Partilha de Recursos de uma Ligação Ponto-a-Ponto

Desempenho e Dimensionamento de Redes

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

DETI-UA, 2020/2021

Recursos de uma ligação digital ponto-a-ponto

Uma ligação digital é um sistema de transmissão que permite a troca de informação em formato digital.

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

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-streaming 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.

Recursos de uma ligação estruturada em circuitos

Uma ligação pode ser explicitamente estruturada em *<u>circuitos</u>* (ou *<u>canais</u>*), que correspondem a partições da capacidade total da ligação.

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

  (1) multiplexagem no tempo (TDM – *Time Division Multiplexing*)

  (2) multiplexagem na frequência (FDM – *Frequency Division Multiplexing*)

  (3) multiplexagem de comprimentos de onda em redes óticas (WDM – *Wavelength Division Multiplexing*)

Uma ligação pode também ser implicitamente usada como se um fosse estruturada em circuitos (noção de circuitos virtuais).

- Cada comunicação usa uma partição da capacidade da ligação.

Partilha de recursos de uma ligação estruturada em circuitos

- Uma *chamada*, quando é estabelecida, corresponde à atribuição temporária de um ou mais circuitos.
- Após estabelecida, a chamada dura um tempo finito após o qual a largura de banda dos circuitos atribuídos é libertada e fica disponível para pedidos futuros de chamadas.
- Assim, uma ligação pode ser partilhada 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 de uma classe de serviço são determinadas:

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

  (ii) pela descrição estatística da duração das chamadas, e

  (iii) pelos recursos (em número de circuitos) requeridos por cada chamada.

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

- 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 entre diferentes classes de serviço faz-se através da função de controle de fluxos (também designada de controlo de admissão de chamadas)

Recurso de comunicações uni-serviço - Distribuição de ErlangB

Um recurso de comunicações:

(i) estruturado por um grupo de *m* circuitos

(ii) ao qual é oferecido um *fluxo de chamadas de Poisson* com taxa λ

(iii) em que cada chamada requer um circuito

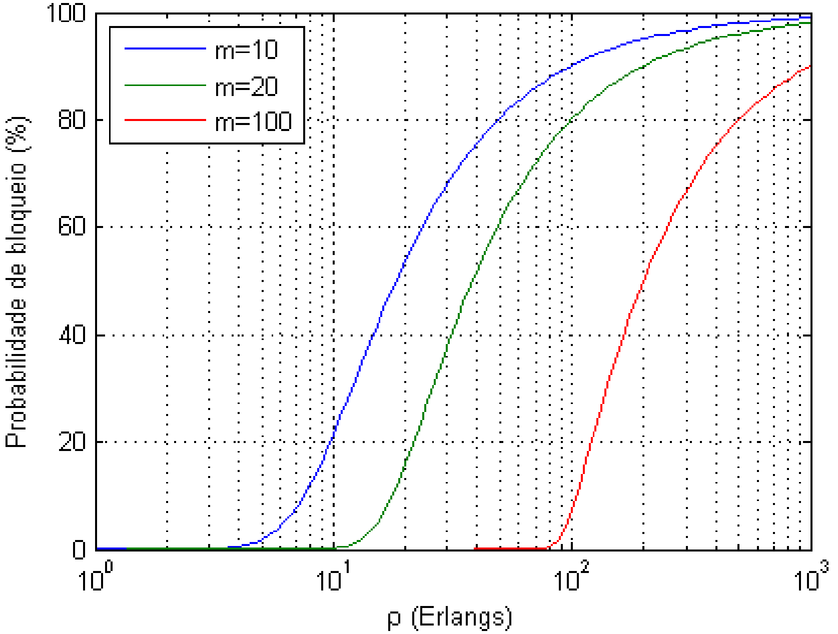
(iv) e *a duração das chamadas é exponencialmente distribuída* com média $1/\mu$

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.

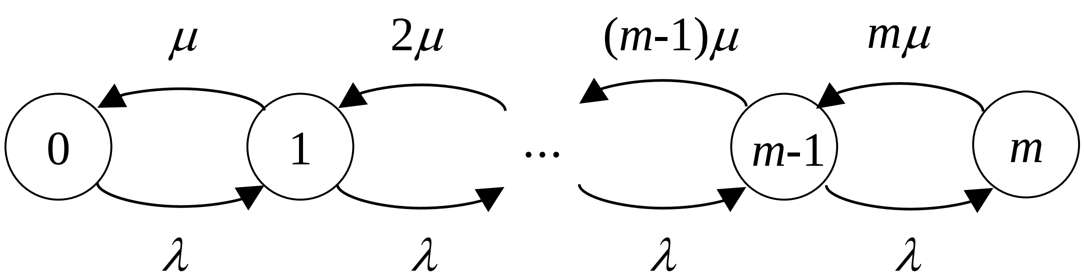
A fórmula de ErlangB, dá a probabilidade de bloqueio:

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



Estabelecimento de circuitos uni-serviço

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

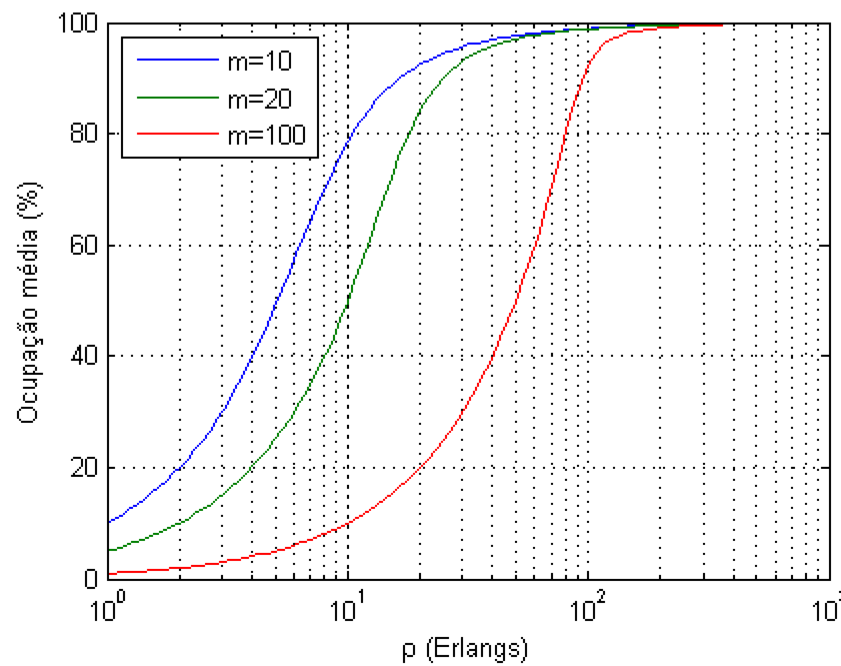


$$\pi_0 = \frac{1}{1+\sum_{i=1}^{m} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}}$$

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

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

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

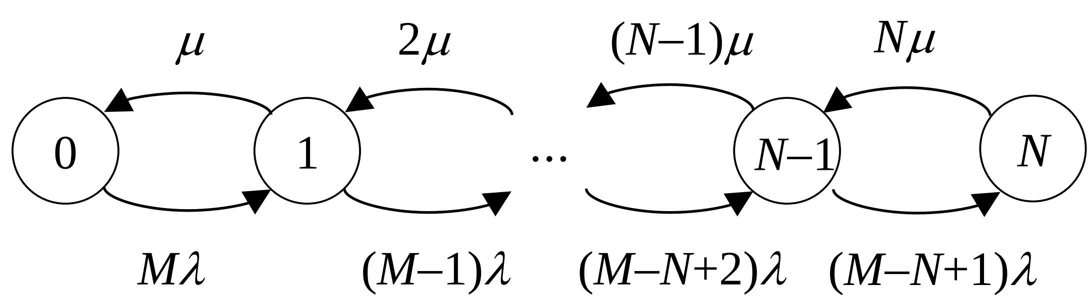


Recurso de comunicações uni-serviço
Distribuição de Engset

- Considere-se uma ligação estruturada em 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 é modelado por um processo de nascimento e morte com os estados n = 0, 1, …, N (representando o número de circuitos ocupados) e com taxas de nascimento e de morte dadas por:

$$\lambda_n = (M-n)\lambda, \quad 0 \le n \le N-1$$

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



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 depende do estado do sistema (*i.e.*, 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 uma ligação com C circuitos que serve as classes de serviço k = 1, 2, …, 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 número 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 **n** = (n_1, …, n_k) e **b** = (b_1, …, 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 – **b·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:

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

- 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 $S\backslash S_k$ (estados para os quais uma chamada da classe *k* é rejeitada).

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

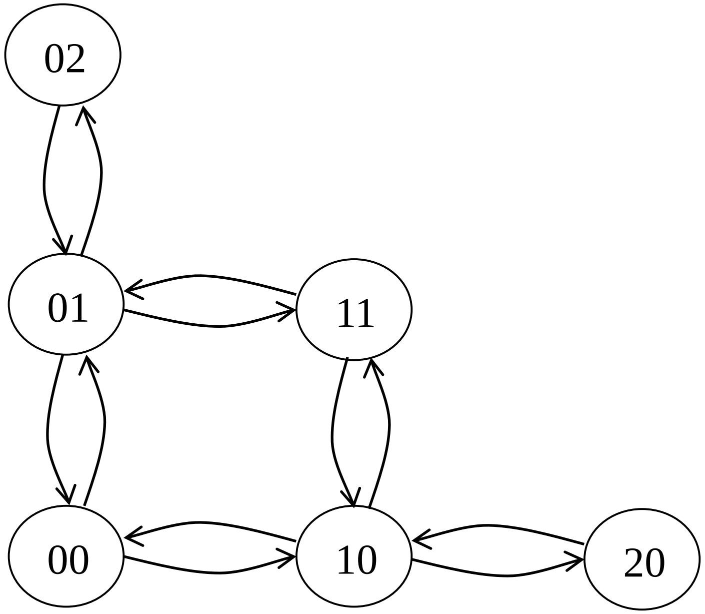
Exemplo 1

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 (tipo 1) e chamadas de apoio ao cliente (tipo 2).

O tráfego VoIP de/para a rede pública é um processo de Poisson com taxa de 5 chamadas/hora para chamadas do tipo 1 e de 2 chamadas/hora para as chamadas do tipo 2.

As chamadas do tipo 1 têm duração exponencial com média de 3 minutos e as chamadas do tipo 2 têm duração exponencial com média de 20 minutos.

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



$\rho_1 = \lambda_1/\mu_1 = 5/60\times3 = 1/4$ Erlangs

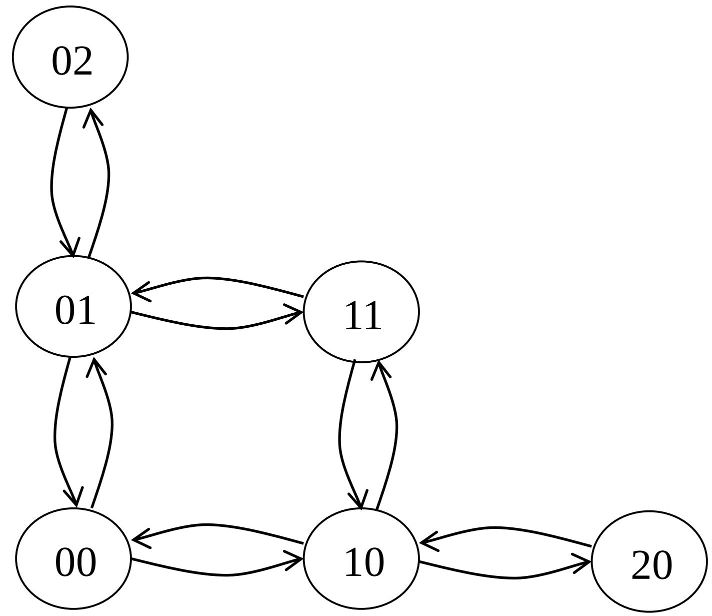
$\rho_2 = \lambda_2/\mu_2 = 2/60\times20 = 2/3$ Erlangs

Exemplo 1 - Resolução

$\rho_1 = \lambda_1/\mu_1 = 5/60\times 3 = 1/4$ Erlangs

$\rho_2 = \lambda_2/\mu_2 = 2/60\times 20 = 2/3$ Erlangs

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



$$B_1 = \frac{\frac{(1/4)^2(2/3)^0}{2!\,0!} + \frac{(1/4)^1(2/3)^1}{1!\,1!} + \frac{(1/4)^0(2/3)^2}{0!\,2!}}{\frac{(1/4)^0(2/3)^0}{0!\,0!} + \frac{(1/4)^1(2/3)^0}{1!\,0!} + \frac{(1/4)^2(2/3)^0}{2!\,0!} + \frac{(1/4)^0(2/3)^1}{0!\,1!} + \frac{(1/4)^1(2/3)^1}{1!\,1!} + \frac{(1/4)^0(2/3)^2}{0!\,2!}}$$

$$B_1 = \frac{\frac{(1/4)^2}{2} + 1/4 \times 2/3 + \frac{(2/3)^2}{2}}{1 + \frac{1}{4} + \frac{(1/4)^2}{2} + 2/3 + 1/4 \times 2/3 + \frac{(2/3)^2}{2}} = 0.1798 = 17.98\% \qquad B_2 = B_1 = 17.98\%$$

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*.

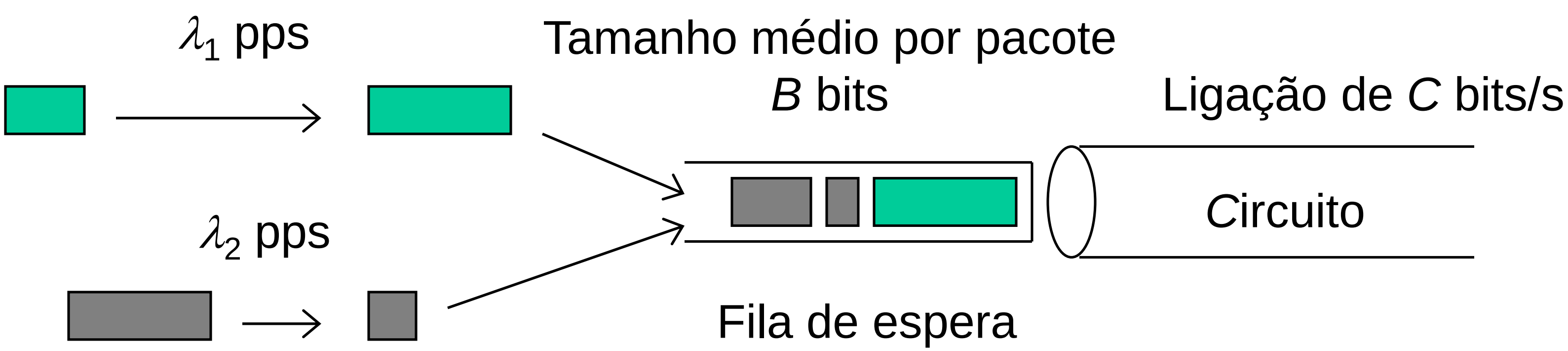
As características de tráfego são determinadas por

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

(ii) descrição estatística do tamanho dos pacotes.

Parâmetros de qualidade de serviço mais importantes são: o *atraso médio por pacote* e a taxa de perda de pacotes.

Multiplexagem estatística de fluxos de pacotes



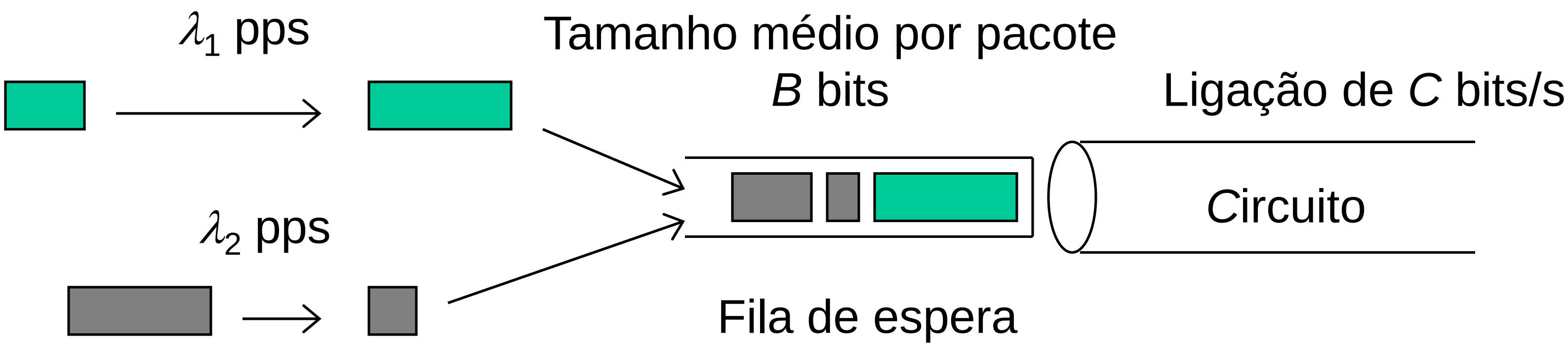
Considere-se que:
(i) as chegadas de pacotes são processos de Poisson,
(ii) o tamanho dos pacote é exponencialmente distribuído,
(iii) a fila de espera é atendida com uma disciplina *First-In-First-Out*.

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

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

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*–1 pacotes, este sistema é modelado pelo sistema de filas de espera *M/M/1/m*:

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

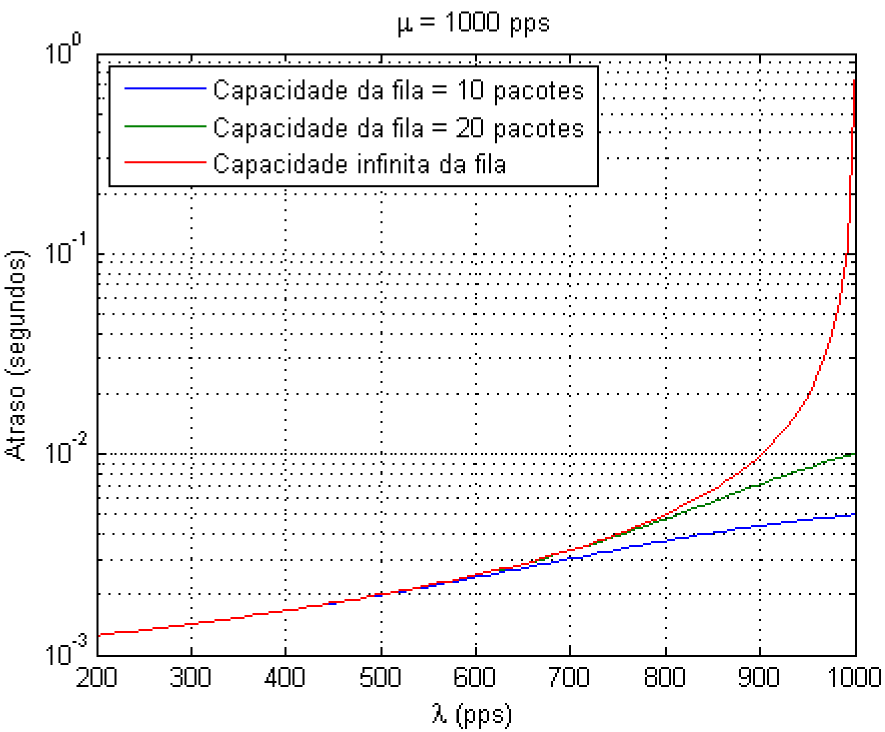
Número médio de pacotes no sistema: $L = \sum_{i=0}^{m} i \times \pi_i = \dfrac{\sum_{i=0}^{m} i \times (\lambda/\mu)^i}{\sum_{j=0}^{m}(\lambda/\mu)^j}$

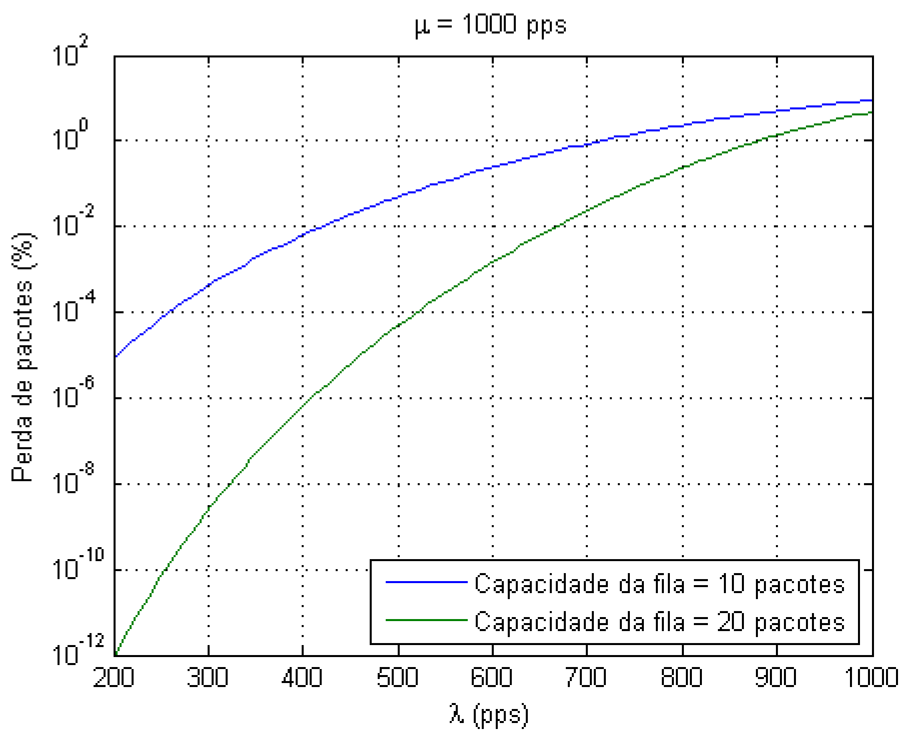
Atraso médio dos pacotes: $W = \dfrac{L}{\lambda(1-\mu_m)}$

Multiplexagem estatística de fluxos de pacotes

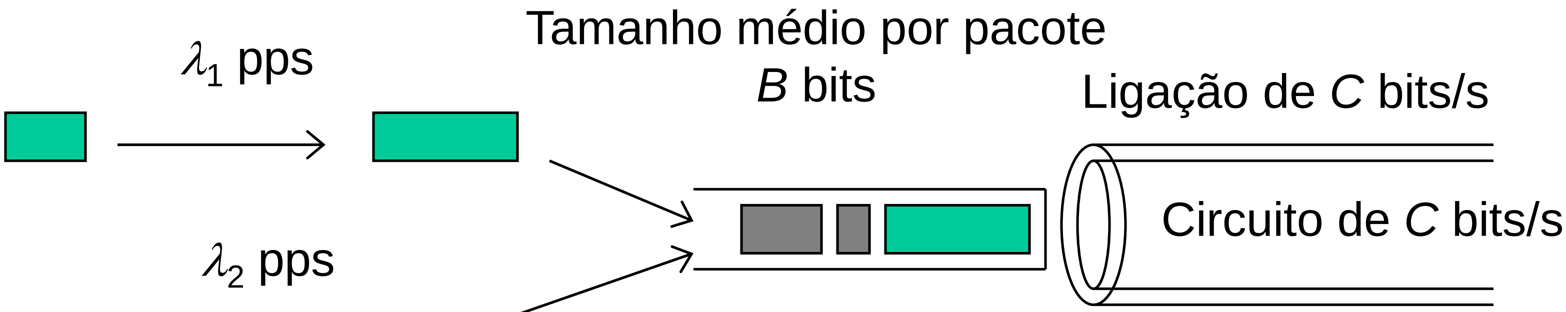
- Se a fila de espera for de tamanho infinito, sistema modelado por *M/M/1*
- Se a fila de espera tiver capacidade para *m*–1 pacotes, sistema modelado por *M/M/1/m*

- Exemplo:
  - ligação de 10 Mbps e tamanho médio de pacotes de 1250 Bytes
  - $\mu = 10^7/(1250\times8) = 1000$ pps

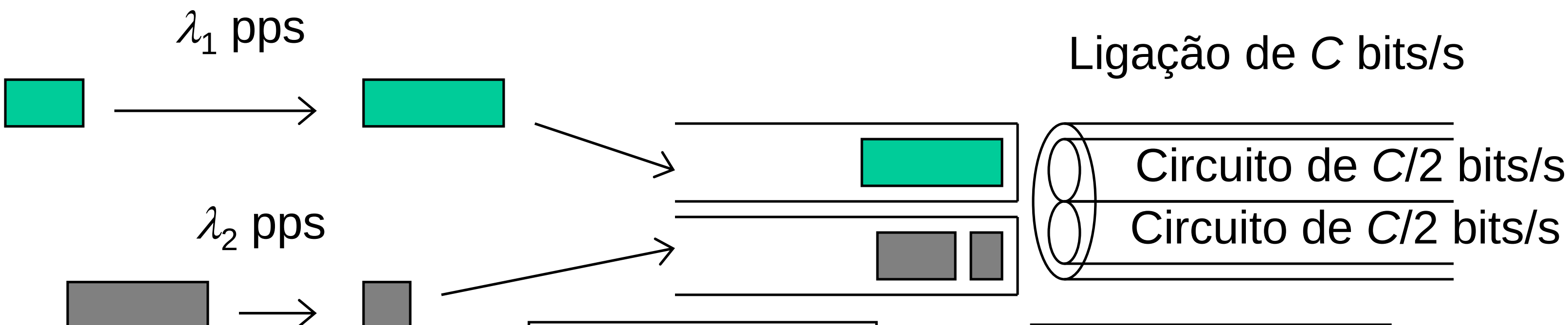




Multiplexagem Estatística vs. Multiplexagem Determinística



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



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

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

Exemplo 2

Considere uma ligação ponto-a-ponto de 10 Mbps que suporta dois fluxos de pacotes: fluxo A de 1 Mbps e fluxo B de 4 Mbps. Ambos os fluxos geram pacotes de tamanho exponencialmente distribuído com média de 1000 bytes. Calcule o atraso médio por pacote de cada fluxo quando:

(a) os fluxos são multiplexados estatisticamente na capacidade total da ligação;

(b) os fluxos são multiplexados deterministicamente em que é atribuída metade da capacidade da ligação a cada fluxo.

Exemplo 2 – resolução

Considere uma ligação ponto-a-ponto de 10 Mbps que suporta dois fluxos de pacotes: fluxo A de 1 Mbps e fluxo B de 4 Mbps. Ambos os fluxos geram pacotes de tamanho exponencialmente distribuído com média de 1000 bytes. Calcule o atraso médio por pacote de cada fluxo quando:

(a) os fluxos são multiplexados estatisticamente na capacidade total da ligação;

$$\mu = \frac{10 \times 10^6}{8 \times 1000} = 1250 \text{ pps} \qquad \lambda_A = \frac{1 \times 10^6}{8 \times 1000} = 125 \text{ pps} \qquad \lambda_B = \frac{4 \times 10^6}{8 \times 1000} = 500 \text{ pps}$$

$$W_A = W_B = \frac{1}{\mu - (\lambda_A + \lambda_B)} = \frac{1}{1250 - (125 + 500)} = 1.6 \times 10^{-3} = 1.6 \text{ ms}$$

(b) os fluxos são multiplexados deterministicamente em que é atribuída metade da capacidade da ligação a cada fluxo.

$$W_A = \frac{1}{\frac{\mu}{2} - \lambda_A} = \frac{1}{\frac{1250}{2} - 125} = 2 \text{ ms} \qquad W_B = \frac{1}{\frac{\mu}{2} - \lambda_B} = \frac{1}{\frac{1250}{2} - 500} = 8 \text{ ms}$$

Multiplexagem Estatística vs. Multiplexagem Determinística

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

Na multiplexagem determinística, a capacidade atribuída 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, todos os fluxos partilham uma única fila de espera.

Assim, a diferença entre o número mínimo e o número máximo de pacotes na fila de espera é maior.

Disciplina com prioridades

Na multiplexagem estatística, os pacotes de cada fluxo não 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 *atraso médio na fila de espera* correspondente aos pacotes da classe k é dado por:

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

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

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

Exemplo 3

Considere uma ligação ponto-a-ponto de 10 Mbps que suporta dois fluxos de pacotes: fluxo A de 0.5 Mbps e fluxo B de 4 Mbps. Ambos os fluxos geram pacotes de tamanho exponencialmente distribuído com média de 1000 bytes. O atraso médio por pacote do fluxo A deverá ser de 2 milissegundos. Pretende-se implementar um esquema de multiplexagem determinística que permita cumprir com a qualidade de serviço do fluxo A. Calcule:

(a) a capacidade mínima da ligação que deve ser atribuída ao fluxo A;

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

Exemplo 3 – resolução

Considere uma ligação ponto-a-ponto de 10 Mbps que suporta dois fluxos de pacotes: fluxo A de 0.5 Mbps e fluxo B de 4 Mbps. Ambos os fluxos geram pacotes de tamanho exponencialmente distribuído com média de 1000 bytes. O atraso médio por pacote do fluxo A deverá ser de 2 milissegundos. Pretende-se implementar um esquema de multiplexagem determinística que permita cumprir com a qualidade de serviço do fluxo A. Calcule:

(a) a capacidade mínima da ligação que deve ser atribuída ao fluxo A;

$$\mu = \frac{10 \times 10^6}{8 \times 1000} = 1250 \text{ pps} \qquad \lambda_A = \frac{0.5 \times 10^6}{8 \times 1000} = 62.5 \text{ pps} \qquad \lambda_B = \frac{4 \times 10^6}{8 \times 1000} = 500 \text{ pps}$$

$$W_A = \frac{1}{\mu_A - \lambda_A} = \frac{1}{\mu_A - 62.5} = 2 \times 10^{-3} \Leftrightarrow$$

$$\Leftrightarrow \mu_A = \frac{1}{2 \times 10^{-3}} + 62.5 = 562.5 \text{ pps} = 562.5 \times (8 \times 1000) = 4.5 \times 10^6 = 4.5 \text{ Mbps}$$

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

$$W_B = \frac{1}{\mu_B - \lambda_B} = \frac{1}{(1250 - 562.5) - 500} = 0.0053 = 5.3 \text{ ms}$$

Exemplo 4

Considere uma ligação ponto-a-ponto de 10 Mbps que suporta dois fluxos de pacotes: fluxo A de 0.5 Mbps e fluxo B de 4 Mbps. Ambos os fluxos geram pacotes de tamanho exponencialmente distribuído com média de 1000 bytes.

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

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

$$\lambda_A = \frac{0.5 \times 10^6}{8 \times 1000} = 62.5 \text{ pps}$$

$$\lambda_B = \frac{4 \times 10^6}{8 \times 1000} = 500 \text{ pps}$$

$$\mu_A = \mu_B = \mu = \frac{10 \times 10^6}{8 \times 1000} = 1250 \text{ pps}$$

$$E(S_A) = E(S_B) = \frac{1}{\mu} = \frac{1}{1250} \text{ seg.}$$

$$E\left({S_A}^2\right) = E\left({S_B}^2\right) = \frac{2}{\mu^2} = \frac{2}{1250^2} \text{ seg.}^2$$

$$W_A = \frac{\lambda_A \times E\left({S_A}^2\right) + \lambda_B \times E\left({S_B}^2\right)}{2 \times \left(1 - \frac{\lambda_A}{\mu_A}\right)} + E(S_A) = 1.2 \text{ ms}$$

$$W_B = \frac{\lambda_A \times E\left({S_A}^2\right) + \lambda_B \times E\left({S_B}^2\right)}{2 \times \left(1 - \frac{\lambda_A}{\mu_A}\right) \times \left(1 - \frac{\lambda_A}{\mu_A} - \frac{\lambda_B}{\mu_B}\right)} + E(S_B) = 1.5 \text{ ms}$$