Partilha de um Recurso de Comunicações

Desempenho e Dimensionamento de Redes

Prof. Amaro de Sousa (asou@ua.pt)

DETI-UA, 2017/2018

Recurso de comunicações digital

Um recurso de comunicações digital é um elemento de rede que permite a troca de informação em formato digital.

Pode ser visto como um *túnel de bits* através do qual pode ser transmitido/recebido um dado valor em bits/s. Este valor é designado por *capacidade do recurso*.

Exemplos:

- Cada sentido de uma ligação ponto-a-ponto bidirecional (um cabo Ethernet a ligar um PC a um *switch* Ethernet) é um recurso de comunicações entre os dois elementos.
- Uma interface de rede de um servidor de vídeo é um recurso de comunicações do servidor para os terminais de vídeo.
- Uma ligação de um router de uma casa ao ISP é constituído por dois recursos de comunicação, um em cada sentido, cuja capacidade é normalmente diferente em cada sentido.

Recurso de comunicações organizado em circuitos

Um recurso de comunicações pode ser explicitamente estruturado em *circuitos* (ou *canais*), que correspondem a partições da capacidade total do recurso.

- Um circuito é caracterizado pela sua largura de banda (um valor também em bits/s).
- A estruturação de uma ligação ponto-a-ponto em circuitos pode ser feita através de:

  (1) multiplexagem no tempo (sistemas TDM)

  (2) multiplexagem na frequência (sistemas FDM)

  (3) multiplexagem de comprimentos de onda (sistemas WDM)

Um recurso de comunicações pode também ser implicitamente usado como se um fosse estruturado em circuitos (noção de circuitos virtuais).

- Cada comunicação usa uma partição da capacidade do recurso.

Recurso de comunicações com estabelecimento de circuitos

Uma *chamada*, quando é estabelecida, corresponde à atribuição temporária de um circuito (ou mais) com uma determinada largura de banda.

Após estabelecida, a chamada dura um tempo finito após o qual a largura de banda do(s) circuito(s) atribuído(s) é libertada e fica disponível para pedidos futuros de chamadas.

Assim, um recurso de comunicações pode ser partilhado por diferentes *fluxos de chamadas*.

Uma *classe de serviço* identifica um fluxo de chamadas com *as mesmas características de tráfego* e *os mesmos requisitos de largura de banda*.

Recurso de comunicações com estabelecimento de circuitos

As características de tráfego são determinadas

(i) pela descrição estocástica do processo de chegada de chamadas,

(ii) pela descrição estocástica do tempo das chamadas, e

(iii) pela largura de banda (em nº de circuitos) requerida por cada chamada.

O parâmetro de qualidade de serviço mais importantes é a *probabilidade de bloqueio* (probabilidade de um pedido de chamada não ser aceite pela ligação por falta de recursos disponíveis).

Um recurso de comunicações pode ser:

(1) *uni-serviço* – se suportar apenas fluxos de chamadas de uma única classe de serviço

(2) *multi-serviço* – se suportar fluxos de chamadas de diferentes classes de serviço

A diferenciação de serviço faz-se através da função de controle de fluxos (também designada de controle de admissão de chamadas)

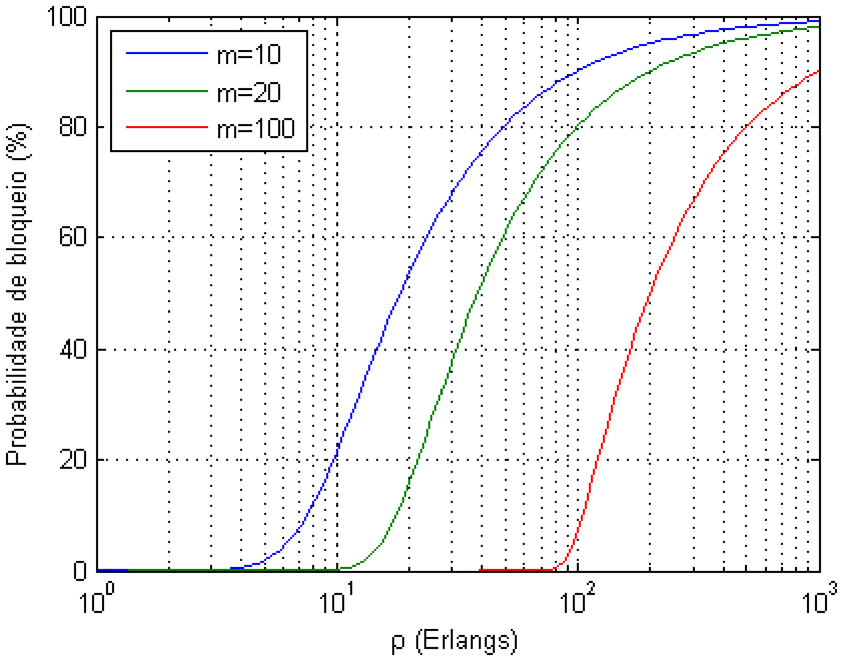
Recurso de comunicações uni-serviço - Distribuição de ErlangB

Um recurso de comunicações modelado por um grupo de *m* circuitos ao qual é oferecido um *fluxo de chamadas de Poisson* com taxa λ e em que *a duração das chamadas é exponencialmente distribuída* com média $1/\mu$ é modelado por um sistema de fila de espera *M/M/m/m*.

Designa-se como unidade de *intensidade de tráfego*, $\rho = \lambda/\mu$, o Erlang.

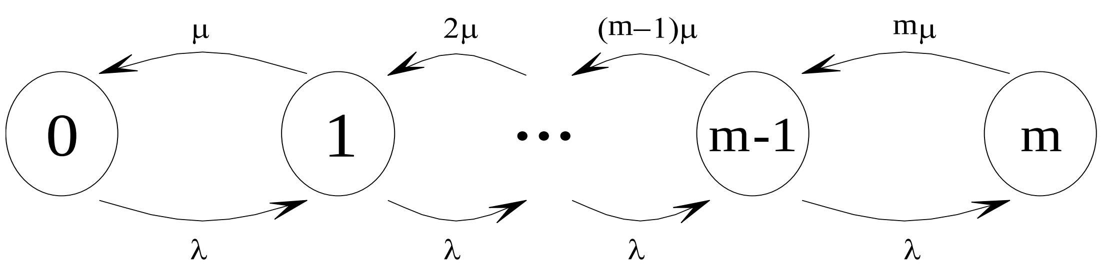
A fórmula de ErlangB, dá a probabilidade de bloqueio:

$$P_m = \frac{\rho^m / m!}{\sum_{n=0}^{m} \rho^n / n!} = E(\rho, m)$$



Estabelecimento de circuitos uni-serviço

Cadeia de Markov de um sistema de fila de espera *M/M/m/m*:

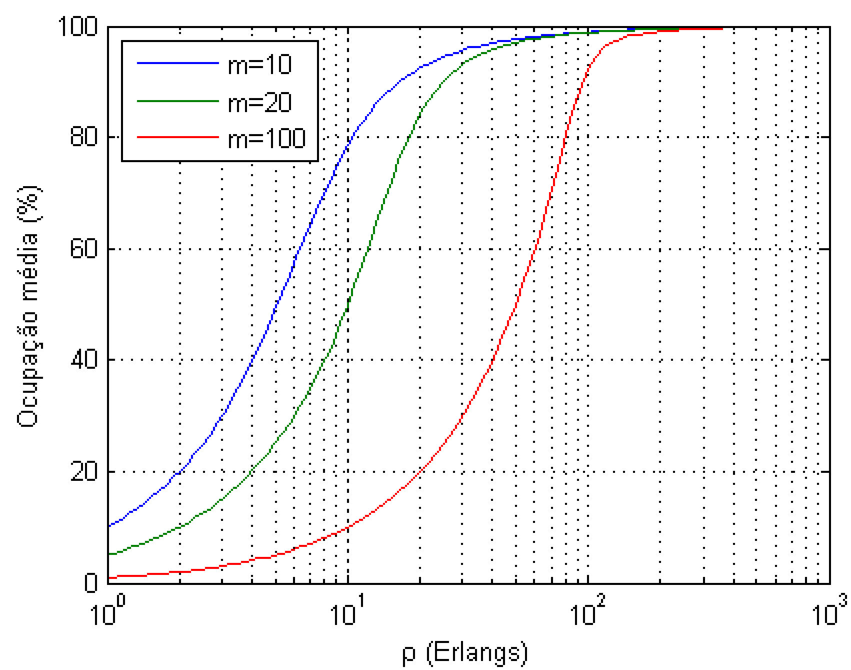


$$\pi_0 = \frac{1}{1+\sum_{i=1}^{m} \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}}$$

$$\pi_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\cdot\pi_0,\ \ n\geq 1$$

A ocupação média da ligação é ($\rho = \lambda/\mu$):

$$\sum_{i=0}^{m}\left(i\times\pi_i\right) = \frac{\sum_{i=1}^{m}\frac{\rho^i}{(i-1)!}}{\sum_{i=0}^{m}\frac{\rho^i}{i!}}$$



Recurso de comunicações uni-serviço
Distribuição de Engset

Considere um recurso de comunicações com capacidade para N circuitos e que serve M clientes ($M > N$).

Cada um dos M clientes está inativo durante um período de tempo exponencialmente distribuído com média $1/\lambda$ e gera uma chamada com uma duração média $1/\mu$.

A cada chamada pedida é atribuído um dos N circuitos disponíveis; se não houver qualquer circuito disponível a chamada é bloqueada.

Este sistema pode ser modelado por um processo de nascimento e morte com os estados $n = 0, 1, \ldots, N$ (representando o número de circuitos ocupados) e com taxas de nascimento e de morte dadas por:

$$\lambda_n = (M - n)\lambda, \qquad 0 \leq n \leq N - 1$$

$$\mu_n = n\mu, \qquad 1 \leq n \leq N$$

Recurso de comunicações uni-serviço
Distribuição de Engset

A probabilidade de n chamadas no sistema é

$$\pi_n = \frac{\binom{M}{n}\left(\frac{\lambda}{\mu}\right)^n}{\sum_{n=0}^{N}\binom{M}{n}\left(\frac{\lambda}{\mu}\right)^n}$$

A probabilidade de bloqueio é

$$P_B = \frac{\binom{M-1}{N}\left(\frac{\lambda}{\mu}\right)^N}{\sum_{n=0}^{N}\binom{M-1}{n}\left(\frac{\lambda}{\mu}\right)^n}$$

Esta é a chamada *distribuição de Engset*.

Note-se que a propriedade PASTA não se verifica ($\pi_n \neq P_B$) dado que a taxa de chegada de chamadas não é estatisticamente independente do estado do sistema.

Esta fórmula degenera na fórmula de ErlangB, quando $M \to \infty$ $\lambda \to 0$, com $M\lambda$ constante.

Estabelecimento de circuitos multi-serviço

Considere um recurso de comunicações com C circuitos que serve as classes de serviço $k = 1, 2, \ldots, K$ (K é o número de classes de serviço).

A cada classe k está associada uma taxa de chegada, λ_k, um tempo médio de serviço, $1/\mu_k$ e uma largura de banda b_k (em numero de circuitos):

- As chamadas das K classes chegam de acordo com processos independentes de Poisson à taxa λ_k.
- Uma chamada da classe k que tenha sido admitida pelo sistema, ocupa b_k circuitos durante o tempo de serviço da chamada, o qual é exponencialmente distribuído com média $1/\mu_k$.

Seja ρ_k a intensidade de tráfego de cada classe, isto é, $\rho_k = \lambda_k / \mu_k$.

Assuma-se que os tempos de serviço das chamadas são independentes entre si e independentes dos processos de chegadas.

Estabelecimento de circuitos multi-serviço

Seja n_k o número de chamadas da classe k no sistema.

Considerem-se os vetores $\mathbf{n} = (n_1, \ldots, n_k)$ e $\mathbf{b} = (b_1, \ldots, b_k)$.

O número total de circuitos ocupados no sistema é:

$$\mathbf{b} \cdot \mathbf{n} = \sum_{k=1}^{K} b_k n_k$$

Uma chamada da classe k é admitida no sistema se $b_k \leq C - \mathbf{b} \cdot \mathbf{n}$; caso contrário é bloqueada e perdida.

O espaço de estados do processo de nascimento e morte multidimensional é:

$$S = \left\{ \mathbf{n} \in I^K : \mathbf{b} \cdot \mathbf{n} \leq C \right\}$$

onde I é o conjunto dos inteiros não negativos.

Seja S_k o subconjunto dos estados para os quais uma chamada da classe k é admitida quando chega à rede, isto é,

$$S_k = \left\{ \mathbf{n} \in S : \mathbf{b} \cdot \mathbf{n} \leq C - b_k \right\}$$

Estabelecimento de circuitos multi-serviço

As probabilidades limite de cada estado são dadas por:

$$P(\mathbf{n}) = \frac{1}{G} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!} \qquad \mathbf{n} \in S$$

onde $$G = \sum_{\mathbf{n} \in S} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}$$

e a probabilidade de bloqueio da classe *k* por:

$$B_k = 1 - \frac{\sum_{\mathbf{n} \in S_k} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}{\sum_{\mathbf{n} \in S} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}} \quad \Leftrightarrow \quad B_k = \frac{\sum_{\mathbf{n} \in S \setminus S_k} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}{\sum_{\mathbf{n} \in S} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}$$

- Primeira expressão: 1 menos a probabilidade dos estados S_k (estados para os quais uma chamada da classe *k* é admitida).
- Segunda expressão: probabilidade dos estados $S \setminus S_k$ (estados para os quais uma chamada da classe *k* é rejeitada).

Este sistema é por vezes designado de *stochastic knapsack*.

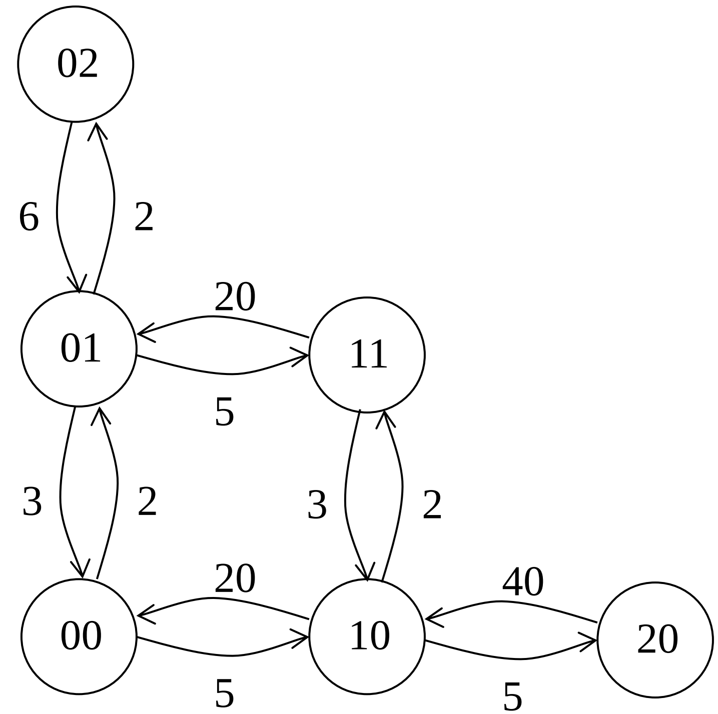
Exemplo 1

Considere uma rede telefónica VoIP de uma empresa em que o número de chamadas VoIP para o exterior é no máximo 2. Existem dois tipos de chamadas VoIP: chamadas pessoais e chamadas de apoio ao cliente.

O tráfego VoIP de/para a rede pública é um processo de Poisson com taxa de 5 chamadas por hora para chamadas pessoais e de 2 chamadas por hora para as chamadas de apoio ao cliente.

As chamadas pessoais têm duração exponencial com média de 3 minutos e as chamadas de apoio ao cliente têm duração exponencial com média de 20 minutos.

Determine a probabilidade de bloqueio de cada tipo de chamada VoIP.



Recurso de comunicações com comutação de pacotes

Diferentes fluxos de pacotes podem ser

(1) suportados por um circuito - *multiplexagem estatística*,

(2) suportados por diferentes circuitos - *multiplexagem determinística*.

Tal como nos fluxos de chamadas, uma *classe de serviço* identifica um conjunto de fluxos de pacotes com *as mesmas características de tráfego* e *os mesmos requisitos de qualidade de serviço*.

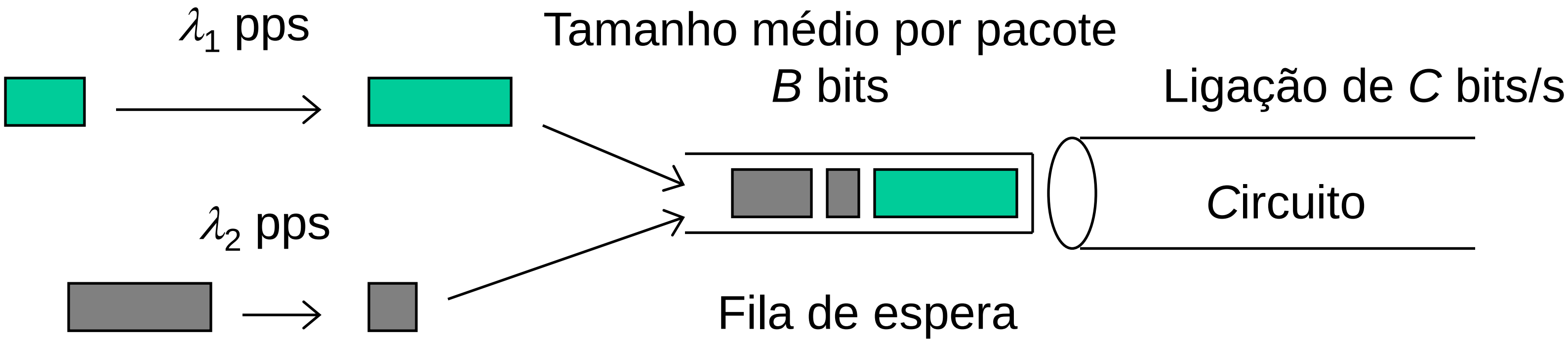
As características de tráfego são determinadas por

(i) descrição estocástica do processo de chegada dos pacotes, e

(ii) descrição estocástica do tamanho dos pacotes.

Um dos parâmetro de qualidade de serviço mais importantes é o *atraso médio por pacote* (existem outros como, por exemplo, a variância do atraso ou a percentagem de perda de pacotes).

Multiplexagem estatística de fluxos de pacotes



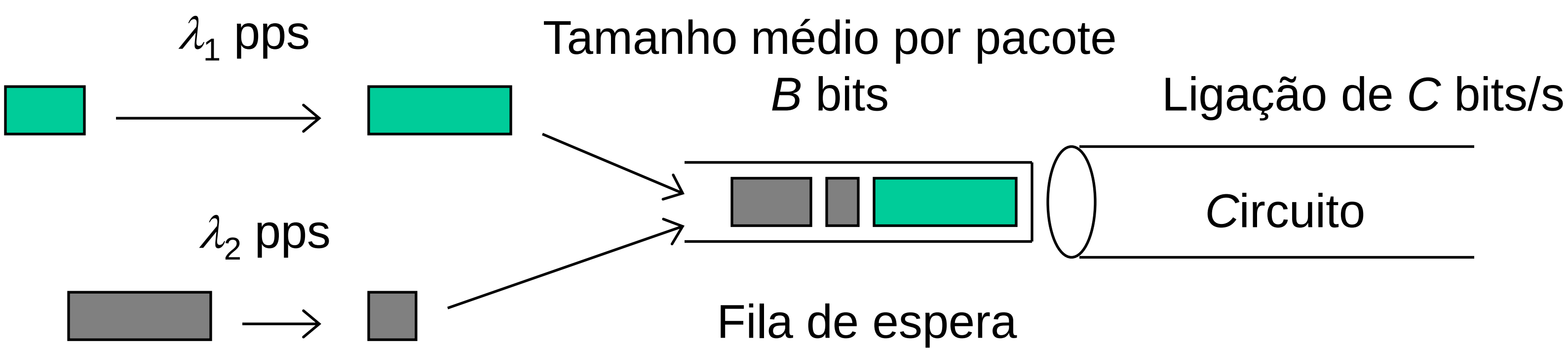
Considere-se que:
(i) as chegadas de pacotes são processos de Poisson,
(ii) o tamanho dos pacote é exponencialmente distribuído,
(iii) a fila de espera é atendida com uma disciplina *First-In-First-Out*.

Se a fila de espera for de tamanho infinito, este sistema é modelado pelo sistema de filas de espera *M/M/1* em que

$\lambda = \lambda_1 + \lambda_2$ pps (pacotes por segundo)
$\mu = C/B$ pps

Atraso médio dos pacotes: $$W = \frac{1}{\mu - \lambda} = \frac{1}{C/B - (\lambda_1 + \lambda_2)}$$

Multiplexagem estatística de fluxos de pacotes



Se a fila de espera for finita com capacidade para *m* pacotes, este sistema é modelado pelo sistema de filas de espera *M/M/1/m*

Percentagem de perda de pacotes: $$\mu_m = \frac{(\lambda/\mu)^m}{\sum_{i=0}^{m}(\lambda/\mu)^m}$$

Atraso médio dos pacotes: $$W = \frac{L}{\lambda(1-\mu_m)} = \frac{\sum_{i=0}^{m} i \times \pi_i}{\lambda(1-\mu_m)}$$

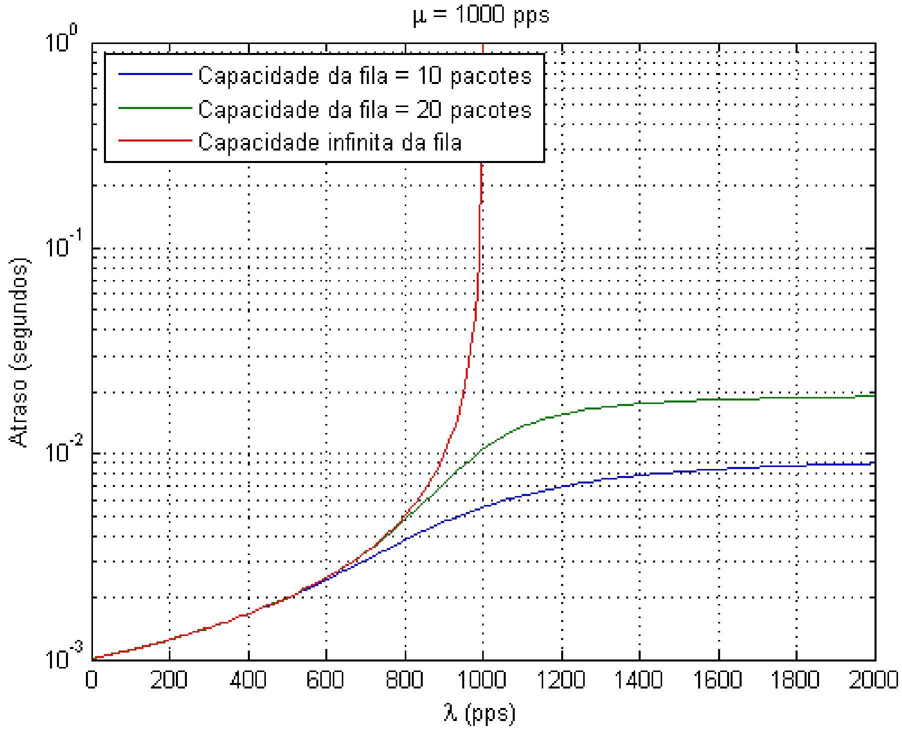
Teorema de Little

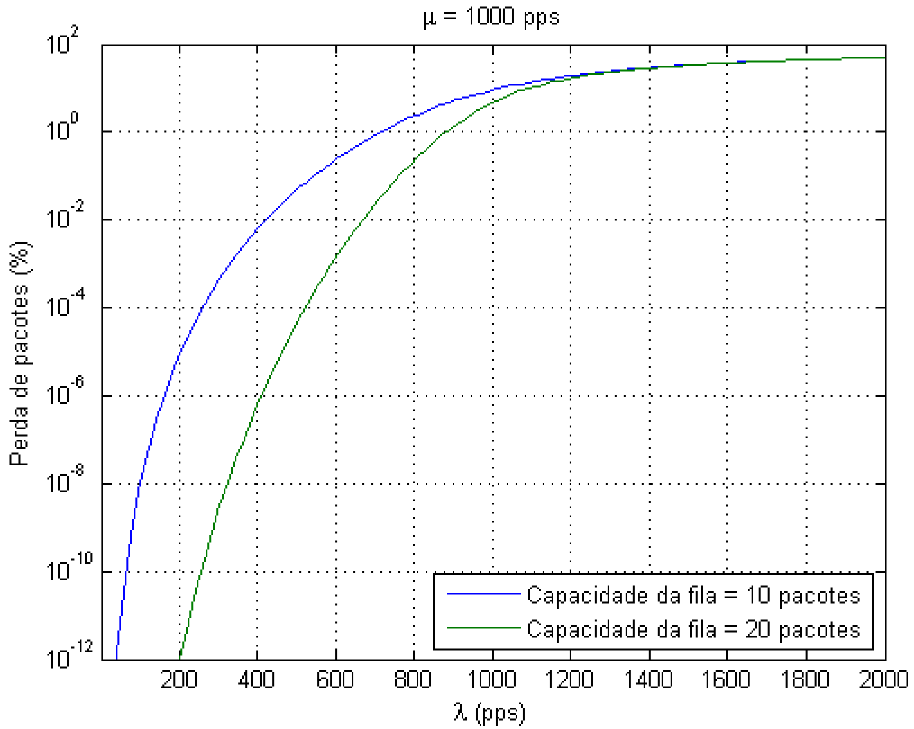
$$= \frac{1}{\lambda(1-\mu_m)} \times \frac{\sum_{i=0}^{m} i \times (\lambda/\mu)^i}{\sum_{i=0}^{m}(\lambda/\mu)^i}$$

Multiplexagem estatística de fluxos de pacotes

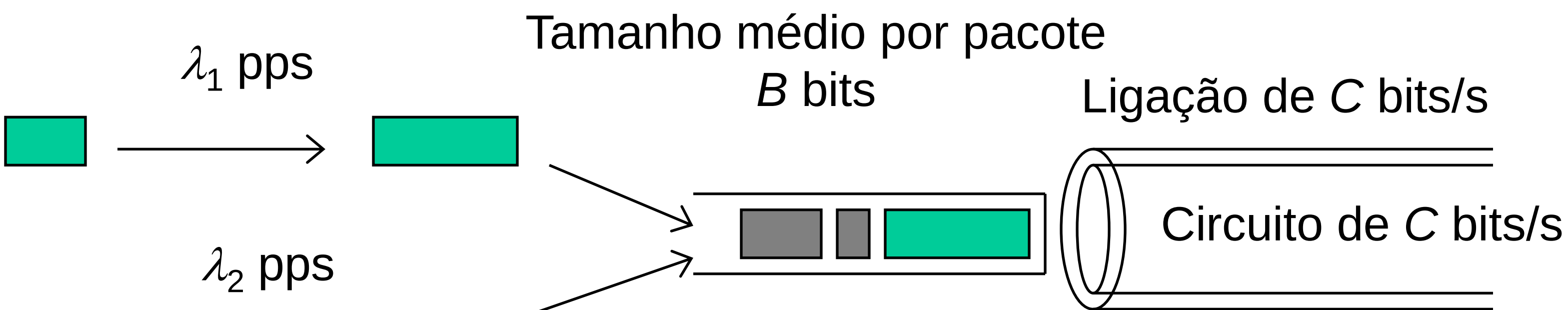
- Se a fila de espera for de tamanho infinito, sistema modelado por *M/M/1*
- Se a fila de espera tiver capacidade para *m* pacotes, sistema modelado por *M/M/1/m*

- Exemplo:
  - ligação de 10 Mbps e tamanho médio de pacotes de 1250 Bytes
  - $\mu = 10^7/(1250\times 8) = 1000$ pps

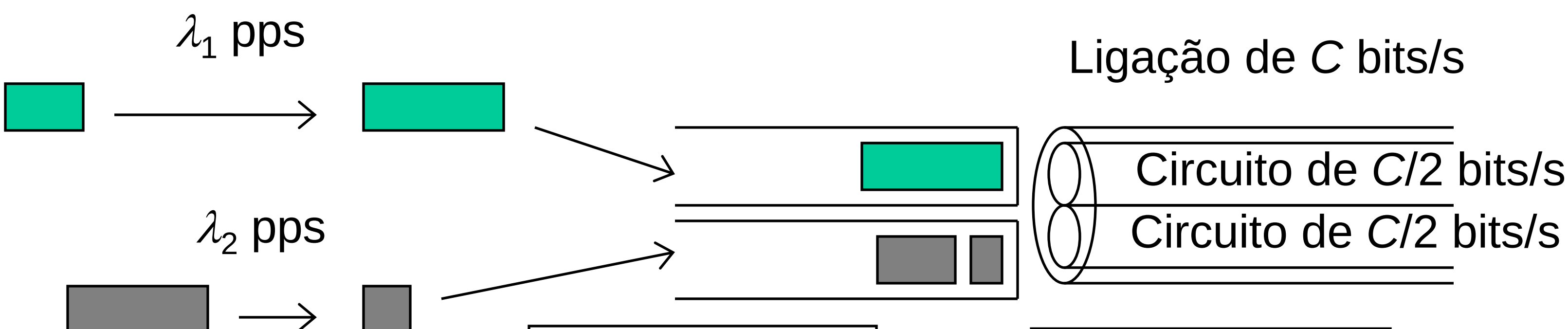




Multiplexagem Estatística vs. Multiplexagem Determinística



$$W = \frac{1}{C/B - (\lambda_1 + \lambda_2)}$$



$$W_1 = \frac{1}{C/2B - \lambda_1}$$

$$W_2 = \frac{1}{C/2B - \lambda_2}$$

Exemplo 2

Considere um sistema de transmissão de pacotes constituído por 2 canais de 64 Kbps com uma fila de espera muito grande. Este sistema suporta dois fluxos de pacotes: fluxo A de 1 pacote/segundo e fluxo B de 25 pacotes/segundo. Ambos os fluxos geram pacotes de tamanho exponencialmente distribuído com média de 250 bytes. Calcule o atraso médio por pacotes de cada fluxo quando:

(a) Ambos os fluxos são multiplexados estatisticamente na capacidade total do sistema de transmissão.

(b) Cada fluxo é encaminhado separadamente por cada um dos canais do sistema de transmissão.

Multiplexagem Estatística vs. Multiplexagem Determinística

- Em geral, a multiplexagem estatística conduz a atrasos médios por pacote inferiores:

  Na multiplexagem determinística, um circuito dedicado a um fluxo e que esteja momentaneamente livre não pode ser usado pelos outros fluxos.

- No entanto, a multiplexagem determinística conduz a variâncias de atraso menores:

  Na multiplexagem estatística, o número médio de clientes em espera é maior.

Disciplina com prioridades

Na multiplexagem estatística, os pacotes de cada fluxo podem ser tratados (i.e., transmitidos) de forma diferenciada.

Uma possibilidade é atribuir prioridades aos fluxos, de tal forma que os pacotes de um fluxo com maior prioridade são sempre transmitidos antes do que os pacotes de um fluxo com menor prioridade (esquema designado de *prioritização estrita*).

O sistema M/G/1 com prioridades pode ser utilizado para modelar este sistema.

Considere um sistema *M*/*G*/1 em que existem *n* classes de serviço.

A *k*-ésima classe é determinada por:

(1) taxa de chegadas: λ_k

(2) 1º e 2º momentos do tempo de serviço: $E(S_k) = \dfrac{1}{\mu_k}$ $\quad E(S_k^2)$

(3) prioridade: *k* (quanto menor este valor, maior a prioridade)

Sistema *M/G/1* com prioridades

O sistema transmite primeiro os pacotes das classes com maior prioridade.

Os pacotes de uma mesma classe são transmitidos por ordem de chegada (disciplina *FIFO - First In First Out*).

Considera-se que as chegadas dos pacotes de cada classe são independentes e de Poisson e independentes dos tempos de transmissão.

A transmissão de um pacote não é interrompida pela chegada de um pacote de uma classe com maior prioridade (disciplina de serviço designada por *não-preemptiva*).

O <u>*atraso médio na fila de espera*</u> correspondente aos pacotes da classe *k* é dado por:

$$W_{Qk} = \frac{\sum_{i=1}^{n} \lambda_i E\left(S_i^2\right)}{2(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)}$$

$$\rho_k = \lambda_k / \mu_k$$

$$\rho_1 + \ldots + \rho_n < 1$$

Disciplina “*Fair Queuing*”

- Na disciplina FIFO, um fluxo de pacotes que decida momentaneamente aumentar a sua taxa de transmissão captura uma fração arbitrariamente grande dos recursos do sistema.
- Na disciplina com prioridades, enquanto a fila de espera de um fluxo de maior prioridade contiver pacotes, os fluxos de menor prioridade nunca serão servidos.

Uma disciplina que permite evitar estas desvantagens é o <u>*fair queuing*</u>. Nesta disciplina, existe uma fila de espera separada para cada fluxo de pacotes e as filas de espera são servidas proporcionalmente aos pesos que lhe foram atribuídos.

Assim, se a fonte de um fluxo envia pacotes a uma taxa demasiado rápida, a sua fila de espera enche sem afetar os restantes fluxos.

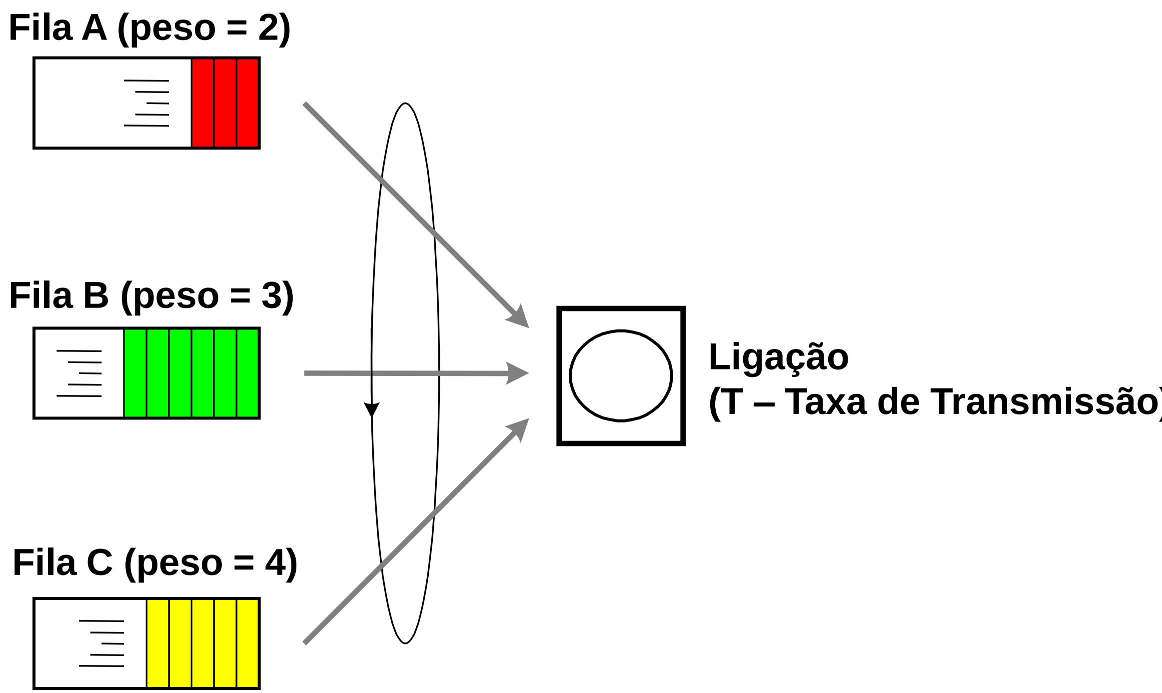
Weighted Fair Queuing (WFQ)

Este algoritmo garante que cada fila de espera consegue uma percentagem da taxa de transmissão do recurso de comunicações pelo menos igual ao seu peso a dividir pela soma dos pesos de todas as filas

$$R_A = \frac{2}{2+3+4}T$$

$$R_B = \frac{3}{2+3+4}T$$

$$R_C = \frac{4}{2+3+4}T$$



Exemplo 3

Considere um sistema de transmissão de pacotes com capacidade de 120 Kbps. Este sistema suporta dois fluxos de tráfego: fluxo A de 1 pacote/segundo de tamanho exponencialmente distribuído com média de 250 bytes; fluxo B de 5 pacotes/segundo de tamanho exponencialmente distribuído com média de 125 bytes. O fluxo B tem um requisito de qualidade de serviço que exige que o atraso médio por pacote seja 10 milissegundos. Pretende-se implementar um esquema de multiplexagem determinística que permita cumprir com a qualidade de serviço do fluxo B. Calcule:

(a) a percentagem da capacidade de transmissão que deve ser dedicada ao fluxo B.

(b) o atraso médio por pacote do fluxo A no sistema resultante.

Exemplo 4

Considere um sistema de transmissão de pacotes com capacidade de 120 Kbps. Este sistema suporta dois fluxos de tráfego: fluxo A de 1 pacote/segundo de tamanho exponencialmente distribuído com média de 250 bytes; fluxo B de 5 pacotes/segundo de tamanho exponencialmente distribuído com média de 125 bytes.

O sistema atribui maior prioridade ao fluxo B com uma disciplina não preemptiva. Calcule o atraso médio por pacote de cada fluxo.