Teoria das Filas

Mário Meireles Teixeira
Departamento de Informática, UFMA
mario@deinf.ufma.br

Filas, filas...

- As filas são a "praga" do mundo atual!
- Espera-se em fila no banco, na padaria, no ponto de ônibus, no trânsito, no restaurante...
- Em sistemas computacionais, há filas por toda parte: para acessar a CPU, o disco, a memória, a rede, a impressora, os servidores e outros recursos
- Os clientes podem ser: pessoas, processos, threads, jobs, pacotes, transações de BD, requisições C/S

- As filas surgem porque a demanda de serviço é maior que a capacidade de atendimento do sistema

O que é a Teoria das Filas?

- É um ramo da Probabilidade que estuda o fenômeno da formação de filas de solicitantes de serviços, fornecidos por um determinado recurso
- Permite estimar importantes medidas de desempenho de um sistema a partir de propriedades mensuráveis das filas
- Dessa forma, pode-se dimensionar um determinado sistema segundo a demanda dos seus clientes, evitando desperdícios ou gargalos ☺
- Contudo, as filas apresentam comportamento estocás-tico... ☹

- Aplicações:
  - fluxo de tráfego (veículos, pessoas, redes de comunicação)
  - escalonamento (pacientes, tarefas industriais, processos)
  - serviços de atendimento (bancos, restaurantes, servidores)

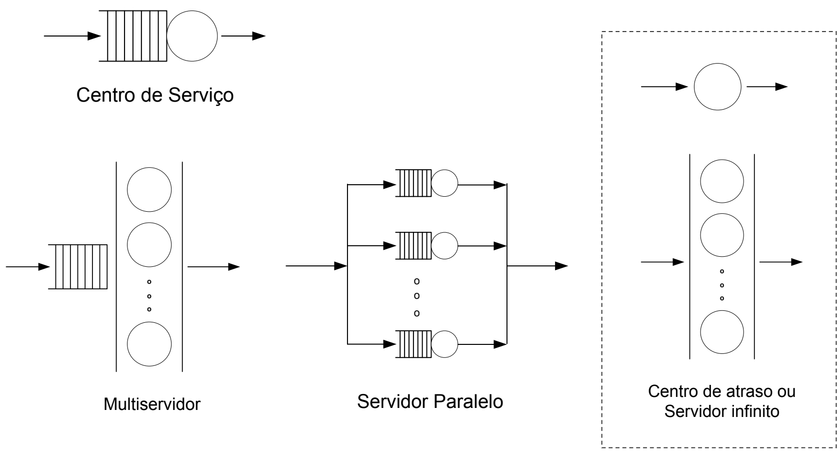
Definições e Terminologia

- Rede de Filas
  - Consiste em um conjunto de entidades interligadas que oferecem serviços (os centros de serviço) e de usuários (os clientes)

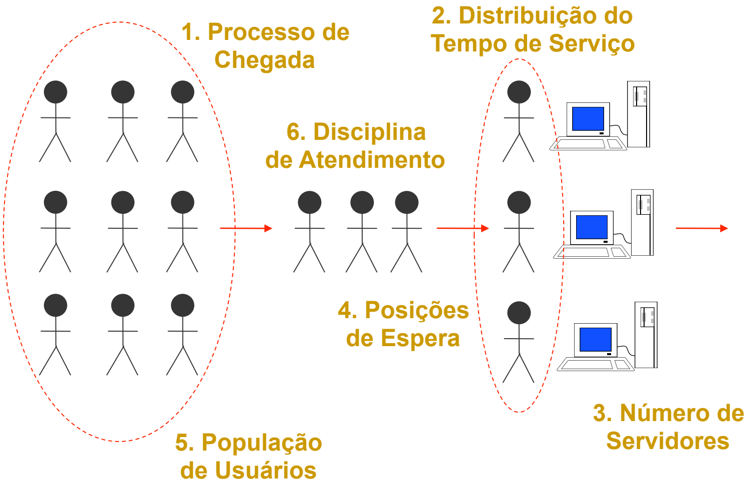
- Centro de Serviço
  - Representa os recursos do sistema
  - Compreende um ou mais servidores e um conjunto de clientes esperando por serviço

- Fila = cliente em serviço + clientes em espera
- Fila de espera = clientes em espera

Simbologia



Componentes de uma Fila



Características das Filas

1. Processo de Chegada: apresenta um comportamento estocástico

   É fundamental conhecer a distribuição de probabilidade dos tempos entre as chegadas:
   - Mais comum: Chegadas de Poisson (tempos entre as chegadas são exponencialmente distribuídos)
   - Outras distribuições: Erlang, hiperexponencial, arbitrária

   E ainda:
   - chegadas de clientes individuais / simultâneas
   - cliente sempre decide ficar / não fica se a fila for muito grande / impaciente / muda de fila
   - padrão de chegada estacionário / não-estacionário

Características das Filas

2. Distribuição dos Tempos de Serviço: os tempos de serviço dos clientes também são variáveis aleatórias independentes e identicamente distribuídas (IID)

   Distribuições comuns: exponencial, Erlang, hiperexponencial, arbitrária (geral)

   E ainda:
   - atendimentos simples / batch
   - serviço independente / dependente do estado
   - serviço estacionário / não-estacionário

Características das Filas

3. Número de Servidores: número de posições de atendimento disponíveis no sistema
   - Servidores idênticos / distintos
   - Fila única / por servidor / por grupo de servidores

4. Capacidade do Sistema: número máximo de clientes que podem permanecer no sistema, devido a restrições de espaço (buffers) ou de tempo de espera
   - Inclui clientes em serviço e esperando por serviço
   - Capacidade pode ser finita / infinita (mais fácil de analisar)

Características das Filas

5. Tamanho da População (fonte): número potencial de clientes que podem chegar a um sistema
   - Tamanho finito / infinito

6. Disciplina de Atendimento (de fila): ordem na qual os clientes são atendidos
   Tipos:
   - FCFS (First-Come First-Served) ou FIFO
   - LCFS (Last-Come First-Served) ou LIFO
   - RR (Round Robin) / PS (Processor Sharing)
   - Prioridades (fila única / múltiplas filas)
   - Não-preemptivo / Preemptivo (resume/repeat) ...

Notação de Kendall

- Para especificar um sistema de filas, é preciso conhecer as seis características anteriores
- Notação de Kendall:

  A/S/m/K/N/Q

  - **A** : distribuição dos tempos entre chegadas
  - **S** : distribuição dos tempos de serviço
  - **m**: número de servidores
  - **K** : capacidade do sistema
  - **N** : tamanho da população
  - **Q** : disciplina de atendimento

Notação de Kendall

- Exemplos:
  - M/G/4/50/2000/LCFS
  - D/M/1/∞/∞/RR
  - D/U/2/1000/∞/FCFS

  - A/S/m ou A/S/m/∞/∞/FCFS (.../∞/∞/FCFS é default)
  - M/M/1 ou M/M/1/∞/∞/FCFS
  - M/M/m
  - M/M/m/K
  - G/G/1

Distribuições de Probabilidade

- Determinista (D)
  - Tempo entre as chegadas e o Tempo de Serviço são constantes
  - Não há variância estatística
  - Pelo menos uma das distribuições (chegada ou serviço) precisa ser aleatória, caso contrário o sistema de filas terá baixa aplicabilidade no mundo real

Distribuições de Probabilidade

- Exponencial (M)
  - Para Chegadas (A): o intervalo entre uma chegada e a próxima é completamente independente do período anterior
  - Para Tempos de Serviço (S): o tempo de serviço atual é independente do tempo de serviço anterior
  - Esses processos são ditos "sem memória", pois seus intervalos não estão correlacionados no tempo. Portanto, podem ser caracterizados por uma distribuição exponencial

Distribuições de Probabilidade

- Uniforme (U)
  - Os tempos de chegada estão limitados por algum valor finito ($a \le x \le b$)
  - A probabilidade de x assumir qualquer dos valores do intervalo é a mesma → média = $(a + b)/2$

- Arbitrária ou Geral (G)
  - Não é especificada uma distribuição de probabilidade para os tempos de chegada e serviço
  - Resultados são válidos para todas as distribuições

Distribuições de Probabilidade

- Erlang (E_k)
  - Generalização da distribuição exponencial
  - Um servidor com *k* estágios, obedecendo à distribuição de Erlang, pode ser representado por uma seqüência de *k* servidores com tempos de serviço exponencialmente distribuídos, de mesma média

- Hiperexponencial (H_k)
  - Cada estágio no modelo de Erlang tem média diferente para os tempos de serviço
  - Os estágios estão organizados em paralelo, mas o serviço é fornecido um por vez

Processos Estocásticos

- São funções ou seqüências aleatórias dependentes do tempo
- Exemplos:
  - *n(t)* – número de jobs na CPU de um sistema
  - *W(t)* – tempo de espera em fila
- Os processos estocásticos são úteis para representar o estado de sistemas de filas

Tipos de Processos Estocásticos

- **Processos de Estado Discreto e Estado Contínuo:**
  Discreto: número de valores de estado possíveis é finito ou contável; também chamado de *cadeia estocástica*. Ex: *n(t)*

  Contínuo: pode assumir qualquer valor entre os números reais. Ex: *w(t)*
- **Processos de Markov:** os estados futuros do sistema independem do passado e dependem exclusivamente do estado atual.
  Um processo de Markov de estados discretos é chamado de *Cadeia de Markov*

Tipos de Processos Estocásticos

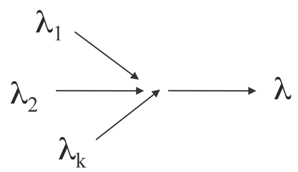
- **Processos de Nascimento e Morte:** processos de Markov de espaço discreto em que as transições entre estados estão restritas a estados vizinhos.
  Ex: número de jobs em um servidor único com chegadas individuais

- **Processos de Poisson:** se os tempos entre as chegadas têm distribuição exponencial, o número de chegadas em um dado intervalo terá uma distribuição de Poisson. O processo de chegadas é chamado de *Processo de Poisson*.
  As chegadas são "sem memória", pois o tempo entre as chegadas é IID e exponencialmente distribuído.

Propriedades dos Fluxos de Poisson

- Superposição

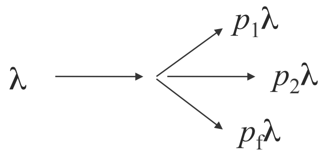
$$\lambda = \sum_{i=1}^{k} \lambda_i$$



- A superposição de fluxos de Poisson dá como resultado um novo fluxo de Poisson cuja taxa de chegada é o somatório das taxas dos fluxos originais

Propriedades dos fluxos de Poisson

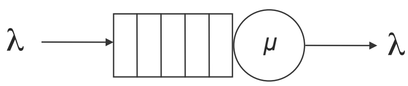
- Divisão



- Se um fluxo de Poisson for dividido em *f* subfluxos com probabilidade p_i de um usuário seguir o subfluxo *i*, então cada subfluxo é também um fluxo de Poisson com taxa média $p_i\lambda$

Propriedades dos fluxos de Poisson

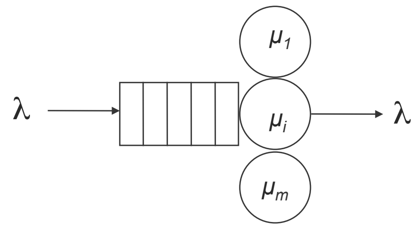
- Se as chegadas a uma fila com um servidor único e tempo de serviço exponencial forem Poisson com taxa média λ, então as partidas também serão Poisson, com a mesma taxa λ (desde que $\lambda < \mu$)



- $\lambda < \mu$: condição de equilíbrio do sistema

Propriedades dos fluxos de Poisson

- Se as chegadas a uma fila com *m* servidores e tempos de serviço exponenciais forem Poisson com taxa média λ, então as partidas também serão Poisson, com a mesma taxa λ (desde que $\lambda < \Sigma\, \mu_i$)

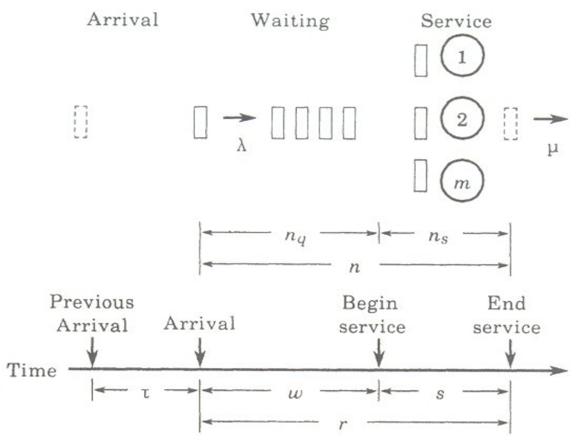


Leis Operacionais

Introdução

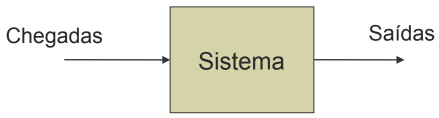
- Relações simples que não necessitam de nenhuma hipótese sobre as distribuições dos tempos de serviço ou dos intervalos entre chegadas
- Foram identificadas inicialmente por Buzen (1976) e posteriormente estendidas por Denning e Buzen (1978)
- A palavra operacional significa que pode ser medida diretamente

Variáveis Aleatórias de uma Fila



Quantidades Operacionais

- São quantidades que podem ser medidas diretamente durante um período finito de observação:
  - Período de observação – T
  - Número de chegadas (arrivals) – A_i
  - Número de términos (completions) – C_i
  - Tempo ocupado (busy time) – B_i



Quantidades Operacionais

$$\text{Taxa de chegada } \lambda_i = \frac{\text{número de chegadas}}{\text{tempo}} = \frac{A_i}{T}$$

$$\text{Throughput } X_i = \frac{\text{número de términos}}{\text{tempo}} = \frac{C_i}{T}$$

$$\text{Utilização } U_i = \frac{\text{tempo ocupado}}{\text{tempo total}} = \frac{B_i}{T}$$

$$\text{Tempo médio de serviço } S_i = \frac{\text{tempo total de serviço}}{\text{número de saídas}} = \frac{B_i}{C_i}$$

Estas quantidades são variáveis que podem mudar de um período de observação para outro, mas as relações permanecem válidas!

Lei da Utilização

$$U_i = \frac{B_i}{T} = \frac{C_i}{T} \times \frac{B_i}{C_i}$$

$$U_i = X_i S_i$$

Exemplo 33.1

- Considere um roteador em que os pacotes chegam a uma taxa de 125 pps e o roteador leva em média 2 ms para encaminhá-los. Qual a utilização do sistema?

$$X_i = \text{taxa de saída} = \text{taxa de chegada} = 125 \text{ pps}$$

$$S_i = 0{,}002 \text{ segundos}$$

$$U_i = X_i S_i = 125 \times 0{,}002 = 0{,}25 = 25$$

Este resultado é válido para qualquer processo de chegada ou atendimento!

Lei de Little

- A Lei de Little relaciona o número de clientes no sistema com o tempo médio despendido no sistema:

$$Q_i = \lambda_i R_i$$

Número médio = Taxa de chegada x Tempo médio de resposta

- $R_i = S_i + W_i$
- Esta lei se aplica sempre que o número de chegadas for igual ao número de saídas (sistema em equilíbrio)
- Pode-se aplicar a lei de Little a qualquer sistema ou subsistema (caixa preta)

Lei de Little

- Se o sistema está em equilíbrio, a taxa de chegada é igual ao throughput, portanto:

$$Q_i = X_i R_i$$

- Exemplo 3.14: Um servidor de arquivos NFS foi monitorado durante 30 minutos e o número observado de operações de I/O foi 10.800. Apurou-se que o número médio de pedidos ativos no NFS era três. Qual o tempo médio de resposta por pedido no servidor?

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 de um servidor de arquivos
  - Assuma 1 CPU, 3 discos e carga balanceada (probabilidade de requisição ser enviada para
    cada disco é 1/3)

Exemplo: Servidor de Arquivos

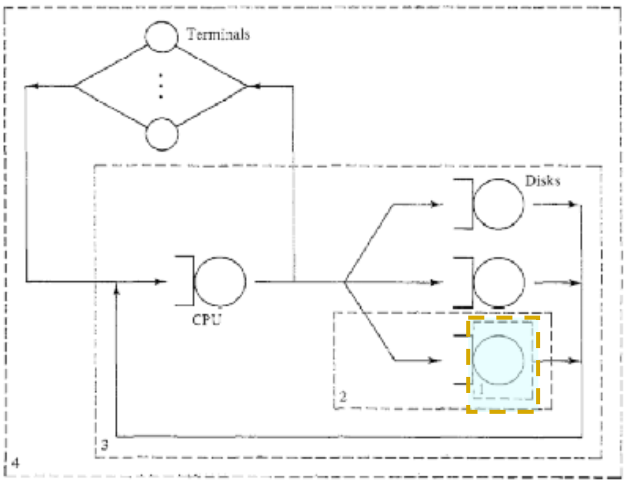


Figure 3.3 — Little's Law Applied at Four Levels

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)

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

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

$U = 90\%$

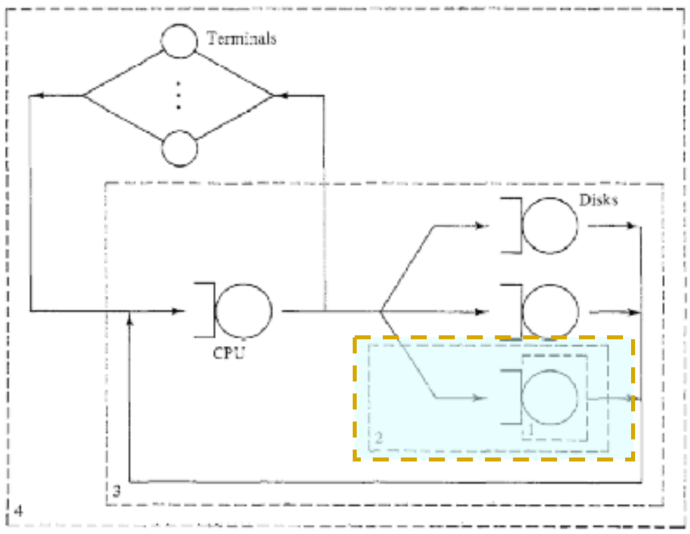


Figure 3.3 — Little's Law Applied at Four Levels

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 de 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 X = 40

N = RX R = N/X = 0.1 segundos (Lei de Little)

Tempo na fila: R - 0.0225 = 0.1 - 0.0225 = 0.0775 seg

Tamanho da fila: N - U = 4 - 0.9 = 3.1

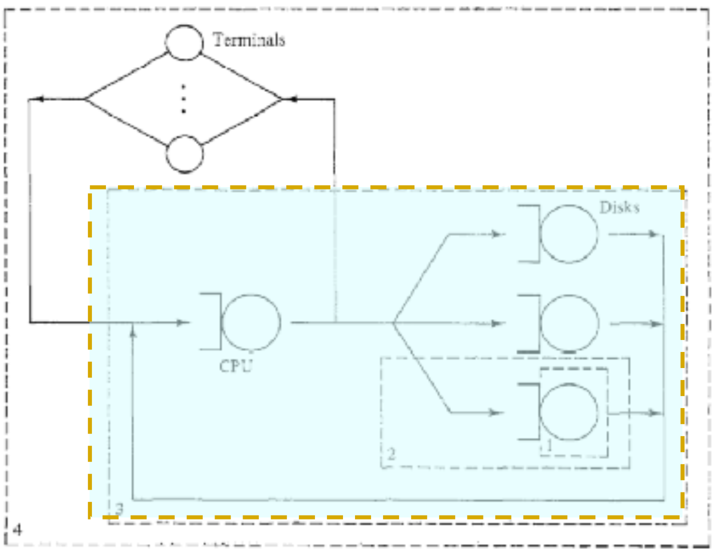


Figure 3.3 — Little's Law Applied at Four Levels

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 X = 1/2

N = RX R = N/X = 15 segundos

Lei do Fluxo Forçado

- Relaciona o throughput global do sistema com o throughput dos dispositivos individuais
- Se o período de observação T for tal que o número de chegadas em cada dispositivo é igual ao número de saídas, i.e., $A_i = C_i$, diz-se que o dispositivo satisfaz a Hipótese de Equilíbrio (*job flow balance*)
- Para um período de observação longo o bastante, a diferença A_i-C_i é normalmente pequena se comparada com C_i

Lei do Fluxo Forçado

- Seja V_i o número médio de visitas ao recurso i por uma tarefa
- Cada pedido que termina precisa passar, em média, V_i vezes pelo recurso i. Assim, se X pedidos foram concluídos por unidade de tempo, temos que $V_i X$ pedidos terão passado pelo recurso i:

$$X_i = V_i X$$

- Esta lei é aplicável sempre que a hipótese de equilíbrio for verdadeira

Lei da Demanda de Serviço

- Combinando as leis da Utilização e do Fluxo Forçado, temos:

$$U_i = X_i S_i = X V_i S_i$$

ou

$$U_i = X D_i$$

- Onde $D_i = V_i S_i$ é a demanda total de serviço no i-ésimo dispositivo
- O dispositivo com a maior demanda de serviço tem a maior utilização e pode tornar-se o gargalo do sistema

Exemplos 3.12 e 3.13

- As transações de um banco de dados realizam uma média de 4,5 operações de I/O no servidor de BD. O servidor foi monitorado durante uma hora e, durante esse período, 7.200 transações foram concluídas.

  a) Qual a taxa média de processamento no disco?
  b) Se cada I/O de disco leva 20 ms em média, qual a utilização do disco?
  c) Qual a demanda de serviço do disco?

Lei Geral do Tempo de Resposta

- Sistemas de tempo compartilhado podem ser divididos em dois subsistemas: o subsistema de terminais e o subsistema central de processamento
- Dados os comprimentos individuais Q_i das filas de cada dispositivo, podemos calcular Q:

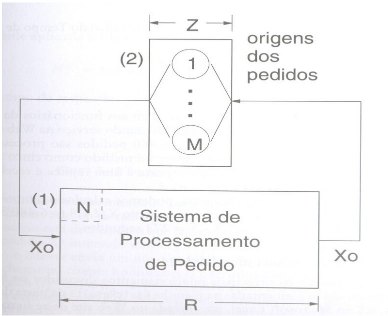
$$Q = Q_1 + Q_2 + \cdots + Q_M$$
$$XR = X_1 R_1 + X_2 R_2 + \cdots + X_M R_M$$

- Dividindo ambos os lados por X e usando a lei do fluxo forçado:

$$R = V_1 R_1 + V_2 R_2 + \cdots + V_M R_M \quad \text{ou} \quad R = \sum_{i=1}^{M} R_i V_i$$

Lei do Tempo de Resposta Interativo

- Num sistema interativo, os usuários geram pedidos que são processados pelo subsistema central e os resultados voltam ao terminal. Após um tempo ocioso Z, o usuário submete o próximo pedido.



45

Lei do Tempo de Resposta Interativo

- Aplicando-se a lei de Little ao subsistema central, temos:

$$Q = XR$$

- Agora, aplicando-se a lei de Little aos M terminais:

$$\overline{M} = XZ$$

- Considerando que um cliente ou está sendo processado ou está ocioso:

$$M = Q + \overline{M} = XR + XZ = X(R + Z)$$

$$R = \frac{M}{X} - Z$$

46

Exemplo: Sistema Timesharing

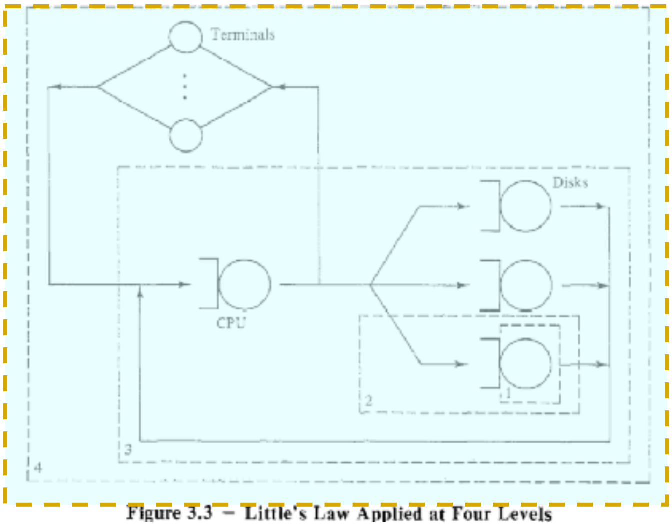


Figure 3.3 — Little's Law Applied at Four Levels

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 R = 15 Z = 5

N = X(R+Z) X = N/(R+Z) = 10/20 = 0.5 interações/seg

Exemplo 3.16

- Um portal corporativo oferece serviços na Web aos funcionários de uma empresa. Em média, 500 funcionários estão on-line solicitando serviços. Uma análise do log do portal revelou que, em média, 6.480 pedidos são processados por hora. O tempo de resposta médio por pedido é de cinco segundos.

  Qual o tempo médio entre o momento em que a resposta a uma réplica é recebida e um novo pedido é enviado por um funcionário?

Box 33.1 Operational Laws

Utilization law	$U_i = X_i S_i = X D_i$
Forced flow law	$X_i = X V_i$
Little's law	$Q_i = X_i R_i$
General response time law	$R = \sum_{i=1}^{M} R_i V_i$
Interactive response time law	$R = N/X - Z$
Asymptotic bounds	$R \geq \max\{D, N D_{max} - Z\}$
	$X \leq \min\{1/D_{max}, N/(D+Z)\}$

Symbols:

D	Sum of service demands on all devices, $= \sum_i D_i$
D_i	Total service demand per job for the ith device, $= S_i V_i$
D_{max}	Service demand on the bottleneck device, $= \max_i\{D_i\}$
N	Number of jobs in the system
Q_i	Number in the ith device
R	System response time
R_i	Response time per visit to the ith device
S_i	Service time per visit to the ith device
U_i	Utilization of the ith device
V_i	Number of visits per job to the ith device
X	System throughput
X_i	Throughput of the ith device
Z	Think time

Exemplo

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 ⇒ carga interativa ⇒ modelo fechado

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

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

Utilização do disco $U_{disco} = 0.30$

$S_{disco} = 0.025$

Qual o valor de R?

Solução

$R = N / X - Z$ (Lei do Tempo de Resposta)

Precisamos do valor de X (thpt do s

$X_{disco} = V_{disco} X$ (Lei do Fluxo

Como calcular X_{disco}?

$U_{disco} = X_{disco} S_{disco}$ (Lei da Utiliza

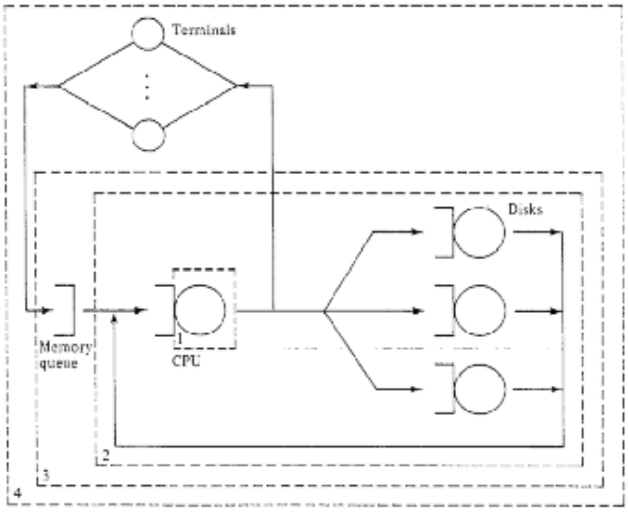
Utilizando as leis fundamentais em conjunto, é possível estimar métricas de desempenho do sistema como um todo (p.ex., tempo de resposta do sistema), conhecendo métricas de carga de um único dispositivo do sistema

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

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

$R = N / X - Z = 25 / 0.6 - 18 = 23.7$ segs

Aplicação da Lei de Little: Sistema Interativo com memória



Exemplo

Suponha que um servidor com três discos atendendo a 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 interações / s
- número médio de requisições ocupando memória: 1.9
- demanda média por CPU para cada interação: 0.63 s

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

Exemplo

Qual o think time médio de um usuário?

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

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

Aplicar Lei de Little na Caixa 3:

$N_{want\_mem} = XR = 0.45 \times 30 = 13.5$ usuários

Em média, quantos usuários estão esperando na fila da memória?

$N_{mem\_queue} = N_{want\_mem} - N_{in\_mem} = 13.5 - 1.9 = 11.6$ usuários

Exemplo

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

Aplicar Lei de Little na Caixa 2:

$N_{in\_mem} = XR_{in\_mem} \quad R_{in\_mem} = N_{in\_mem} / X = 1.9 / 0.45 = 4.2$ s

Qual o tempo médio gasto na fila da memória?

$R_{mem\_queue} = R - R_{in\_mem} = 30 - 4.2 = 25.8$ segundos

Qual a utilização de CPU pela carga de timesharing?

Aplicar Lei da Utilização na Caixa 1:

$U_{CPU} = XD_{CPU} = 0.45 \times 0.63 = 28\%$

Qual o gargalo aparente deste sistema?

A memória

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

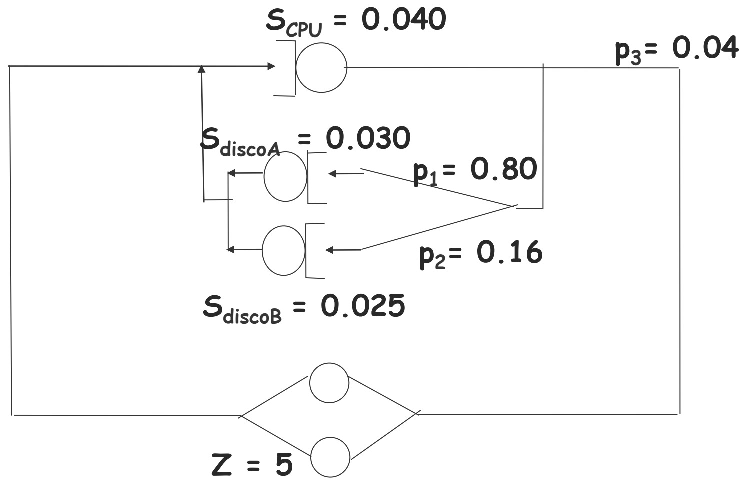
- $\lambda_k \equiv A_k / T$
- $X_k \equiv C_k / T$
- $U_k \equiv B_k / T$
- $S_k \equiv B_k / C_k \equiv U_k T / C_k$
- $V_k \equiv C_k / C$
- $D_k \equiv V_k S_k \equiv B_k / C \equiv U_k T / C$

Exercício

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 disso, foram medidos o think time médio dos usuários de 5 segundos, tempos médios de serviços dos discos A e B de 30 e 25 ms, respectivamente, e tempo médio de serviço por visita à CPU de 40 ms. Responda:

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

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

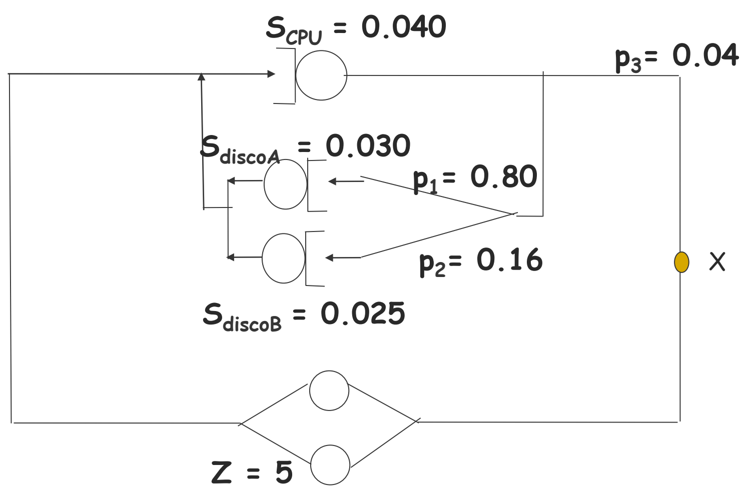


a) Se a utilização do disco A é de 60%, qual a utilização da CPU e do disco B?

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

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

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



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

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

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

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

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

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

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

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

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

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

V_{discoB} = 16/4 = 4 acessos por interação

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

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

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

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

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

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

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

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

$R = N/X - Z = 20/1 - 5 = 15$ segundos