Relações Adicionais

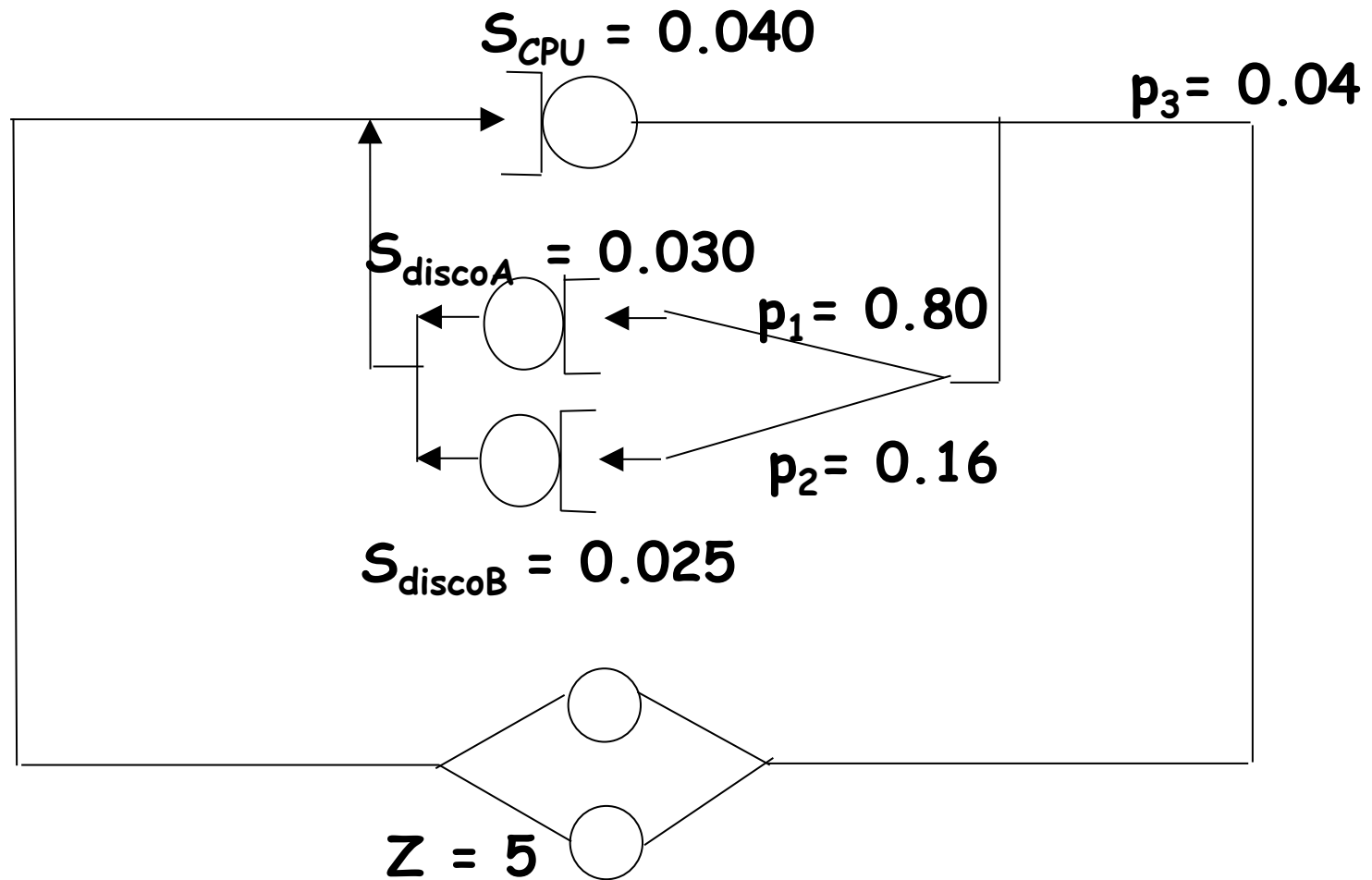
- $\lambda_k \equiv A_k / T$
- $X_k \equiv C_k / T$
- $U_k \equiv B_k / T$
- $S_k \equiv B_k / C_k \equiv U_k T / C_k$
- $V_k \equiv C_k / C$
- $D_k \equiv V_k S_k \equiv B_k / C \equiv U_k T / C$

Exercício 1

Seja um servidor de arquivos com dois discos. Sabe-se que as probabilidades de uma requisição completando serviço na CPU fazer um acesso ao disco A, ao disco B ou de finalizar são 0.80, 0.16 e 0.04, respectivamente. Além disto, foram medidos think time médio dos usuários de 5 segundos, tempos médios de serviços dos disco A e B de 30 e 25 milisegundos, respectivamente, e tempo médio de serviço por visita à CPU de 40 milisegundos. Responda:

- 1) Se a utilização do disco A é de 60%, qual as utilizações da CPU e do disco B?
- 2) Se a utilização do disco B é de 10%, qual o tempo de resposta médio quando há 20 usuários no sistema?

Exercício 1



Exercício 1

- 1) Se a utilização do disco A é de 60%, qual as utilizações da CPU e do disco B?

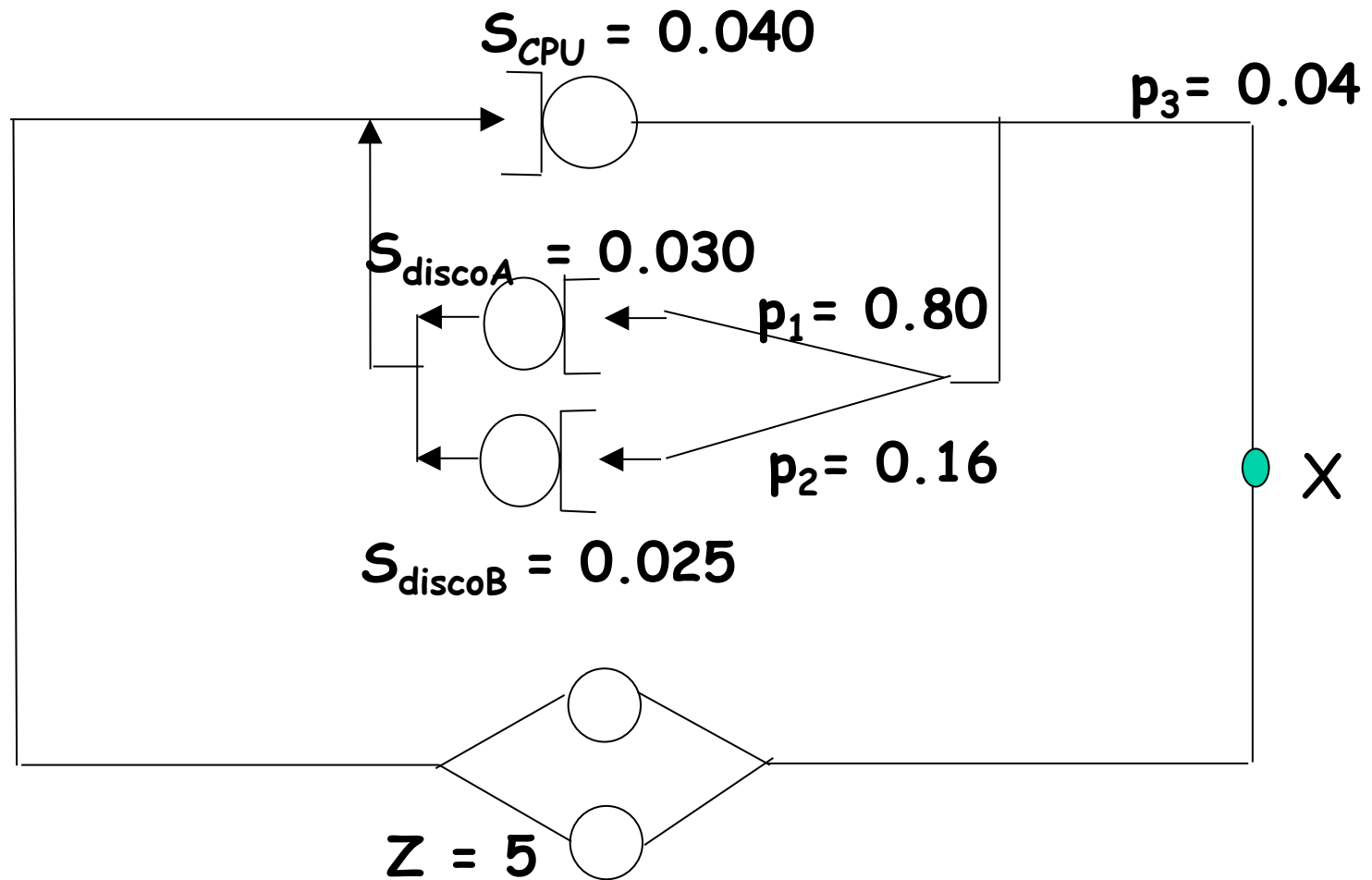
$$S_{CPU} = 0.040 \quad S_{discoA} = 0.030 \quad S_{discoB} = 0.025 \quad Z = 5$$

$$U_{discoA} = X_{discoA} S_{discoA}$$

$$X_{discoA} = U_{discoA} / S_{discoA} = 0.60 / 0.03 = 20$$

$$X_{discoA} = X V_{discoA} \quad V_{discoA} = ?$$

Exercício 1



Exercício 1

1) Se a utilização do disco A é de 60%, qual as utilizações da CPU e do disco B?

$$U_{\text{discoA}} = X_{\text{discoA}} S_{\text{discoA}}$$

$$X_{\text{discoA}} = U_{\text{discoA}} / S_{\text{discoA}} = 0.60 / 0.03 = 20$$

$$X_{\text{discoA}} = X V_{\text{discoA}} \quad V_{\text{discoA}} = ?$$

A cada 100 visitas à CPU, 96 permanecem no sistema (discos) e 4 retornam para os terminais

100 visitas à CPU implicam 80 visitas a disco A e 4 interações

$$V_{\text{discoA}} = 80 / 4 = 20 \text{ acessos por interação}$$

Exercício 1

- 1) Se a utilização do disco A é de 60%, qual as utilizações da CPU e do disco B?

$$X_{\text{discoA}} = XV_{\text{discoA}} \quad X = X_{\text{discoA}} / V_{\text{discoA}} = 20/20 = 1 \text{ inter./seg}$$

$$U_{\text{discoB}} = X_{\text{discoB}} S_{\text{discoB}} = XV_{\text{discoB}} S_{\text{discoB}}$$

$$V_{\text{discoB}} = 16/4 = 4 \text{ acessos por interação}$$

$$U_{\text{discoB}} = XV_{\text{discoB}} S_{\text{discoB}} = 1 \times 4 \times 0.025 = 0.10 = 10\%$$

$$U_{\text{CPU}} = X_{\text{CPU}} S_{\text{CPU}} = XV_{\text{CPU}} S_{\text{CPU}}$$

$$V_{\text{CPU}} = 100/4 = 25 \text{ acessos por interação}$$

$$U_{\text{CPU}} = XV_{\text{CPU}} S_{\text{CPU}} = 1 \times 25 \times 0.040 = 1.0 = 100\%$$

Exercício 1

- 1) Se a utilização do disco B é de 10%, qual o tempo de resposta médio quando há 20 usuários no sistema?

$$U_{\text{discoB}} = X_{\text{discoB}} S_{\text{discoB}} = X V_{\text{discoB}} S_{\text{discoB}} = 0.10$$

$$V_{\text{discoB}} = 16/4 = 4 \text{ acessos por interação}$$

$$X = U_{\text{discoB}} / V_{\text{discoB}} S_{\text{discoB}} = 0.1 / (4 \times 0.025) = 1 \text{ inter. /seg}$$

$$R = N/X - Z = 20/1 - 5 = 15 \text{ segundos}$$

Exercício 2

Um servidor Web, composto de uma CPU e um disco, foi monitorado por 1 hora. A carga do servidor é basicamente composta de requisições para arquivos de imagens.

Durante o intervalo de medição, 10,000 requisições foram processadas. Uma análise do log do servidor mostra que os arquivos têm, em média, 15,000 bytes. Além disto, observou-se que em média o servidor mantinha 5 threads abertas, processando requisições

O tempo médio de serviço no disco é de 12 msec por bloco de 1,000 bytes.

A demanda por CPU para requisições HTTP é dada pela expressão:

$$\text{CPUDemand} = 0.008 + 0.002\text{RequestSize}$$

onde RequestSize é dado pelo número de blocos de 1,000 bytes processados.

Esta expressão indica que existe um tempo constante associado ao processamento de uma requisição (abertura de conexão TCP, análise da requisição e abertura do arquivo).

O segundo componente é proporcional ao tamanho do arquivo, uma vez que processamento de CPU é realizado para cada operação de I/O.

- a) Quais as utilizações da CPU e do disco durante o período de monitoração
- b) Qual o tempo de resposta médio observado pelas requisições monitoradas?
- c) O sistema suportaria uma carga 5 vezes mais pesada?

Exercício 2

Modelo aberto ou fechado?

Carga especificada pelo número de reqs em um período de medição (taxa):
modelo aberto

$$\lambda = 10000 / 3600 = 2.78 \text{ reqs/s}$$

Próximo passo: calcular demandas na CPU e disco:

Arquivos de imagem = 15000 bytes = 15 blocos

$$D_{\text{CPU}} = 0.008 + 0.002 * 15 = 0.038 \text{ s} \quad D_{\text{disk}} = 0.012 * 15 = 0.18 \text{ s}$$

Utilização de CPU:

Throughput $X = \lambda = 2.78 \text{ (reqs/s)}$ (Premissa de Equilíbrio de Fluxos)

$$U_{\text{CPU}} = X D_{\text{CPU}} = 2.78 * 0.038 = 10.5\% \quad U_{\text{disk}} = X D_{\text{disk}} = 2.78 * 0.18 = 50\%$$

Exercício 2

Tempo de Resposta Médio:

$$X = 2.78$$

Número de threads abertas = número de requisições em processamento simultâneo (N)

$$N = 5$$

$$R = N/X = 5 / 2.78 = 1.8 \text{ seg (Little)}$$

O servidor aguenta uma carga 5 x mais pesada?

$$\lambda^{\text{new}} = 5 * 2.78 = 13.9 \text{ reqs/s}$$

A CPU aguenta?

$$U_{\text{CPU}} = X D_{\text{CPU}} = \lambda^{\text{new}} D_{\text{CPU}} = 13.9 * 0.038 = 0.53 \text{ (ok!)}$$

O disco aguenta?

$$U_{\text{disk}} = X D_{\text{disk}} = \lambda^{\text{new}} D_{\text{disk}} = 13.9 * 0.18 = 2.502 \gg 1 \text{ (Não é factível)}$$

Exercício 3

Suponha um servidor que mantenha atualmente uma média de 250 conexões TCP abertas, simultaneamente. Suponha também que o intervalo de tempo entre abertura e fechamento de uma conexão seja dominado pelo tempo necessário para transmitir os dados para o usuário.

Assumindo que todos os usuários atuais estejam conectados por uma linha ISDN (128Kbps), quantas conexões TCP estariam abertas (simultaneamente) se todos decidissem mudar para um conexão de 1Mbps, assumindo nenhum aumento na taxa de submissão de requisições ??

$$N = 250 = XR$$

$$N' = X'R' = XR(128/1000) = 32$$

Exercício 4

Suponha um roteador que receba em média 1,000,000 pacotes por segundo e cada pacote exige um processamento que gasta, em média, 2.5 milissegundos. Quantos pacotes em média o roteador deve manter em seu buffer?

O gerente da rede cogita reduzir a memória disponível no roteador para armazenamento para 200KB. Isto é recomendável? Quais implicações? Suponha que cada pacote tem 100 bytes.

$$X = 1,000,000 \text{ pacotes/s} \quad R = 2.5 \text{ ms} = 0.0025$$

$$N = XR = 1,000,000 * 0.0025 = 2,500$$

$$\text{Memória necessária} = N * 100 = 250,000 \text{ bytes} \sim 245 \text{ KB}$$

Não é recomendável , se não haverá perda de pacotes.

Exercício 5

Suponha um desktop PC conectado a uma LAN 100 Mbps. Um software de video-conferencia utiliza um codificador que requer uma media de 390,321 bits/frame e envia 33 frames/seg. Qual a utilização da LAN devido a uma única conexão? Quantas conexoes de video-conferencia simultaneas a LAN suportaria?

$$\lambda = X = 2 * 33 = 66 \text{ frames/seg (upload/download)}$$

$$D = 390,321 \text{ bits/frame} / 100\text{Mbps} = 0.00390321 \text{ seg/frame}$$

$$U = \lambda D = 66 * 0.00390321 = 0.2576 = 25.76\%$$

Ate no maximo 3 conexoes (considerando apenas a rede)