Desempenho de Redes com Comutação de Pacotes

Modelação e Desempenho de Redes e Serviços

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

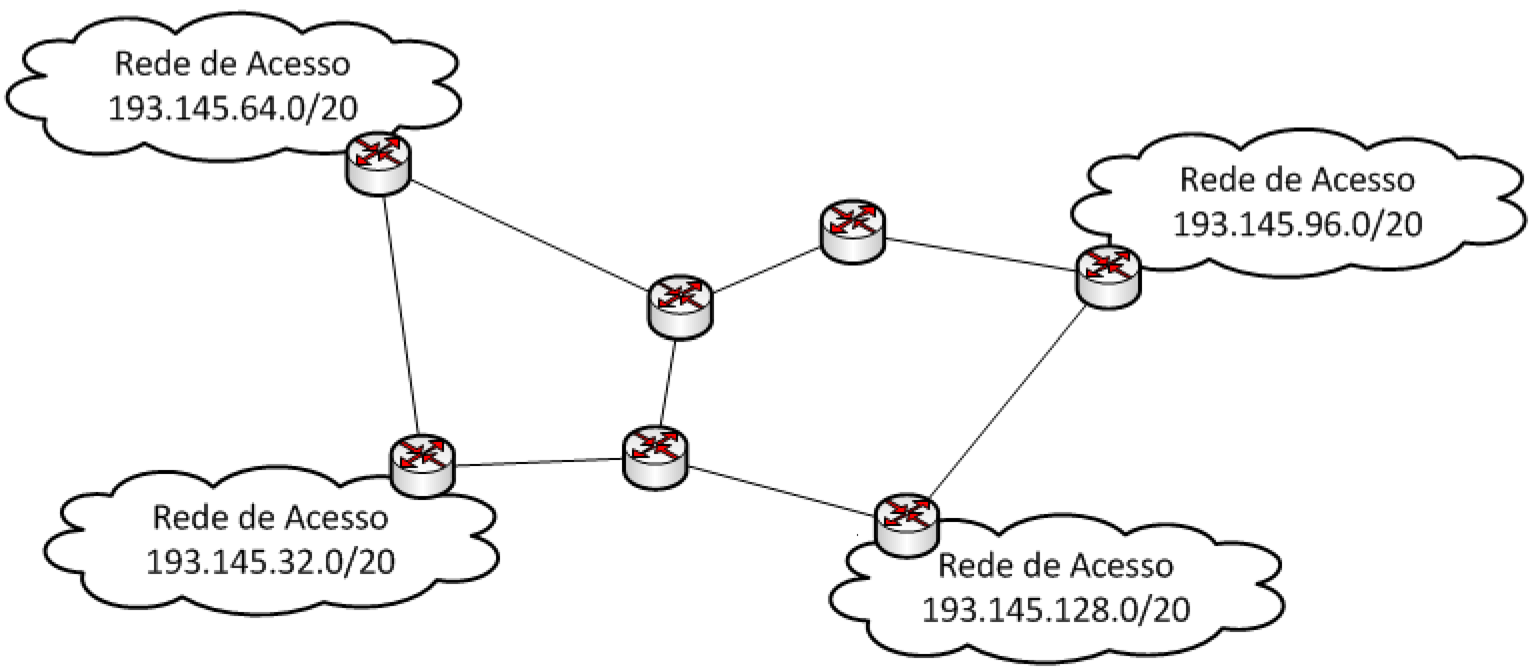
DETI-UA, 2023/2024

Encaminhamento em redes com comutação de pacotes

Existem 2 tipos de redes com comutação de pacotes:

- redes de *circuitos virtuais*
- redes de datagramas

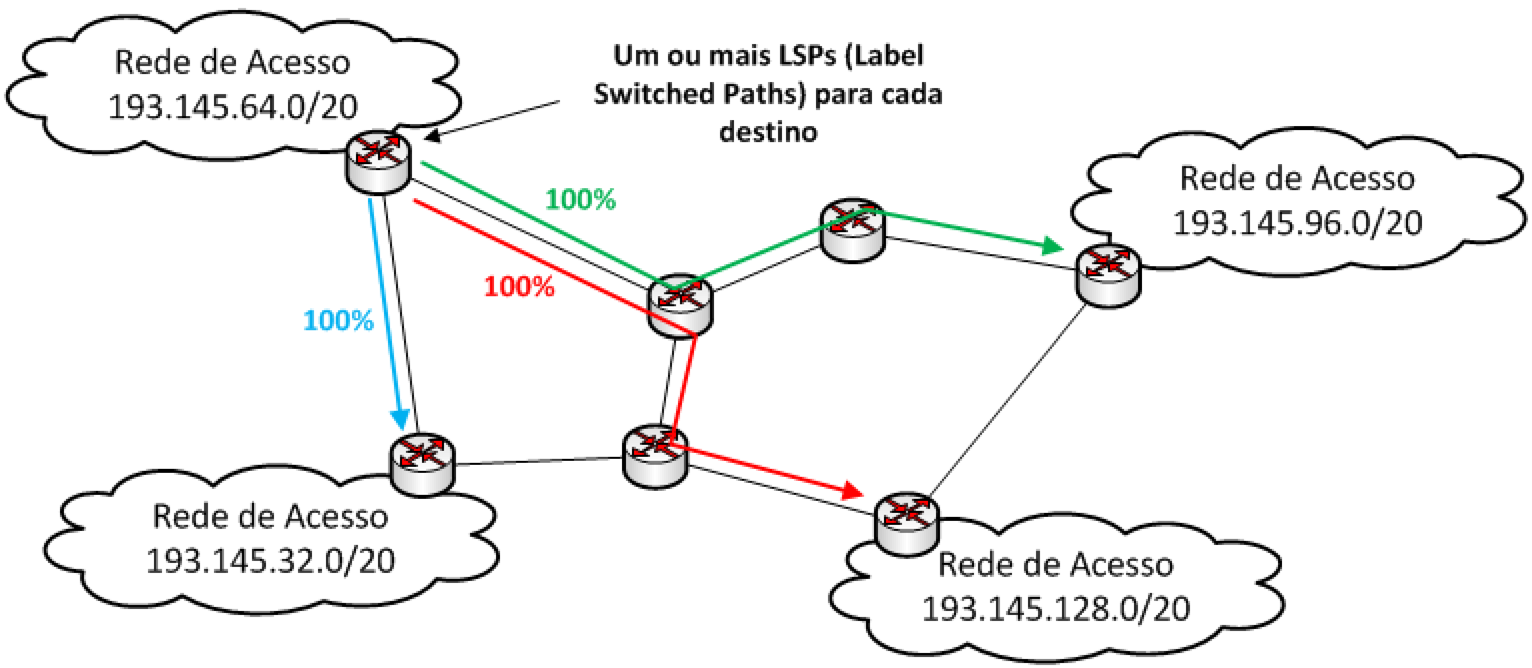
Considere-se o seguinte exemplo de uma rede de um ISP (*Internet Service Provider*) que liga 4 redes de acesso:



Encaminhamento em redes com comutação de pacotes – <u>redes de circuitos virtuais</u>

- A cada fluxo de pacotes, é atribuído pelo menos um circuito virtual.
- Os percursos dos circuitos virtuais são inicialmente estabelecidos.
- Após o estabelecimento dos circuitos virtuais, os pacotes de cada fluxo são encaminhados pelos circuitos virtuais atribuídos.

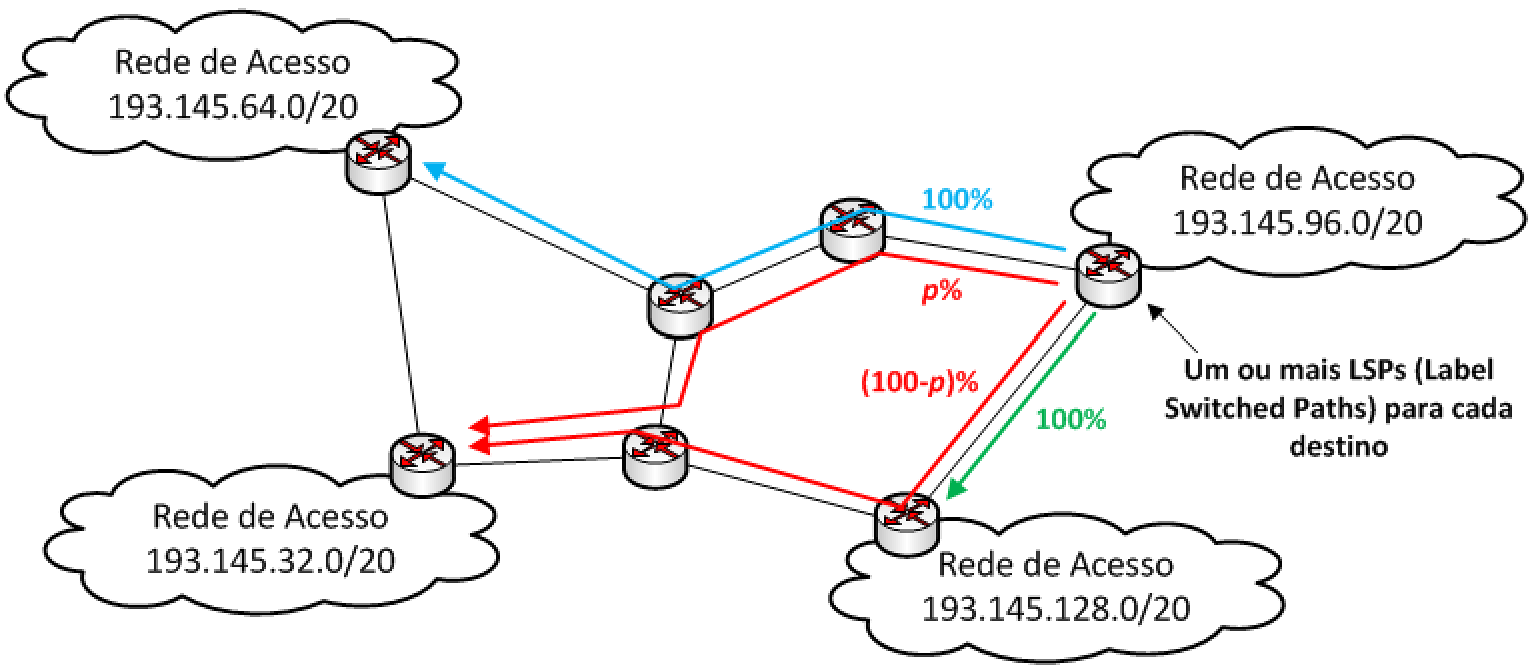
Exemplo: redes IP/MPLS em que os circuitos virtuais se designam por LSPs (*Label Switched Paths*).



Encaminhamento em redes com comutação de pacotes – <u>redes de circuitos virtuais</u>

- A cada fluxo de pacotes, é atribuído pelo menos um circuito virtual.
- Os percursos dos circuitos virtuais são inicialmente estabelecidos.
- Após o estabelecimento dos circuitos virtuais, os pacotes de cada fluxo são encaminhados pelos circuitos virtuais atribuídos.

Exemplo: redes IP/MPLS em que os circuitos virtuais se designam por LSPs (*Label Switched Paths*).



Encaminhamento em redes com comutação de pacotes – <u>redes de datagramas</u>

- As decisões de encaminhamento são efetuadas pacote a pacote.
- Assim, dois pacotes do mesmo par origem-destino podem seguir percursos distintos na rede.

Exemplo: redes IP com o protocolo de encaminhamento RIP ou OSPF.

Nas redes IP, o encaminhamento é baseado em *<u>percursos de custo mínimo</u>* de cada nó (router) para cada rede destino

- No OSPF, é atribuído a cada ligação um número positivo designado por *<u>custo</u>* da ligação.
- No RIP, o custo é 1 para cada ligação.
- Cada percurso de um router para um destino tem um custo igual à soma dos custos das ligações que o compõem.
- Em cada router, cada pacote IP é encaminhado por um dos percursos de custo mínimo para a rede destino do pacote.

Encaminhamento em redes com comutação de pacotes – redes de datagramas

Cada pacote IP é encaminhado por um dos percursos de custo mínimo para o destino do pacote:

Método estático: o custo das ligações é fixo (o caso do RIP e do OSPF).

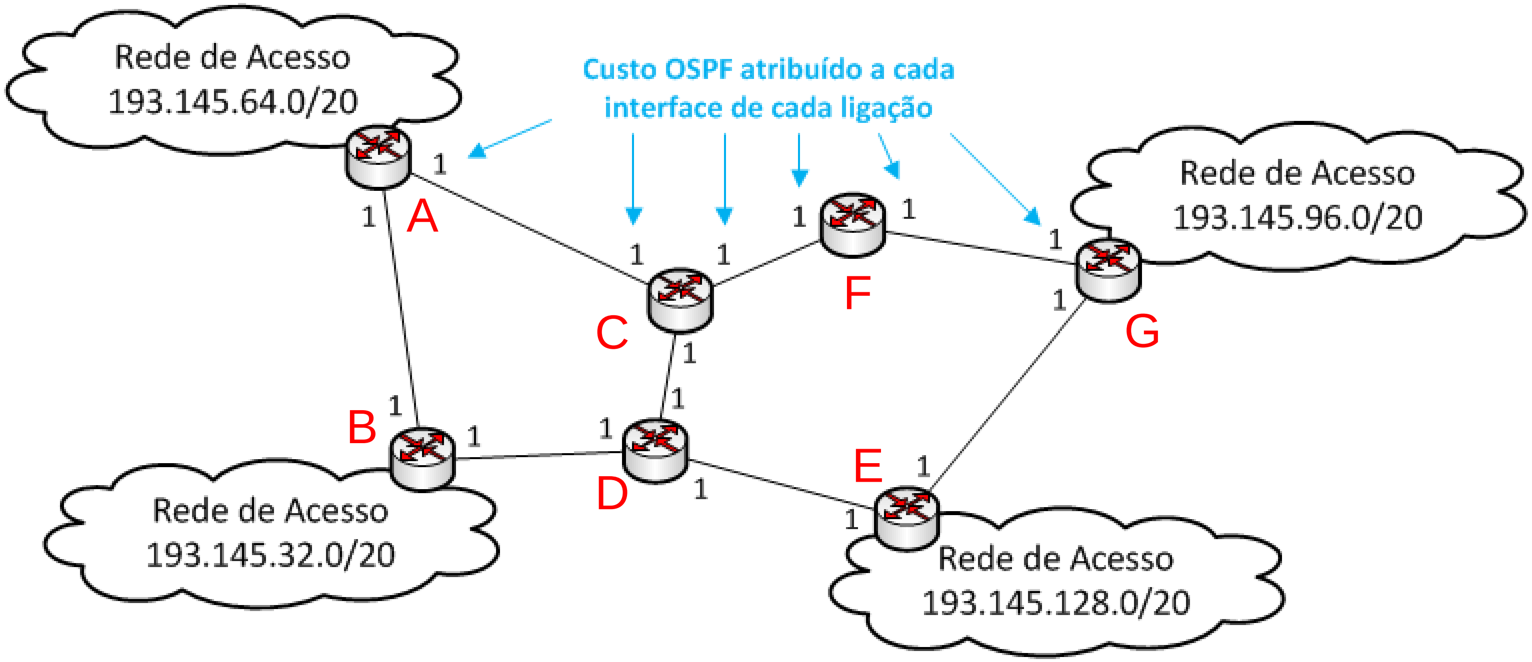
Método dinâmico: o custo das ligações varia ao longo do tempo em função do seu nível de utilização (exemplo: protocolos IGRP e EIGRP)

– o percurso de custo mínimo adapta-se a situações de sobrecarga obrigando os pacotes a evitarem as ligações mais utilizadas;
– introduz um efeito de realimentação que pode levar a oscilações indesejáveis.

Quando existem múltiplos percursos de custo mínimo de um nó para um destino, é usada a técnica ECMP (*Equal Cost Multi-Path*):

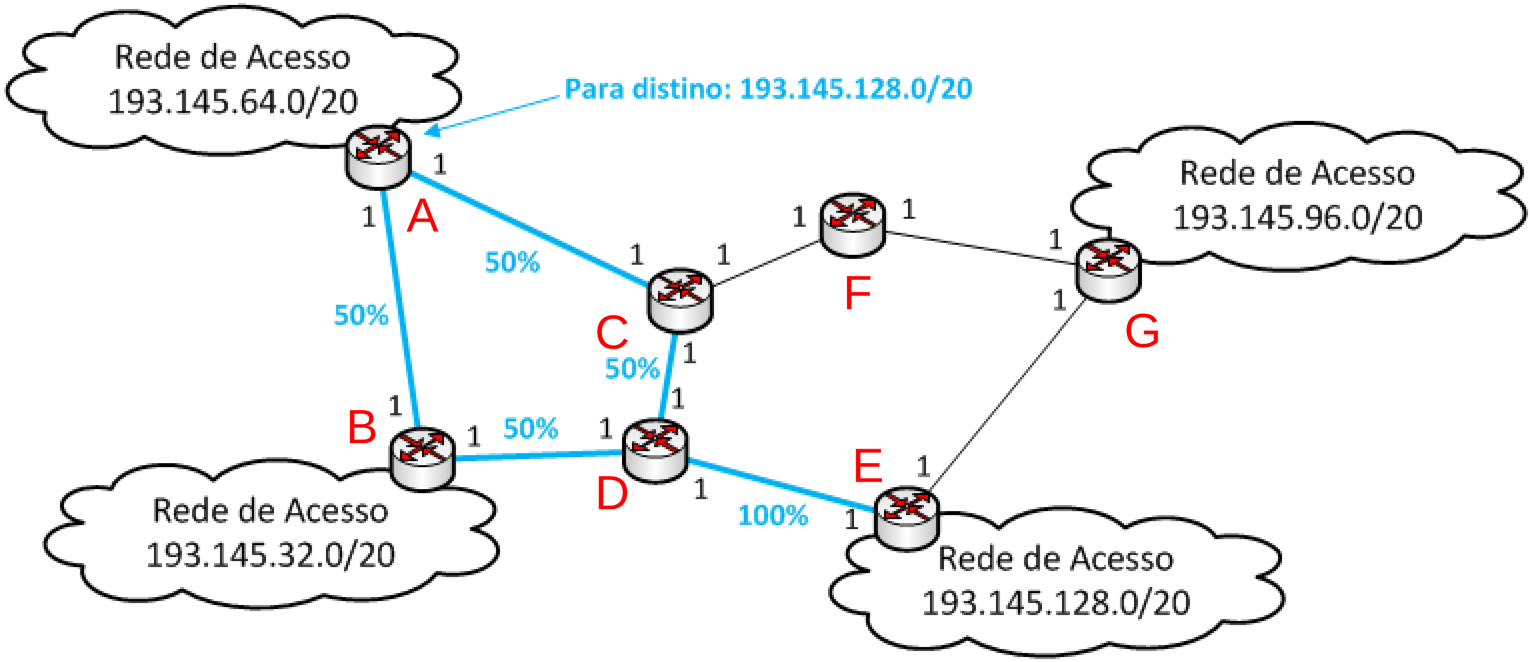
– em cada nó, o tráfego é bifurcado em igual percentagem por todas as ligações de saída que proporcionam percursos de custo mínimo

Encaminhamento em redes IP com encaminhamento OSPF (I)



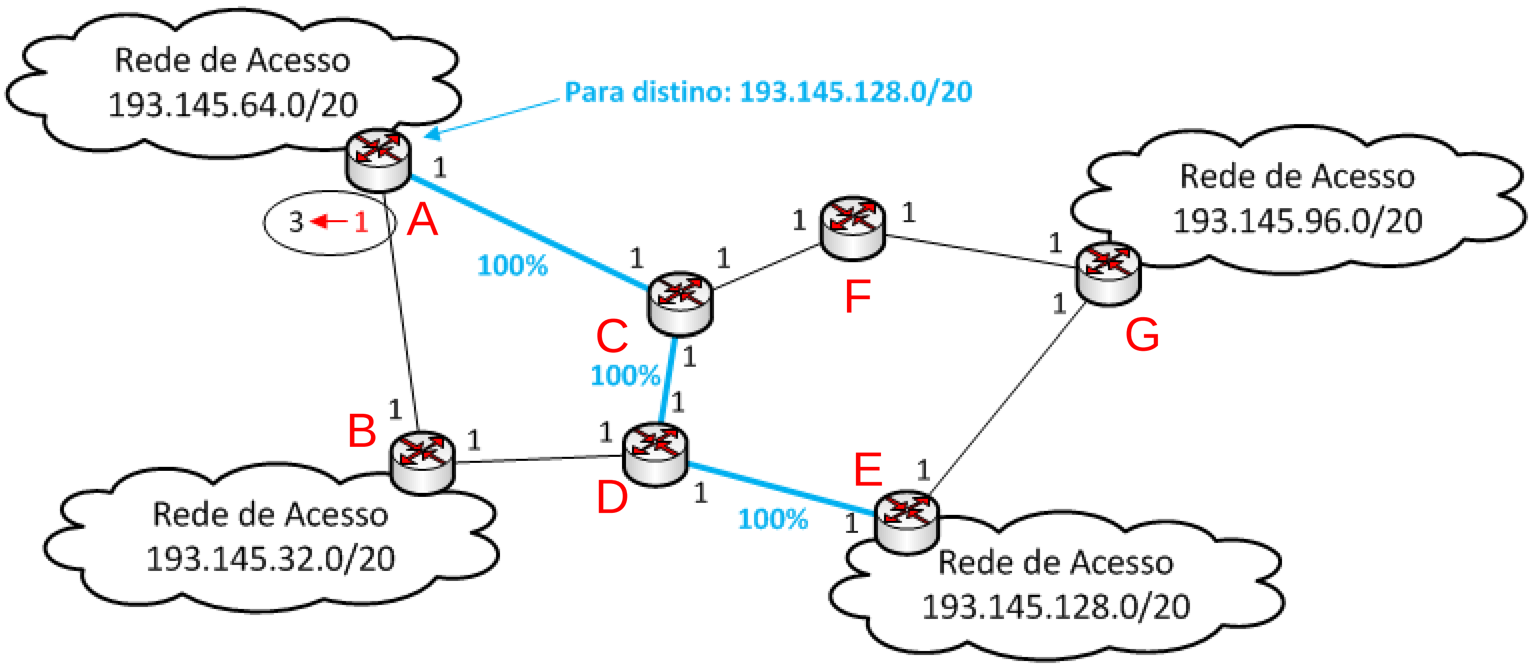
Neste exemplo, todos os custos OSPF estão configurados a 1 (equivalente ao RIP).

Encaminhamento em redes IP com encaminhamento OSPF (II)



Pelo ECMP, o router A encaminha os pacotes IP com destino para um endereço IP da rede 193.145.128.0/20 em igual percentagem pelos percursos que passam por B e por C.

Encaminhamento em redes IP com encaminhamento OSPF (III)

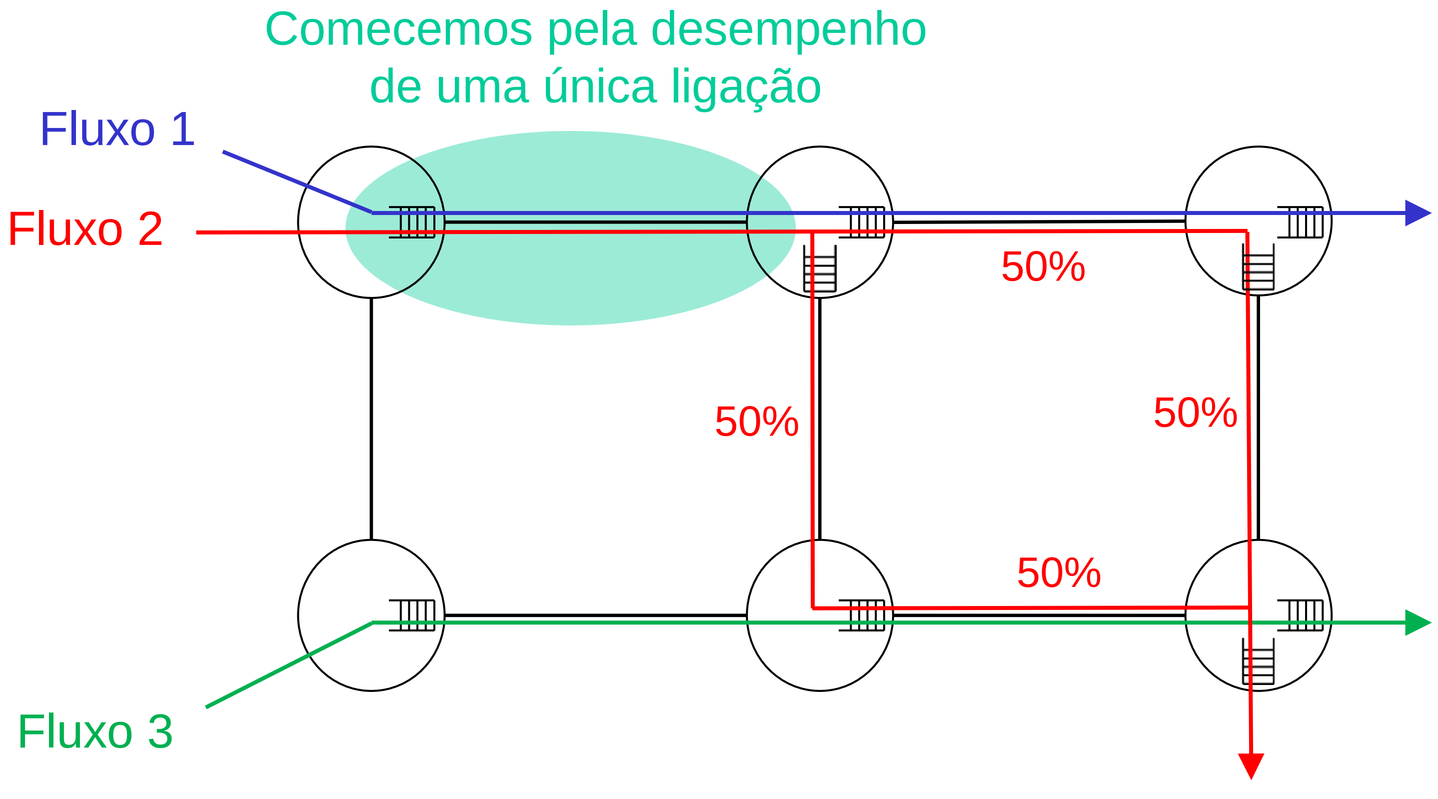


Mudando o custo da ligação de A para B de 1 para 3, o router A encaminha os pacotes IP com destino para um endereço IP da rede 193.145.128.0/20 pelo único percurso de custo mínimo.

Encaminhamento em redes com comutação de pacotes

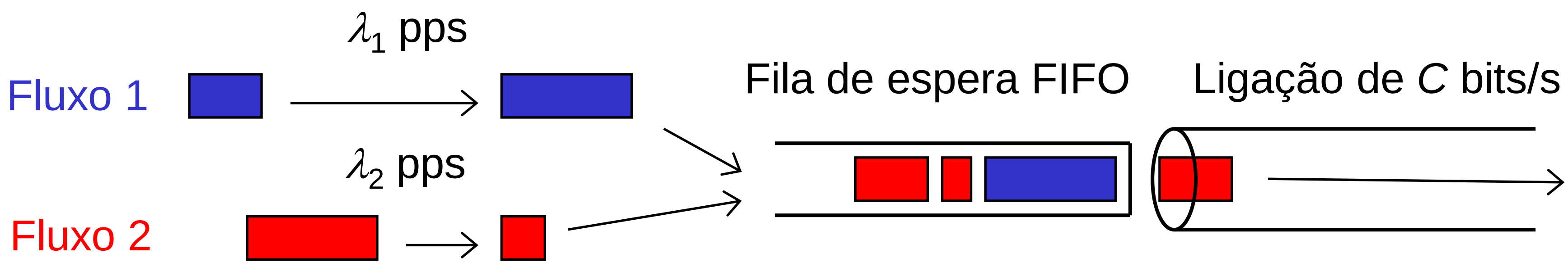
Uma rede é modelada por um conjunto de nós (representando os routers) e um conjunto de ligações entre nós.

O encaminhamento define a sequência de ligações por onde os pacotes de cada fluxo passam do nó origem até ao nó destino.



Multiplexagem estatística de fluxos de pacotes

Quando todos os fluxos são atendidos por uma única fila de espera FIFO, diz-se que os fluxos são multiplexados estatisticamente pela ligação.



Considerando-se que:

(i) as chegadas de pacotes de todos os fluxos são processos de Poisson,

(ii) a fila de espera é de tamanho infinito,

então a ligação é modelada por um **sistema *M*/*G*/1**.

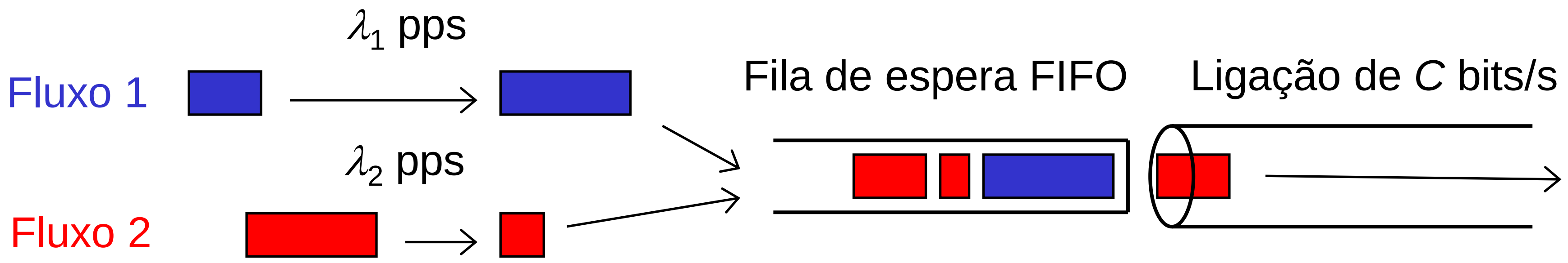
Os pacotes de todos os fluxos sofrem o mesmo atraso médio de fila de espera:

$$W_Q = \frac{\lambda E[S^2]}{2(1 - \lambda E[S])}$$

$\lambda = \lambda_1 + \lambda_2$ pps (pacotes por segundo)

$E[S]$ e $E[S^2]$ são a média do tempo de transmissão e do tempo de transmissão ao quadrado dos pacotes de todos os fluxos

Multiplexagem estatística de fluxos de pacotes



Considerando-se que:

(i) as chegadas de pacotes de todos os fluxos são processos de Poisson,

(ii) a fila de espera é de tamanho infinito,

(iii) <u>o tamanho dos pacote é exponencialmente distribuído de media *B* bits em todos of fluxos</u>,

então a ligação é modelada por um **sistema *M*/*M*/1**.

O atraso médio por pacote do agregado dos fluxos é:

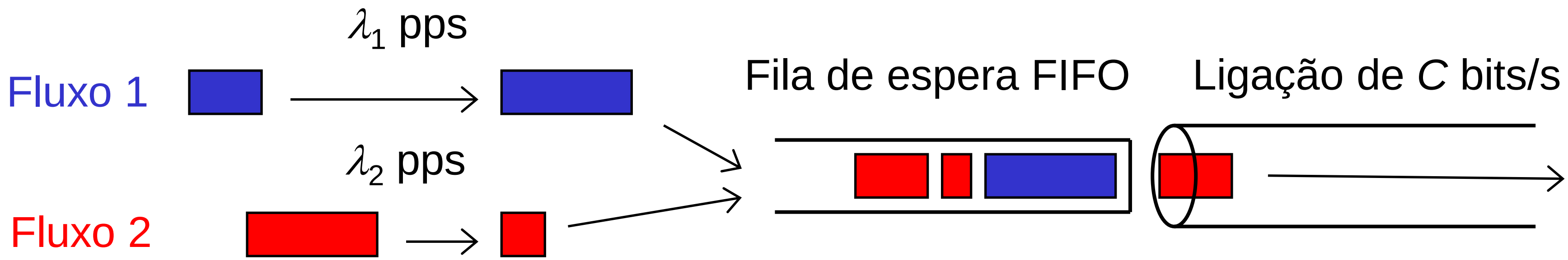
$$W = \frac{1}{\mu - \lambda}$$

$$\lambda = \lambda_1 + \lambda_2 \text{ pps}$$
$$\mu = C / B \text{ pps}$$

Os pacotes de todos os fluxos sofrem o mesmo atraso médio de fila de espera:

$$W_Q = W - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}$$

Multiplexagem estatística de fluxos de pacotes



Considerando-se que:

(i) as chegadas de pacotes de todos os fluxos são processos de Poisson,

(ii) <u>a fila de espera tem capacidade para *m* – 1 pacotes</u>,

(iii) o tamanho dos pacote é exponencialmente distribuído de media *B* bits em todos of fluxos,

então a ligação é modelada por um **sistema *M*/*M*/1/*m***.

– Taxa de perda de pacotes do agregado: $$\theta_m = \frac{(\lambda/\mu)^m}{\sum_{j=0}^{m}(\lambda/\mu)^j}$$

propriedade PASTA

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

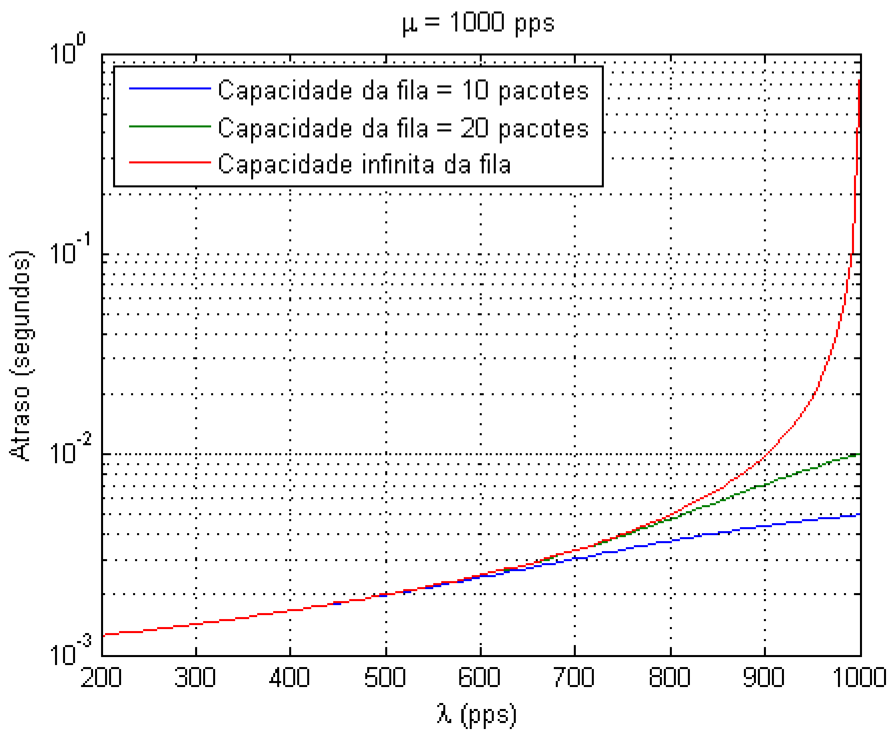
– Atraso médio dos pacotes do agregado: $$W = \frac{L}{\lambda(1-\theta_m)}$$

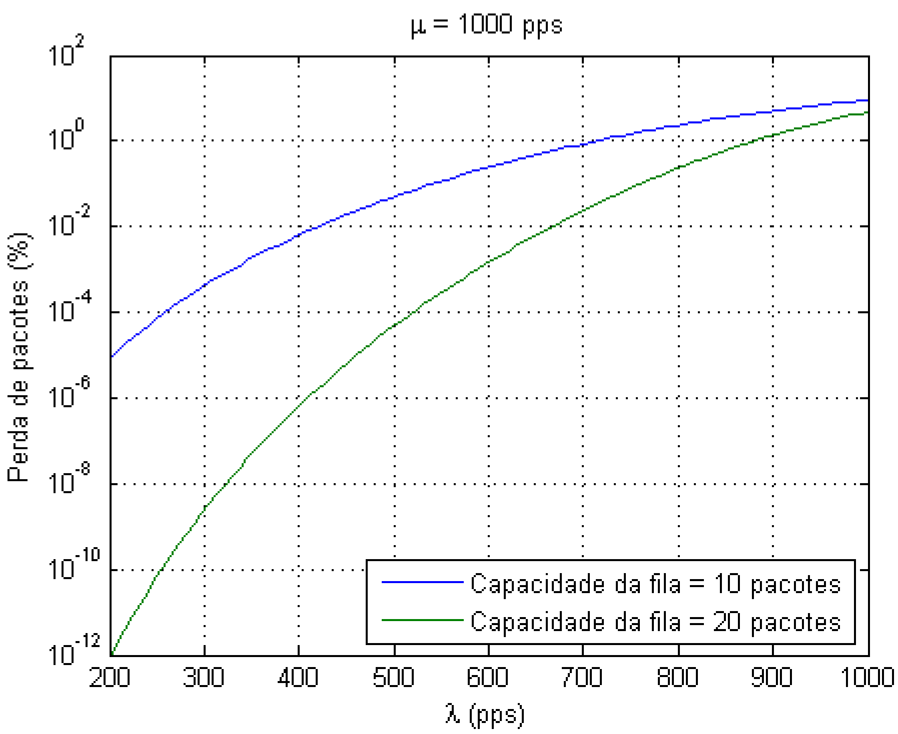
Teorema de Little

– Os pacotes de todos os fluxos sofrem o mesmo atraso médio de fila de espera: $$W_Q = W - \frac{1}{\mu}$$

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
  - $\mu = 10^7/(1250\times 8) = 1000$ pps





Disciplina com prioridades

- Na multiplexagem estatística, os pacotes de cada fluxo sofrem o mesmo atraso médio em fila de espera.
- Uma possibilidade para diferenciar o tratamento dos pacotes de diferentes fluxos é atribuir prioridades aos fluxos, i.e., os pacotes de um fluxo com determinada prioridade serem transmitidos antes dos pacotes dos fluxos com menor prioridade.

O sistema *M*/*G*/1 com prioridades pode ser utilizado para modelar este sistema assumindo que as chegadas de pacotes de todos os fluxos são processos de Poisson.

Considere um sistema *M*/*G*/1 com *n* prioridades em que 1 corresponde à prioridade mais alta e *n* corresponde à prioridade mais baixa.

Considere o agregado de fluxos de pacotes da prioridade k, $1 \leq k \leq n$, definido por:

- taxa de chegadas de pacotes: λ_k
- média (ou 1º momento) e 2º momento do tempo de transmissão dos pacotes: $E[S_k]$ e $E\left[S_k^2\right]$

Sistema *M/G/1* com prioridades

O sistema transmite primeiro os pacotes de maior prioridade.

Os pacotes dos fluxos com a mesma prioridade são transmitidos por ordem de chegada (disciplina *FIFO - First In First Out*).

Considera-se que as chegadas dos pacotes de cada prioridade são independentes (entre prioridades) e de Poisson, e independentes dos tempos de transmissão.

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

O *atraso médio por pacote na fila de espera* correspondente aos pacotes da prioridade *k* é dado por:

$$W_{Qk} = \begin{cases} \dfrac{\sum_{i=1}^{n}\left(\lambda_i E\left[S_i^2\right]\right)}{2(1-\rho_1)} & ,k=1 \\ \dfrac{\sum_{i=1}^{n}\left(\lambda_i E\left[S_i^2\right]\right)}{2(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)} & ,k>1 \end{cases} \qquad \rho_k = \lambda_k E[S_k]$$

Condição de validade: $\rho_1 + \cdots + \rho_n < 1$

Exemplo 1

Considere uma ligação ponto-a-ponto de 10 Mbps com uma fila de espera muito grande e que suporta dois fluxos de pacotes: fluxo A de 1 Mbps e fluxo B de 6 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 ligação;

(b) o fluxo A tem maior prioridade na fila de espera que o fluxo B.

Exemplo 1 – resolução (a)

Considere uma ligação ponto-a-ponto de 10 Mbps com uma fila de espera muito grande e que suporta dois fluxos de pacotes: fluxo A de 1 Mbps e fluxo B de 6 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 ligação;

$$\mu = \frac{10 \times 10^6}{8 \times 1000} = 1250 \text{ pps} \qquad \lambda_A = \frac{1 \times 10^6}{8 \times 1000} = 125 \text{ pps} \qquad \lambda_B = \frac{6 \times 10^6}{8 \times 1000} = 750 \text{ pps}$$

$$W_A = W_B = \frac{1}{\mu - (\lambda_A + \lambda_B)} = \frac{1}{1250 - (125 + 750)} = 2.67 \times 10^{-3} = 2.67 \text{ ms}$$

Exemplo 1 – resolução (b)

Considere uma ligação ponto-a-ponto de 10 Mbps com uma fila de espera muito grande e que suporta dois fluxos de pacotes: fluxo A de 1 Mbps e fluxo B de 6 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:

(b) o fluxo A tem maior prioridade na fila de espera que o fluxo B.

$$\lambda_A = \frac{1\times10^6}{8\times1000} = 125 \text{ pps}$$

$$\lambda_B = \frac{6\times10^6}{8\times1000} = 750 \text{ pps}$$

$$\mu_A = \mu_B = \mu = \frac{10\times10^6}{8\times1000} = 1250 \text{ pps}$$

$$E[S_A] = E[S_B] = \frac{1}{\mu} = \frac{1}{1250} \text{ seg.}$$

$$E[S_A^2] = E[S_B^2] = \frac{2}{\mu^2} = \frac{2}{1250^2} \text{ seg.}^2$$

Relembrar que: $E[S^2] = Var[S] + (E[S])^2$

$$W_{Qk} = \begin{cases} \dfrac{\sum_{i=1}^{n}\left(\lambda_i E[S_i^2]\right)}{2(1-\rho_1)} & , k = 1 \\[2ex] \dfrac{\sum_{i=1}^{n}\left(\lambda_i E[S_i^2]\right)}{2(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)} & , k > 1 \end{cases} \qquad \rho_k = \lambda_k E[S_k]$$

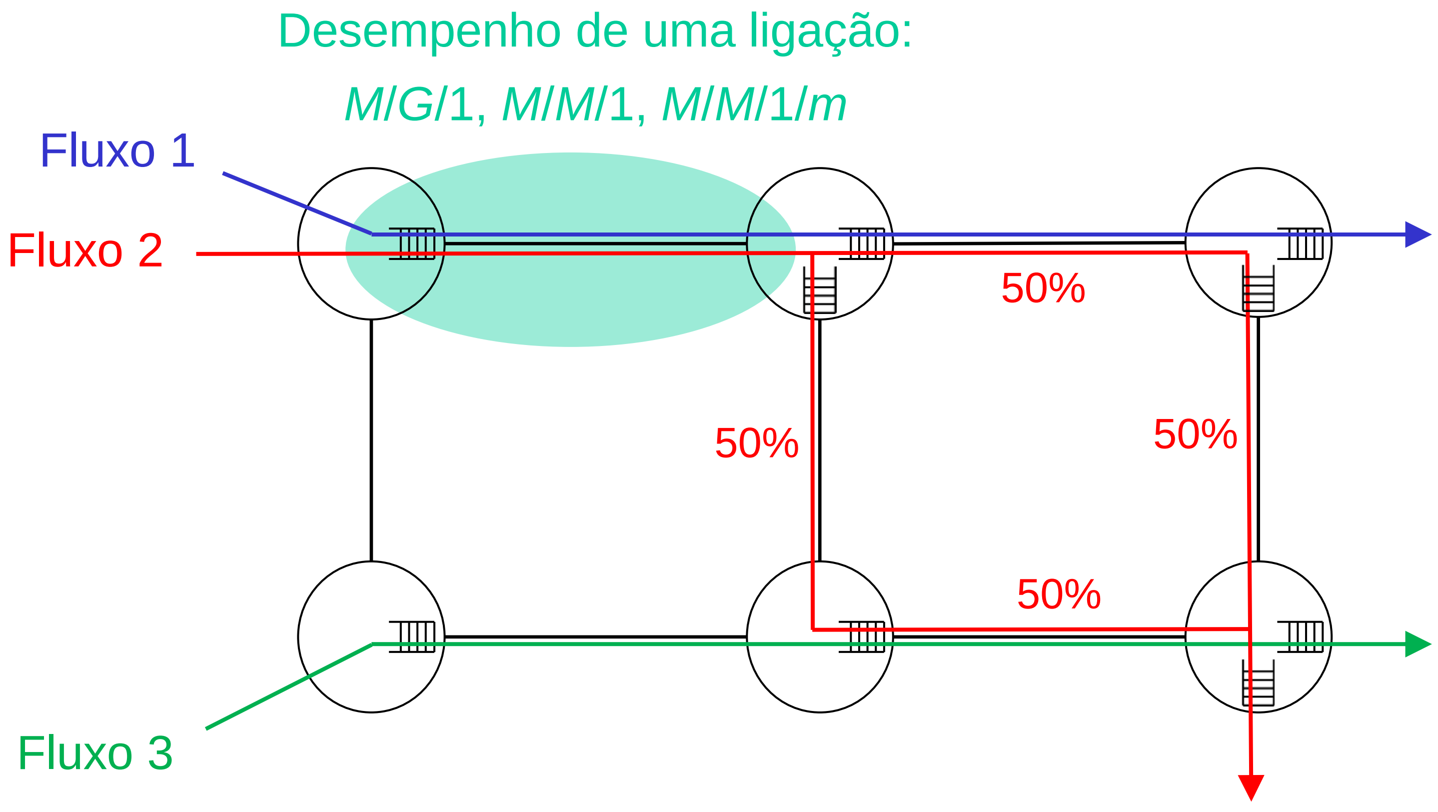
$$W_A = \frac{\lambda_A \times E[S_A^2] + \lambda_B \times E[S_B^2]}{2\times(1-\rho_A)} + E[S_A] = 1.42 \text{ ms}$$

$$W_B = \frac{\lambda_A \times E[S_A^2] + \lambda_B \times E[S_B^2]}{2\times(1-\rho_A)\times(1-\rho_A-\rho_B)} + E[S_B] = 2.87 \text{ ms}$$

Encaminhamento em redes com comutação de pacotes

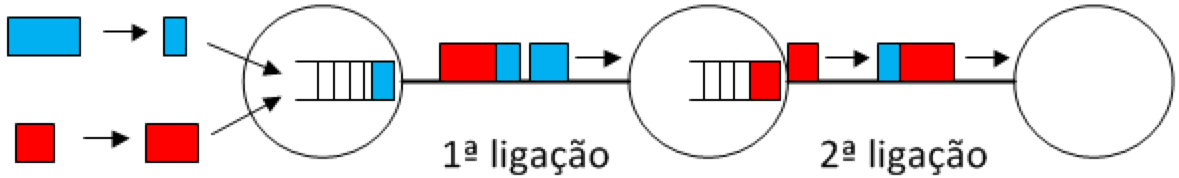
Uma rede é modelada por um conjunto de nós (representando os routers) e um conjunto de ligações entre nós.

O encaminhamento define a sequência de ligações por onde os pacotes de cada fluxo passam do nó origem até ao nó destino.



Redes de ligações ponto-a-ponto

Numa rede de ligações ponto-a-ponto os intervalos entre chegadas de pacotes estão correlacionados com o comprimento dos pacotes, após a passagem pela primeira ligação. Este facto dificulta a análise.



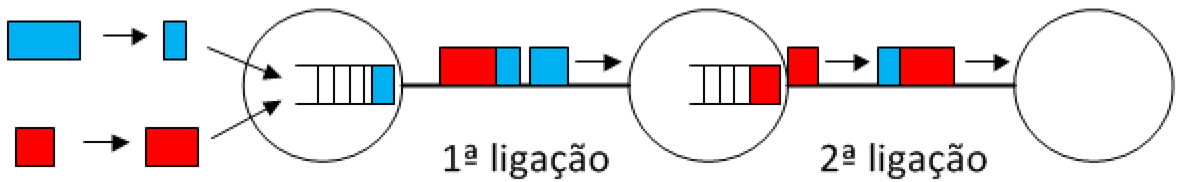
Exemplo:

- Considerem-se duas ligações ponto-a-ponto em cascata.
- Considere-se um conjunto de fluxos de pacotes com origem no nó à esquerda e destino no nó à direita.
- Considere-se que os pacotes destes fluxos chegam segundo um processo de Poisson e o comprimento dos pacotes é exponencialmente distribuído em ambos os fluxos com o mesmo tamanho médio.

Redes de ligações ponto-a-ponto

Numa rede de ligações ponto-a-ponto os intervalos entre chegadas de pacotes estão correlacionados com o comprimento dos pacotes, após a passagem pela primeira ligação. Este facto dificulta a análise.

no 2º link ja n é processo poisson



- a 1ª fila de espera é do tipo *M*/*M*/1
- no entanto, a 2ª fila de espera não é do tipo *M*/*M*/1:
  - o intervalo entre a chegada de dois pacotes consecutivos à 2ª fila de espera é sempre superior ou igual ao tempo de transmissão do segundo pacote na 1ª ligação (ou seja, não é uma distribuição exponencial);
  - assim, tipicamente pacotes maiores demoram mais tempo a ser transmitidos na 1ª ligação e esperam menos tempo na 2ª fila de espera que pacotes mais pequenos.

Aproximação de Kleinrock

A *aproximação de Kleinrock* consiste em assumir que as chegadas de pacotes são processos de Poisson em todos as ligações

- i.e., ignora a correlação entre comprimento dos pacotes e intervalos entre chegadas de pacotes

Nas ligações com filas de espera muito grandes:

- quando o tamanho dos pacotes é exponencialmente distribuído com a mesma média em todos os fluxos – ligação modelada por um *M*/*M*/1
- caso contrário – ligação modelada por um *M*/*G*/1

G modela o tempo de transmissao

Nas ligações em que as filas de espera não são muito grandes:

- assumindo o tamanho dos pacotes exponencialmente distribuído com a mesma média em todos os fluxos – ligação modelada por um *M*/*M*/1/*m*

De notar que:

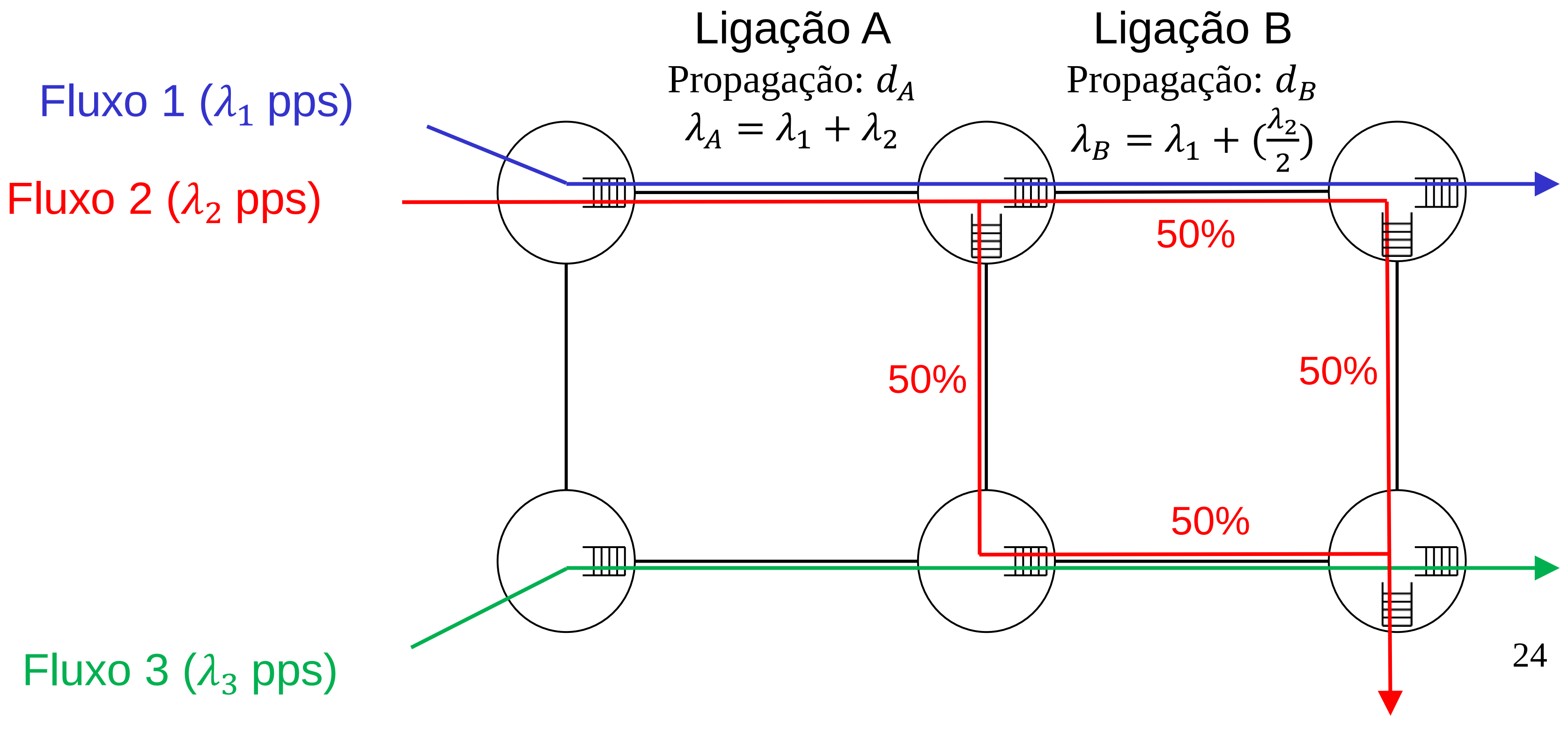
- os fluxos de pacotes são unidirecionais e as ligações das redes de comutação de pacotes são bidirecionais
- assim, uma ligação de rede entre os nós *i* e *j* é representada pelos pares ordenados (*i*,*j*) e (*j*,*i*) que indicam cada sentido da ligação

Atraso médio por pacote de cada fluxo

Num fluxo com um único percurso de encaminhamento (por exemplo, o caso do Fluxo 1), o atraso médio por pacote do fluxo é a soma dos atrasos médios em cada ligação do percurso.

$$\boxed{W_1 = W_{A1} + d_A + W_{B1} + d_B}$$

W_{A1} – atraso médio em fila de espera mais o tempo médio de transmissão dos pacotes do fluxo 1 na ligação A



Atraso médio por pacote de cada fluxo

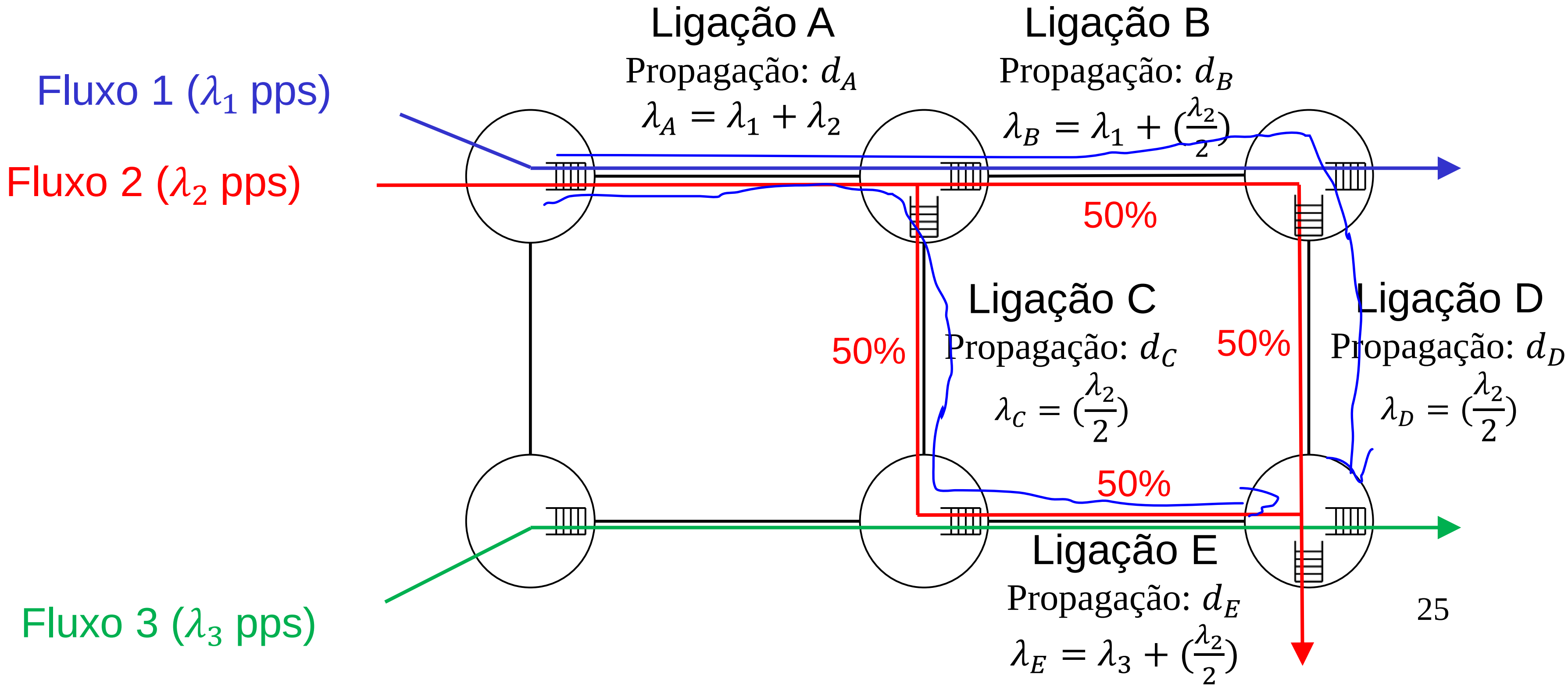
Num fluxo com diferentes percursos de encaminhamento (no exemplo, o caso do Fluxo 2), o atraso médio por pacote do fluxo é:

- a média pesada do atraso médio por pacote de cada percurso
- o peso é a percentagem de pacotes encaminhados por cada percurso.

$$\boxed{\begin{aligned} W_2 = 0.5 \times (W_{A2} + d_A + W_{B2} + d_B + W_{D2} + d_D) + \\ + 0.5 \times (W_{A2} + d_A + W_{C2} + d_C + W_{E2} + d_D) \end{aligned}}$$

Caminho por cima

Caminho por baixo



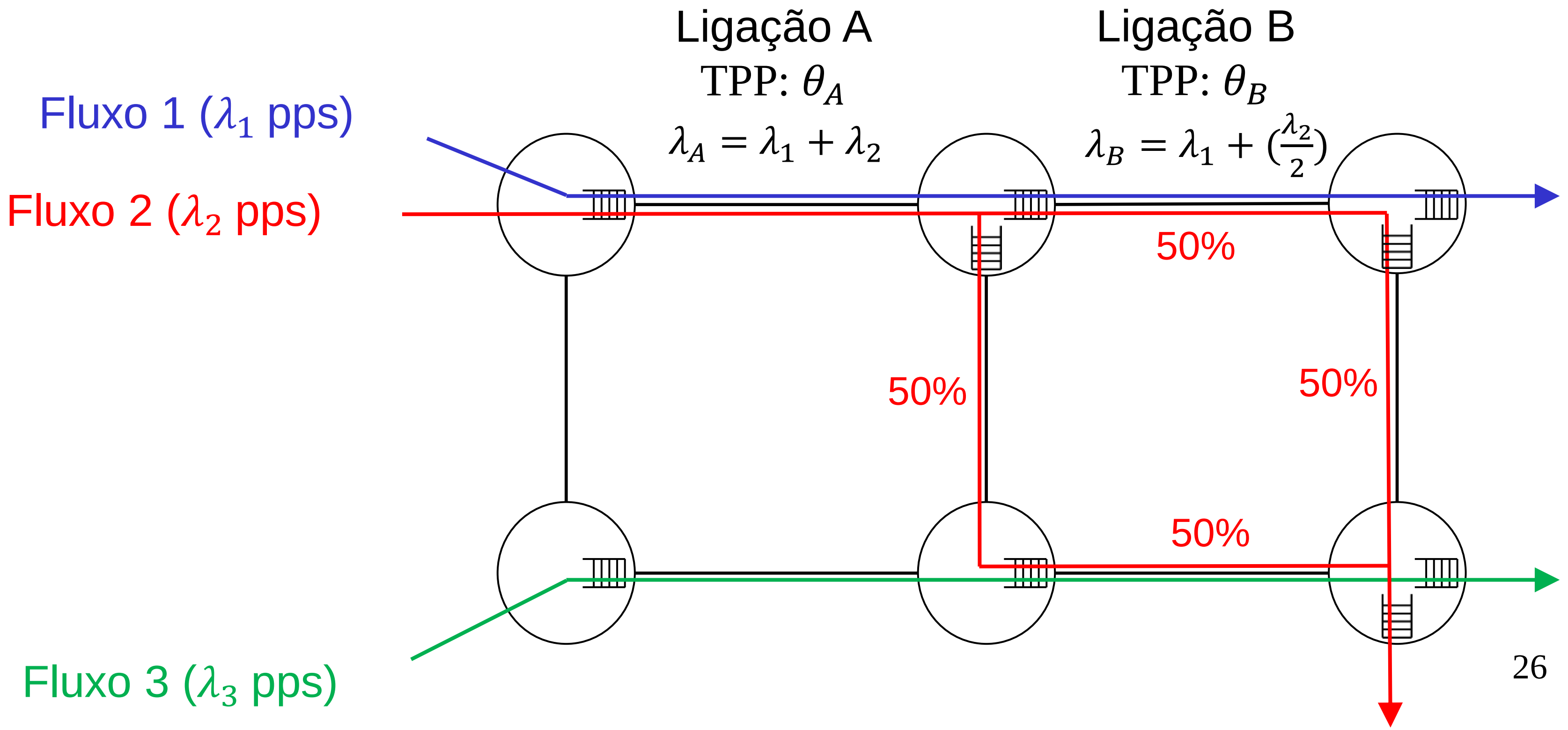
Taxa de perda de pacotes (TPP) de cada fluxo

Num fluxo com um único percurso de encaminhamento (por exemplo, o caso do Fluxo 1), a taxa de perda de pacotes do fluxo é a probabilidade de cada pacote ser descartado na 1ª ligação, ou na 2ª ligação, etc.

Prob de perder na 1º posiçao + prob de nao perder na primeira e perder na 2º posição

$$\boxed{\theta_1 = \theta_A + (1 - \theta_A) \times \theta_B}$$

TPP – taxa de perda de pacotes



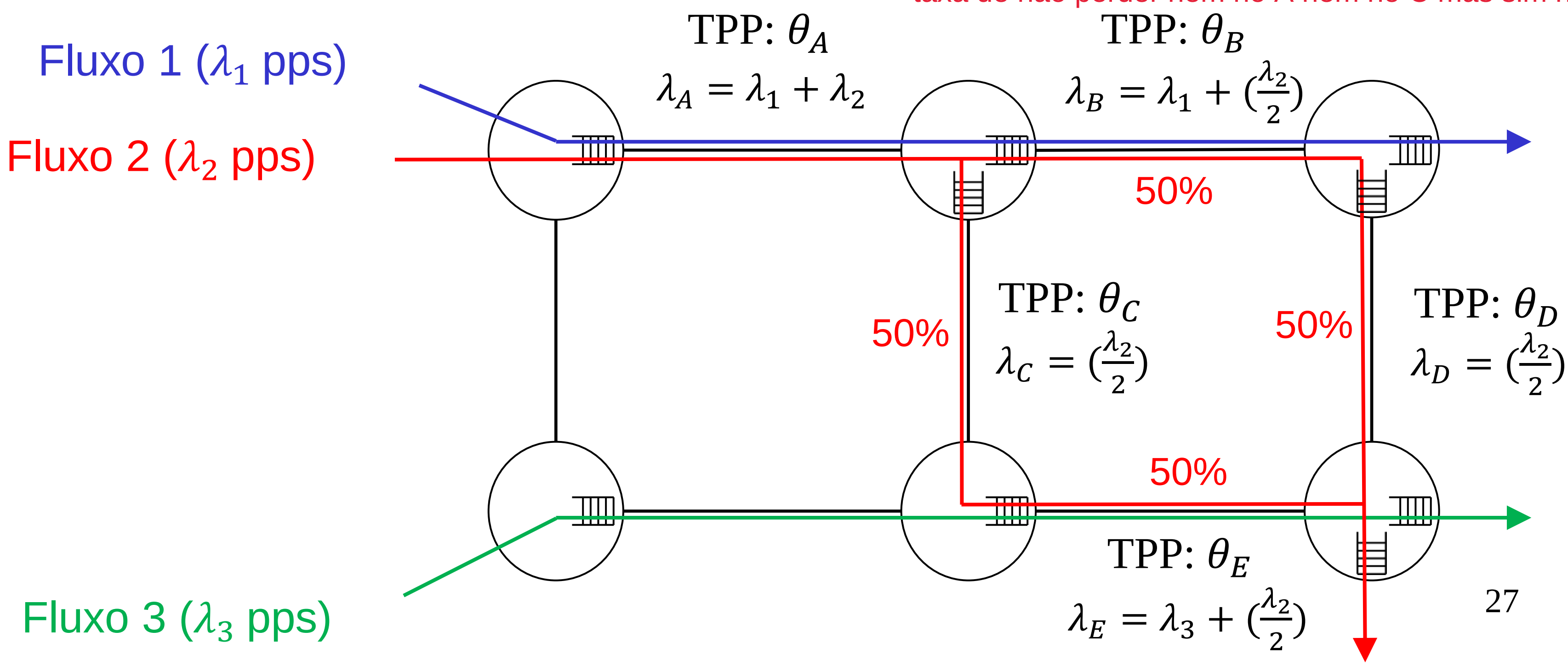
Taxa de perda de pacotes (TPP) de cada fluxo

Num fluxo com diferentes percursos de encaminhamento (no exemplo, o caso do Fluxo 2), a taxa de perda de pacotes do fluxo é:

- a média pesada da taxa de perda de pacotes de cada percurso
- o peso é a percentagem de pacotes encaminhados por cada percurso.

$$\theta_2 = 0.5 \times [\theta_A + (1-\theta_A)\times\theta_B + (1-\theta_A)\times(1-\theta_B)\times\theta_D] + \\ + 0.5 \times [\theta_A + (1-\theta_A)\times\theta_C + (1-\theta_A)\times(1-\theta_C)\times\theta_E]$$

taxa de não perder nem no A nem no C mas sim no E



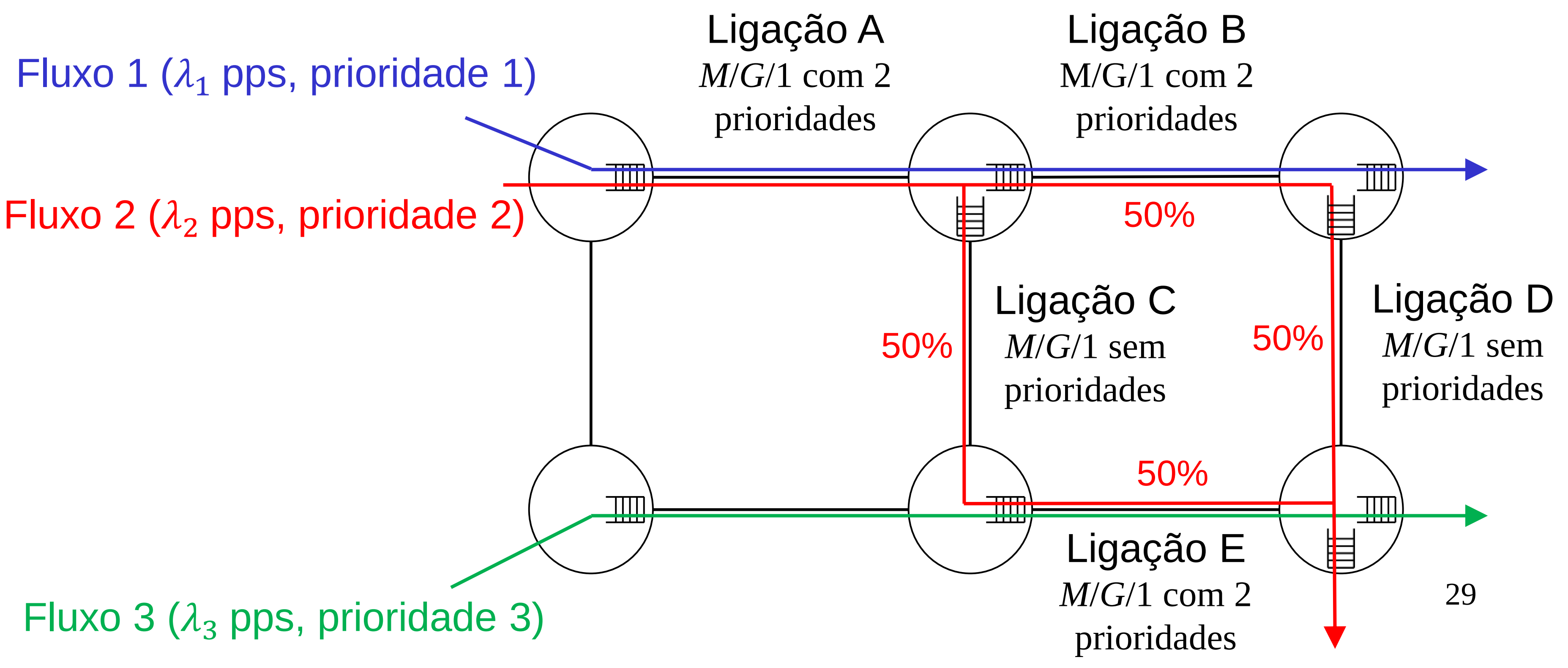
Impacto da TPP no atraso médio por pacote de cada fluxo

- A taxa de perda de pacotes (TPP) de um fluxo numa ligação reduz a taxa de entrada de pacotes do fluxo nessa ligação e nas ligações subsequentes do percurso de encaminhamento até ao nó destino.
- Na prática, as taxas de perda de pacotes são valores muito pequenos (tipicamente < 1%)
  - em serviços de dados, o protocolo TCP tem um mecanismo de controlo de fluxo que reduz a taxa de emissão de pacotes quando os pacotes não chegam ao destino;
  - em serviços *real-time*, os codecs conseguem reconstruir o áudio (e/ou o vídeo) para pequenas taxas de perda de pacotes; caso contrário, interrompem o serviço.
- Para valores de TPP pequenos, o impacto na redução da taxa de entrada de pacotes é insignificante pelo que pode ser ignorado na determinação do atraso médio por pacote de cada fluxo (conforme descrito anteriormente).

Prioridades globais e prioridades em cada ligação

Considerando que a cada fluxo da rede é atribuída uma prioridade global:

- em cada ligação aplica-se o modelo *M*/*G*/1 com as prioridades apenas com os fluxos suportados pela ligação
- no exemplo, o Fluxo 2 tem menor prioridade que o Fluxo 1 nas ligações A e B mas tem maior prioridade que o Fluxo 3 na ligação E



Rede de ligações *M*/*M*/1

Considere-se uma rede de ligações ponto-a-ponto em que a fila de espera de todas as ligações é muito grande.

Considere-se que a rede suporta diferentes fluxos de pacotes $s = 1 \ldots S$ com a mesma prioridade entre si.

Considere-se que o tamanho dos pacotes é exponencialmente distribuído com a mesma média em todos os fluxos.

Neste caso, todas as ligações são modeladas por um *M*/*M*/1.

Considere-se que cada fluxo $s = 1 \ldots S$ é suportado por um percurso único na rede, formado por uma sequência de ligações (i,j) definida pelo conjunto R_s. Da tds as ligaçoes

Seja λ_s a taxa de chegada de pacotes do fluxo s, em pacotes/segundo.

Então a taxa total de chegada de pacotes à ligação (i,j) é:

$$\lambda_{ij} = \sum_{s:(i,j)\in R_s} \lambda_s$$

Rede de ligações *M/M/*1

Considere-se agora o caso em que pode haver múltiplos percursos associados a cada fluxo de pacotes s:

- Seja $f_{ij}(s)$ a fração de pacotes do fluxo s que atravessa a ligação (i,j).
- Neste caso, o conjunto R_s inclui todas as ligações (i,j) tais que $f_{ij}(s)>0$. Diz as percentagens atribuidas a cada pacote

Então a taxa total de chegada de pacotes à ligação (i,j) é:

$$\lambda_{ij} = \sum_{s:(i,j)\in R_s} f_{ij}(s)\lambda_s$$

Considerando μ_{ij} a capacidade da ligação (i,j) em número médio de pacotes/segundo, o número médio de pacotes em todas as ligações é (relembrar o modelo *M/M/*1):

$$L = \sum_{(i,j)} \frac{\lambda_{ij}}{\mu_{ij} - \lambda_{ij}}$$

Rede de ligações *M*/*M*/1

Usando o teorema de Little, o atraso médio por pacote na rede é:

$$W = \frac{1}{\gamma}\sum_{(i,j)}\frac{\lambda_{ij}}{\mu_{ij}-\lambda_{ij}} \qquad \gamma = \sum_{s}\lambda_s$$

Nos casos em que o atraso de propagação nas ligações não é desprezável, o atraso médio por pacote na rede passa a ser

$$W = \frac{1}{\gamma}\sum_{(i,j)}\left(\frac{\lambda_{ij}}{\mu_{ij}-\lambda_{ij}}+\lambda_{ij}d_{ij}\right) \qquad \gamma = \sum_{s}\lambda_s$$

em que d_{ij} é o atraso de propagação da ligação (*i*, *j*).

Rede de ligações *M*/*M*/1

No caso em que a cada fluxo *s* está associado um percurso único na rede, o atraso médio por pacote do fluxo de tráfego *s* é:

$$W_s = \sum_{(i,j)\in R_s} \left(\frac{1}{\mu_{ij} - \lambda_{ij}} + d_{ij} \right)$$

No caso em que há diferentes percursos associados a cada fluxo de pacotes *s*, o atraso médio por pacote do fluxo *s* é:

- a média pesada do atraso de cada percurso (fórmula acima)
- o peso de cada percurso é a percentagem da taxa de chegada do fluxo *s*, λ_s, que é encaminhado pelo percurso.

- Nas redes com um percurso por fluxo, a maior fonte de erro associada à aproximação de Kleinrock deve-se à correlação entre os comprimentos dos pacotes e os intervalos entre chegadas.
- Nas redes com múltiplos percursos por fluxo, pode existir um fator adicional de erro, dependendo da forma como os fluxos são bifurcados nos nós.

Exemplo 2

Considere a rede IP da figura com todas as ligações bidirecionais de 10 Mbps. A rede suporta 4 fluxos de pacotes:

- de A para C com uma taxa de Poisson de 1000 pps,
- de A para D com uma taxa de Poisson de 250 pps,
- de B para D com uma taxa de Poisson de 1000 pps e
- de B para E com uma taxa de Poisson de 750 pps.

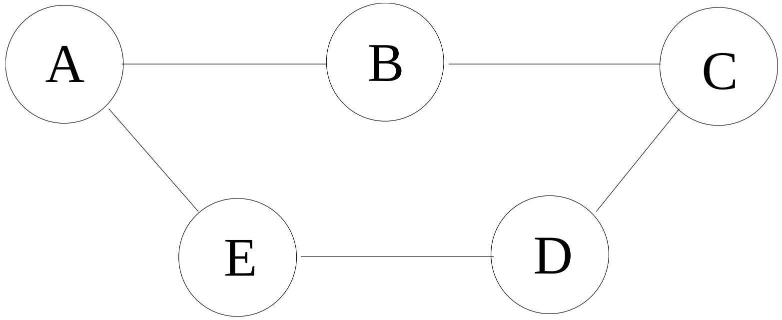
O tamanho dos pacotes é exponencialmente distribuído com média 500 bytes em todos os fluxos. O tempo de propagação da ligação B-C é de 10 ms em cada sentido e desprezável nas outras ligações.

O protocolo de encaminhamento nos routers é o RIP. Utilizando a aproximação de Kleinrock, calcule:

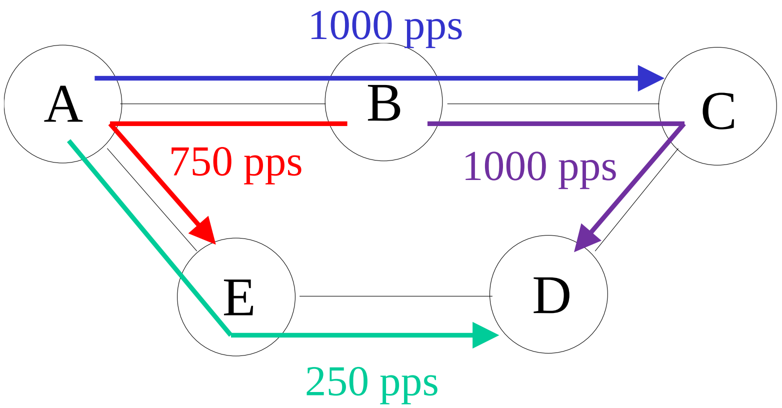
(a) o atraso médio por pacote de cada fluxo;

(b) o atraso médio por pacote de todos os fluxos;

(c) a utilização (em percentagem) de cada ligação em cada sentido.



Exemplo 2



- Ligações bidirecionais de 10 Mbps
- Pacotes de 500 bytes, em média
- Tempo de propagação da ligação B-C de 10 ms em cada sentido
- <u>Encaminhamento RIP</u>

$$W_s = \sum_{(i,j)\in R_S} \left(\frac{1}{\mu_{ij} - \lambda_{ij}} + d_{ij} \right)$$

(a) O atraso médio por pacote de cada fluxo.

$$\mu_{AB} = \mu_{BA} = \mu_{BC} = \ldots = \mu = \frac{10 \times 10^6 \text{ bps}}{500 \times 8 \text{ bpp}} = 2500 \text{ pps}$$

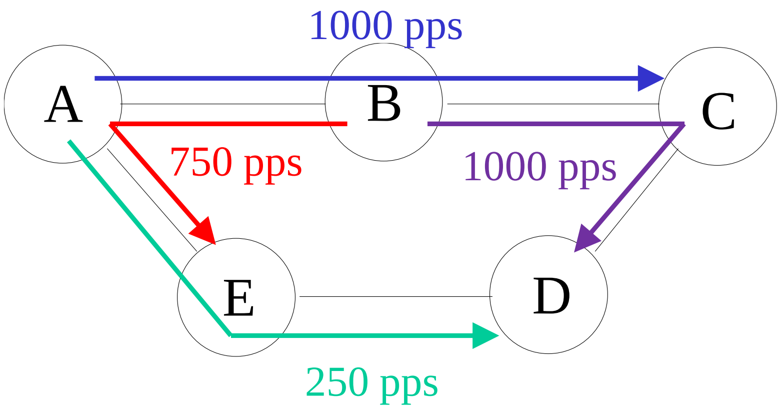
$$W_{A\to C} = \frac{1}{\mu_{AB} - \lambda_{AB}} + d_{AB} + \frac{1}{\mu_{BC} - \lambda_{BC}} + d_{BC} = \frac{1}{2500 - 1000} + 0 + \frac{1}{2500 - (1000 + 1000)} + 0.01 = 0.0127 \text{ seg.}$$

$$W_{A\to D} = \frac{1}{\mu_{AE} - \lambda_{AE}} + d_{AE} + \frac{1}{\mu_{ED} - \lambda_{ED}} + d_{ED} = \frac{1}{2500 - (750 + 250)} + 0 + \frac{1}{2500 - 250} + 0 = 0.0011 \text{ seg.}$$

$$W_{B\to D} = \frac{1}{\mu_{BC} - \lambda_{BC}} + d_{BC} + \frac{1}{\mu_{CD} - \lambda_{CD}} + d_{CD} = \frac{1}{2500 - (1000 + 1000)} + 0.01 + \frac{1}{2500 - 1000} + 0 = 0.0127 \text{ seg.}$$

$$W_{B\to E} = \frac{1}{\mu_{BA} - \lambda_{BA}} + d_{BA} + \frac{1}{\mu_{AE} - \lambda_{AE}} + d_{AE} = \frac{1}{2500 - 750} + 0 + \frac{1}{2500 - (750 + 250)} + 0 = 0.0012 \text{ seg.}$$

Exemplo 2



- Ligações bidirecionais de 10 Mbps
- Pacotes de 500 bytes, em média
- Tempo de propagação da ligação B-C de 10 ms em cada sentido
- <u>Encaminhamento RIP</u>

$$W = \frac{1}{\gamma}\sum_{(i,j)}\left(\frac{\lambda_{ij}}{\mu_{ij}-\lambda_{ij}} + \lambda_{ij} d_{ij}\right)$$

$$\gamma = \sum_s \lambda_s$$

(b) O atraso médio por pacote de todos os fluxos.

$$\mu_{AB} = \mu_{BA} = \mu_{BC} = \ldots = \mu = \frac{10\times10^6 \text{ bps}}{500\times 8 \text{ bpp}} = 2500 \text{ pps}$$

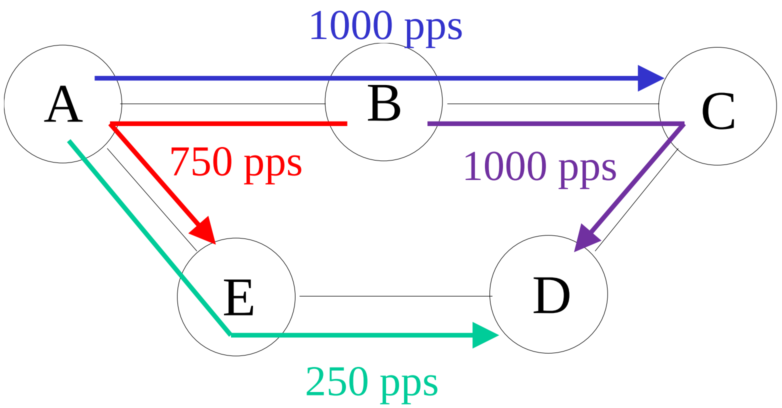
$$\gamma = \lambda_{A\to C} + \lambda_{A\to D} + \lambda_{B\to D} + \lambda_{B\to E} = 1000 + 250 + 1000 + 750 = 3000 \text{ pps}$$

$$W = \frac{1}{\gamma}\times\left(\frac{\lambda_{AB}}{\mu_{AB}-\lambda_{AB}} + \frac{\lambda_{BA}}{\mu_{BA}-\lambda_{BA}} + \frac{\lambda_{BC}}{\mu_{BC}-\lambda_{BC}} + \lambda_{BC} d_{BC} + \frac{\lambda_{CD}}{\mu_{CD}-\lambda_{CD}} + \frac{\lambda_{AE}}{\mu_{AE}-\lambda_{AE}} + \frac{\lambda_{ED}}{\mu_{ED}-\lambda_{ED}}\right)$$

$$W = \frac{1}{3000}\times\left(\frac{1000}{2500-1000} + \frac{750}{2500-750} + \frac{2000}{2500-2000} + 2000\times 0.01 + \frac{1000}{2500-1000} + \frac{1000}{2500-1000} + \frac{250}{2500-250}\right)$$

$$W = 0.00865 \text{ seg.}$$

Exemplo 2



- Ligações bidirecionais de 10 Mbps
- Pacotes de 500 bytes, em média
- Tempo de propagação da ligação B-C de 10 ms em cada sentido
- <u>Encaminhamento RIP</u>

(c) A utilização (em percentagem) de cada ligação em cada sentido.

$$\mu_{AB} = \mu_{BA} = \mu_{BC} = \ldots = \mu = \frac{10 \times 10^6 \text{ bps}}{500 \times 8 \text{ bpp}} = 2500 \text{ pps}$$

$$U_{AB} = \frac{\lambda_{AB}}{\mu_{AB}} = \frac{1000}{2500} = 0.4 = 40\%$$

$$U_{BA} = \frac{\lambda_{BA}}{\mu_{BA}} = \frac{750}{2500} = 0.3 = 30\%$$

$$U_{BC} = \frac{\lambda_{BC}}{\mu_{BC}} = \frac{2000}{2500} = 0.8 = 80\%$$

$$U_{CD} = \frac{\lambda_{CD}}{\mu_{CD}} = \frac{1000}{2500} = 0.4 = 40\%$$

$$U_{AE} = \frac{\lambda_{AE}}{\mu_{AE}} = \frac{1000}{2500} = 0.4 = 40\%$$

$$U_{ED} = \frac{\lambda_{ED}}{\mu_{ED}} = \frac{250}{2500} = 0.1 = 10\%$$

Exemplo 3

Considere a rede IP da figura com todas as ligações bidirecionais de 10 Mbps. A rede suporta 4 fluxos de pacotes:

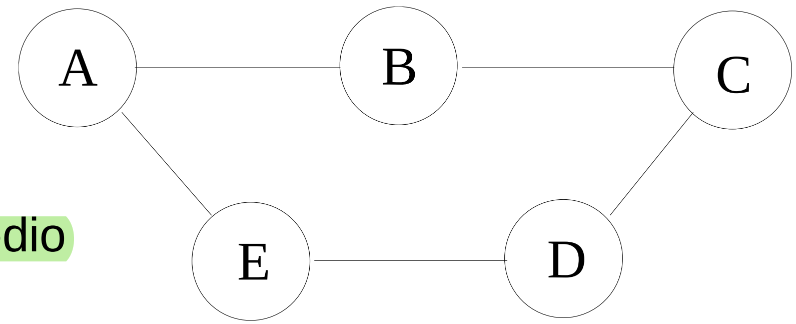
- de A para C com uma taxa de Poisson de 1000 pps,
- de A para D com uma taxa de Poisson de 250 pps,
- de B para D com uma taxa de Poisson de 1000 pps e
- de B para E com uma taxa de Poisson de 750 pps.

O tamanho dos pacotes é exponencialmente distribuído com média 500 bytes em todos os fluxos. O tempo de propagação da ligação B-C é de 10 ms em cada sentido e desprezável nas outras ligações.

O protocolo de encaminhamento nos routers é o OSPF.

(a) Determine os custos OSPF que permitem minimizar a utilização da ligação mais carregada.

(b) Utilizando a aproximação de Kleinrock, determine o atraso médio por pacote de todos os fluxos na solução anterior.



Exemplo 3

- Ligações bidirecionais de 10 Mbps
- Pacotes de 500 bytes, em média
- Propagação da ligação B-C de 10 ms em cada sentido
- <u>Encaminhamento OSPF</u>

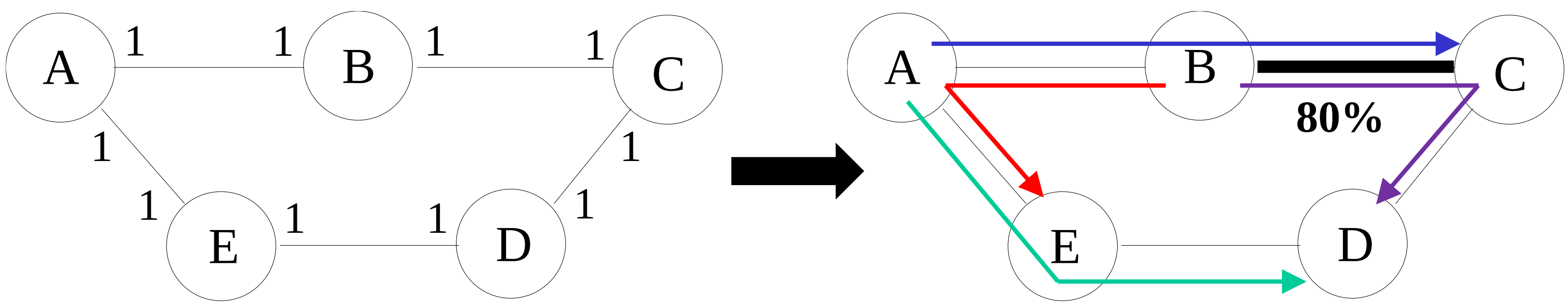
$\lambda_{A \to C} = 1000 \text{ pps}$

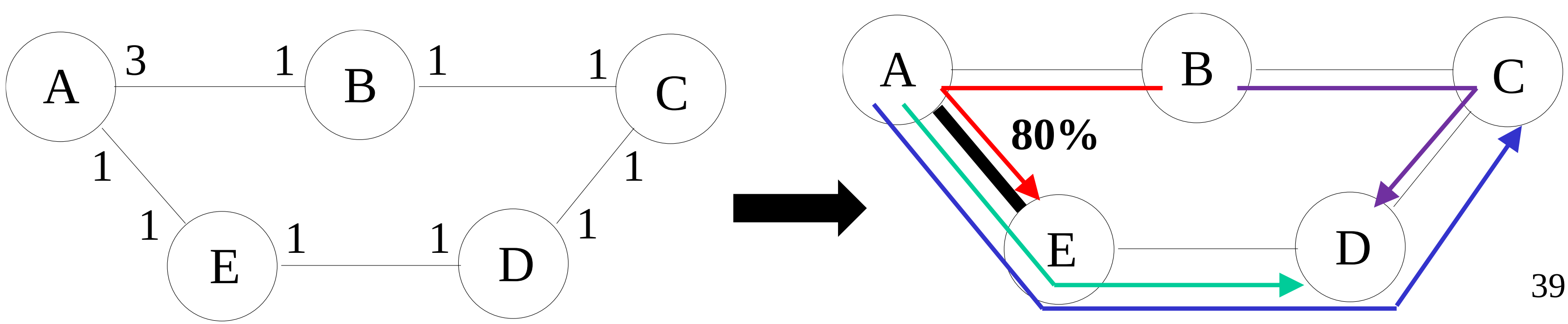
$\lambda_{A \to D} = 250 \text{ pps}$

$\lambda_{B \to D} = 1000 \text{ pps}$

$\lambda_{B \to E} = 750 \text{ pps}$

(a) Determine os custos OSPF que permitem minimizar a utilização da ligação mais carregada.





Exemplo 3

- Ligações bidirecionais de 10 Mbps
- Pacotes de 500 bytes, em média
- Propagação da ligação B-C de 10 ms em cada sentido
- <u>Encaminhamento OSPF</u>

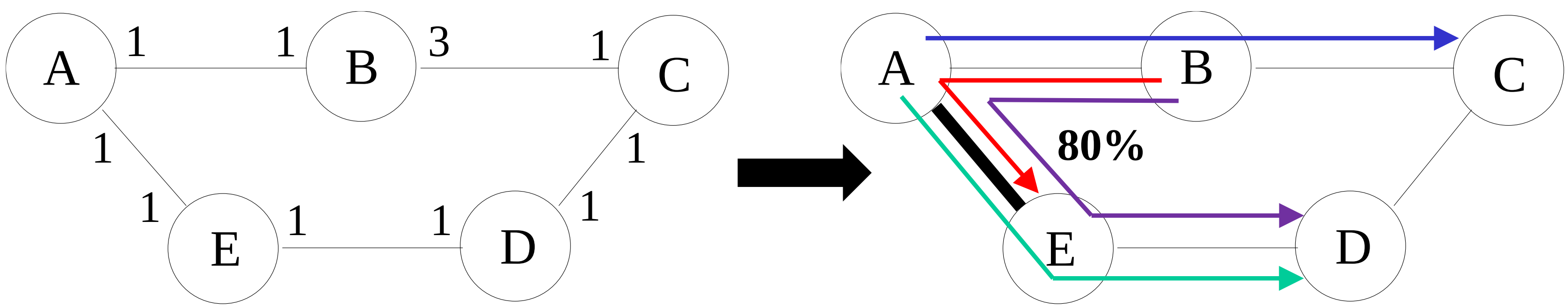
$\lambda_{A\to C} = 1000\ \text{pps}$

$\lambda_{A\to D} = 250\ \text{pps}$

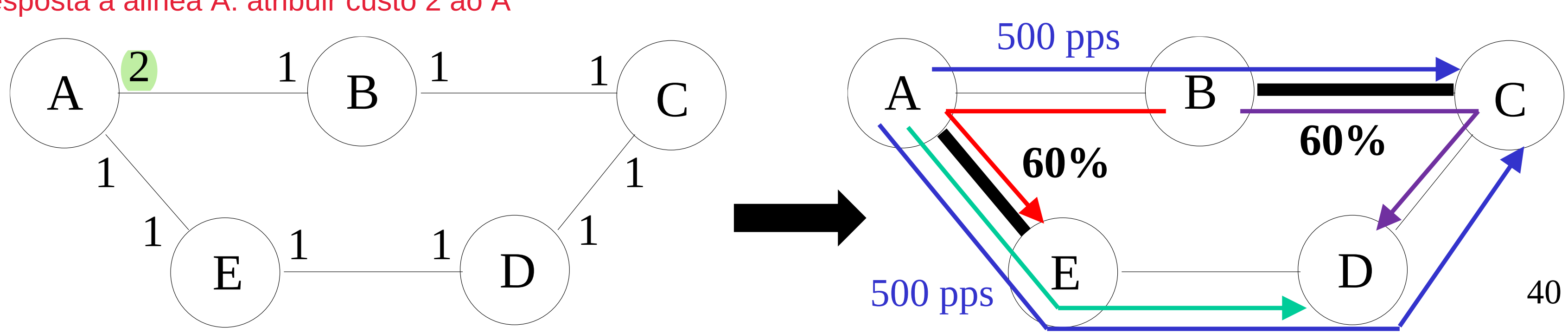
$\lambda_{B\to D} = 1000\ \text{pps}$

$\lambda_{B\to E} = 750\ \text{pps}$

(a) Determine os custos OSPF que permitem minimizar a utilização da ligação mais carregada.



Resposta à alinea A: atribuir custo 2 ao A



Exemplo 3

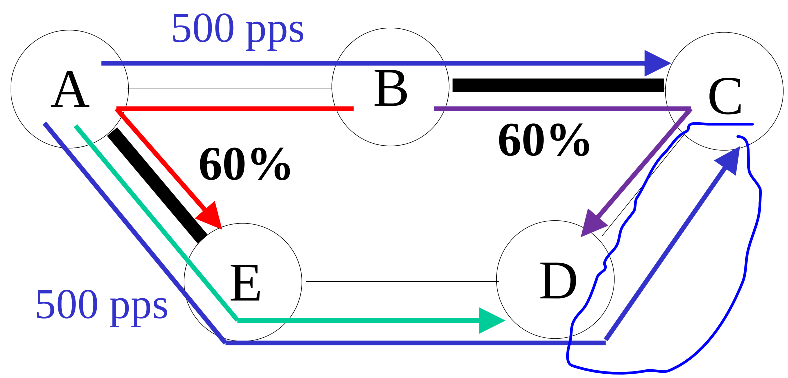
$\lambda_{A \to C} = 1000 \text{ pps}$

$\lambda_{A \to D} = 250 \text{ pps}$

$\lambda_{B \to D} = 1000 \text{ pps}$

$\lambda_{B \to E} = 750 \text{ pps}$

- Ligações bidirecionais de 10 Mbps
- Pacotes de 500 bytes, em média
- Propagação da ligação B-C de 10 ms em cada sentido
- Encaminhamento OSPF



(b) Utilizando a aproximação de Kleinrock, determine o atraso médio por pacote de todos os fluxos na solução anterior.

$$\mu_{AB} = \mu_{BA} = \mu_{BC} = \ldots = \mu = \frac{10 \times 10^6 \text{ bps}}{500 \times 8 \text{ bpp}} = 2500 \text{ pps}$$

$$\gamma = \lambda_{A \to C} + \lambda_{A \to D} + \lambda_{B \to D} + \lambda_{B \to E} = 1000 + 250 + 1000 + 750 = 3000 \text{ pps}$$

$$W = \frac{1}{\gamma} \times \left(\frac{\lambda_{AB}}{\mu_{AB} - \lambda_{AB}} + \frac{\lambda_{BA}}{\mu_{BA} - \lambda_{BA}} + \frac{\lambda_{BC}}{\mu_{BC} - \lambda_{BC}} + \lambda_{BC} d_{BC} + \frac{\lambda_{CD}}{\mu_{CD} - \lambda_{CD}} + \frac{\lambda_{DC}}{\mu_{DC} - \lambda_{DC}} + \frac{\lambda_{AE}}{\mu_{AE} - \lambda_{AE}} + \frac{\lambda_{ED}}{\mu_{ED} - \lambda_{ED}} \right)$$

$$W = \frac{1}{3000} \times \left(\frac{500}{2500 - 500} + \frac{750}{2500 - 750} + \frac{1500}{2500 - 1500} + 1500 \times 0.01 + \frac{1000}{2500 - 1000} + \frac{500}{2500 - 500} + \frac{1500}{2500 - 1500} + \frac{750}{2500 - 750} \right)$$

$W = 0.00667 \text{ seg.}$ $\quad$ $(\text{Exemplo 2: } W = 0.00865 \text{ seg.})$

Métodos de Modelação Estocástica

Modelação e Desempenho de Redes e Serviços

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

DETI-UA, 2023/2024

Experiência aleatória

- Numa experiência aleatória, o *espaço de resultados*, S, é o conjunto de todos os resultados possíveis da experiência
- Qualquer subconjunto E do espaço de resultados S designa-se por evento ou *acontecimento*

- Dados dois acontecimentos E e F, podem-se definir outros acontecimentos:
  - A *união dos acontecimentos*, $E \cup F$, é o conjunto de resultados possíveis que pertence a pelo menos um dos acontecimentos
  - A *intersecção dos acontecimentos*, EF, é o conjunto de resultados possíveis que pertence simultaneamente aos dois acontecimentos

---

- Quando $EF = \emptyset$ ($\emptyset$ é o conjunto vazio) os acontecimentos dizem-se *mutuamente exclusivos*
- O *complemento de E*, E^c, é o conjunto de resultados possíveis de S que não pertencem a E

Probabilidades definidas sobre acontecimentos

- Para cada acontecimento E de S, admite-se a existência de um número $P(E)$ designado por probabilidade de E, se satisfaz as seguintes condições:

  (1) $0 \leq P(E) \leq 1$

  (2) $P(S) = 1$

  (3) Para qualquer conjunto de acontecimentos mutuamente exclusivos E_1, E_2, E_3, ...

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

- Corolários:

  $P(E) + P(E^c) = 1$

  $P(E \cup F) = P(E) + P(F) - P(EF)$

Probabilidades condicionadas

- Dados dois acontecimentos E e F, a *probabilidade condicionada de E ocorrer dado que F ocorreu* designa-se por $P(E|F)$ e é definida por

$$P(E|F) = P(EF) / P(F)$$

- Dois acontecimentos E e F dizem-se *acontecimentos independentes* se

$$P(EF) = P(E)P(F)$$

- Se E e F são independentes, então:

$$P(E|F) = P(EF) / P(F) = P(E)P(F) / P(F) = P(E)$$

$$P(F|E) = P(FE) / P(E) = P(F)P(E) / P(E) = P(F)$$

ou seja, se o conhecimento que um acontecimento ocorreu não afetar a probabilidade do outro ter ocorrido.

Regra de Bayes

Sejam F_1, F_2, …, F_n acontecimentos mutuamente exclusivos tais que a sua união forma o espaço de resultados S. Então:

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E \mid F_i) P(F_i)$$

Tendo ocorrido o acontecimento E, a probabilidade de F_j (j = 1, 2, …, n) ter ocorrido é dada por:

$$P(F_j \mid E) = \frac{P(EF_j)}{P(E)} = \frac{P(E \mid F_j) P(F_j)}{P(E)} = \frac{P(E \mid F_j) P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i) P(F_i)}$$

Probabilidades condicionadas – Exemplo 1

Num teste de escolha múltipla, um estudante sabe a resposta certa com probabilidade p e adivinha a resposta com probabilidade $1 - p$. Ao adivinhar a resposta, o estudante acerta com probabilidade $1/m$, sendo m o número de alternativas de escolha múltipla.

Determine a probabilidade de um estudante *(i)* responder corretamente a uma pergunta e *(ii)* saber a resposta dado que a respondeu corretamente.

Acontecimentos: E – o aluno responde corretamente
F_1 – o aluno sabe a resposta
F_2 – o aluno não sabe a resposta

(i) $P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2)$

$= 1 \times p + 1/m \times (1 - p) =$

$= p + (1 - p)/m$

(ii) $P(F_1|E) = P(E|F_1)P(F_1) \,/\, P(E)$

$= 1 \times p \,/\, [p + (1 - p)/m] =$

$= p\, m \,/\, [1 + (m - 1)\, p]$

Se $p = 50\%$ e $m = 4$, então *(i)* $P(E) = 62.5\%$ e *(i)* $P(F_1|E) = 80\%$

Probabilidades condicionadas – Exemplo 2

Numa ligação sem fios (wireless) entre dois equipamentos, a probabilidade dos pacotes de dados serem recebidos com erros é de 0.1% em condições normais ou de 10% quando há interferências. A probabilidade de haver interferência é de 2%. Os equipamentos têm a capacidade de verificar na receção se os pacotes de dados foram recebidos com erros ou não.

Determine: *(i)* a probabilidade de um pacote ser recebido com erros e *(ii)* se um pacote for recebido com erros, qual a probabilidade da ligação estar com interferência.

Acontecimentos: E – o pacote é recebido com erros
F_1 – a ligação está em condições normais
F_2 – a ligação está com interferência

(i) $P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2)$

$= 0.001\times(1 - 0.02) + 0.1\times0.02$

$= 0.00298 = 0.298\%$

(ii) $P(F_2|E) = P(E|F_2)P(F_2) \,/\, P(E)$

$= 0.1 \times 0.02 \,/\, 0.00298$

$= 0.671 = 67.1\%$

Variáveis aleatórias

- Uma *variável aleatória X é uma função* que atribui um número real a cada ponto do espaço de resultados S de uma experiência aleatória.

- A *função distribuição* (ou função de distribuição cumulativa) da v.a. X é:

$$F(x) = P(X \leq x) \quad , -\infty < x < +\infty$$

- *Propriedades* da função distribuição:

  (1) $0 \leq F(x) \leq 1$ para todo o x

  (2) se $x_1 \leq x_2$ então $F(x_1) \leq F(x_2)$ (função não decrescente)

  (3) $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$

  (4) $P(a < X \leq b) = F(b) - F(a)$, para $a < b$

Variáveis aleatórias discretas

- Uma *variável aleatória X diz-se discreta* se puder tomar, quando muito, um número contável de valores $x_1, x_2, \ldots, x_i, \ldots$

- Define-se *função probabilidade* (ou função massa de probabilidade) da v.a discreta X por

$$f(x_i) = P(X = x_i) \qquad \text{para todos os valores de } i = 1,2,3,\ldots$$

- Obrigatoriamente, tem de acontecer que: $$\sum_{i=1}^{\infty} f(x_i) = 1$$

- A *função distribuição* da v.a discreta X é:

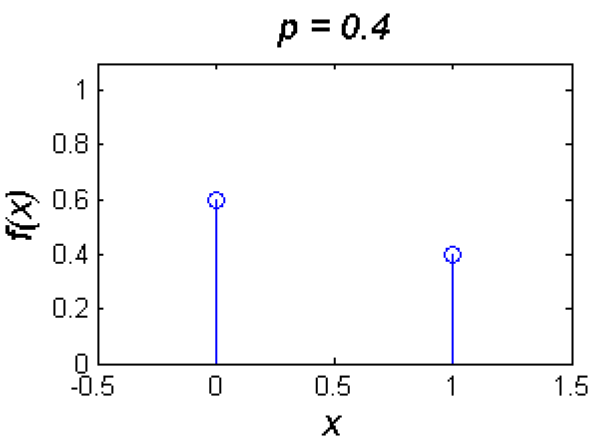
$$F(x) = \sum_{x_i \le x} f(x_i) \qquad , -\infty < x < +\infty$$

Exemplos de variáveis aleatórias discretas

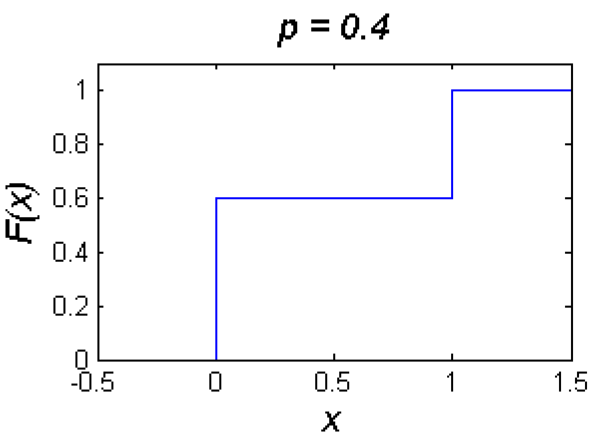
Variável aleatória de Bernoulli: experiência que pode resultar em sucesso com probabilidade *p* ou insucesso com probabilidade 1 – *p*.

Se X = 1 representar um sucesso e X = 0 um insucesso, a função probabilidade é:

$$f(i) = p^i (1-p)^{1-i}, i = 0,1$$



f(x) - função probabilidade



F(x) - função distribuição

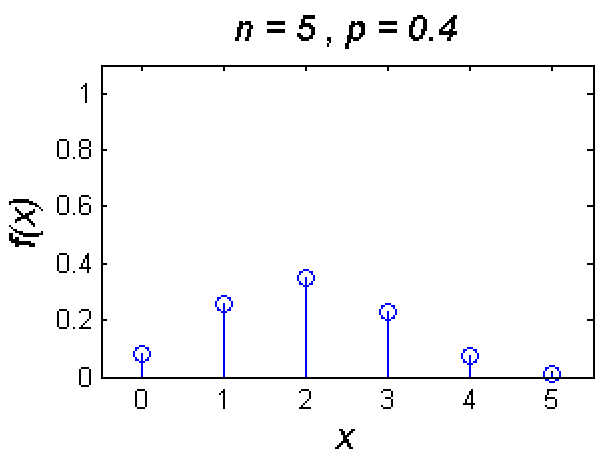
Exemplos de variáveis aleatórias discretas

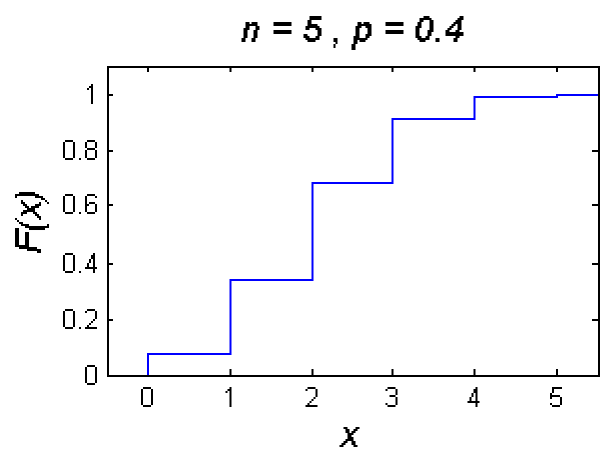
<u>*Variável aleatória binomial*</u>: conjunto de *n* experiências de Bernoulli independentes, cada uma das quais resulta num sucesso com probabilidade *p* ou num insucesso com probabilidade 1 – *p*.

Se *X* representar o número de sucessos em *n* experiências, a função probabilidade é:

$$f(i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, 2, \ldots, n$$

onde $\binom{n}{i} = \frac{n!}{i!(n-i)!}$



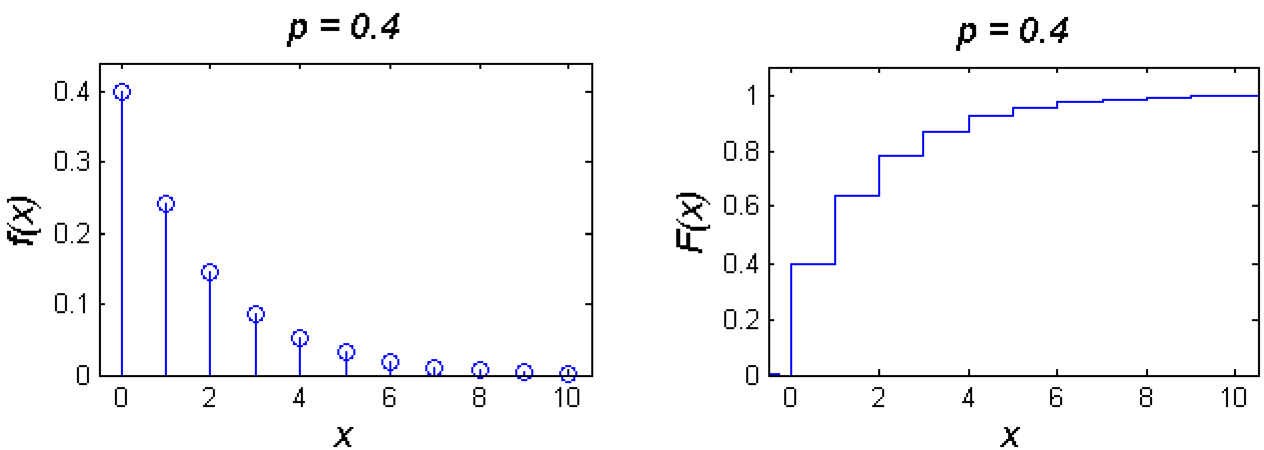


Exemplos de variáveis aleatórias discretas

Variável aleatória geométrica: são realizadas experiências de Bernoulli independentes com parâmetro *p* (probabilidade de sucesso) até que ocorra um sucesso.

Se *X* representar o número de insucessos antes do sucesso, a função probabilidade é

$$f(i) = (1-p)^i p \ , \ i = 0, 1, 2, \ldots$$



Se *X* representar o número de experiências até ao sucesso, a função probabilidade é

$$f(i) = (1-p)^{i-1} p \ , \ i = 1, 2, \ldots$$

Variáveis aleatórias discretas – Exemplo 3

Numa dada ligação de dados, a probabilidade de erro de bit (BER – *Bit Error Rate*) é 10^{-5} e os erros em diferentes bits são estatisticamente independentes.

Determine: *(i)* a probabilidade de um pacote de dados de 100 Bytes ser recebido sem bits errados e *(ii)* a probabilidade de um pacote de dados de 1000 Bytes ser recebido com 2 ou mais bits errados.

O número de bits errados num pacote é uma variável aleatória binomial em que a probabilidade de sucesso é o BER e o número de experiências de Bernoulli é o número de bits do pacote

$$f(i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, 2, ..., n$$

(i) $$f(0) = \binom{n}{0} p^0 (1-p)^{n-0} = \binom{100\times 8}{0} \times \left(1-10^{-5}\right)^{100\times 8} = 0.992 = 99.2\%$$

(ii) $$1 - f(0) - f(1) = 1 - \binom{n}{0} p^0 (1-p)^{n-0} - \binom{n}{1} p^1 (1-p)^{n-1}$$

$$= 1 - \left(1-10^{-5}\right)^{8000} - 8000 \times 10^{-5}\left(1-10^{-5}\right)^{7999} = 3.034\text{E-}3 = 0.3\%$$

Variáveis aleatórias contínuas

- Uma *variável aleatória X diz-se contínua* se existir uma função não negativa *f*(*x*) tal que para qualquer conjunto de números reais *B*:

$$P(X \in B) = \int_B f(x)dx \qquad \int_{-\infty}^{+\infty} f(x)dx = 1$$

f(*x*) é a *função densidade de probabilidade* da v.a contínua *X*

- Resulta então que: $P\{a \leq X \leq b\} = \int_a^b f(x)dx$

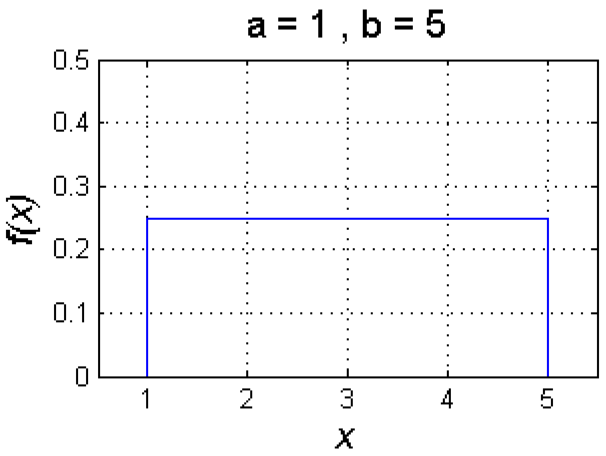
- A *função distribuição* da v.a contínua *X* é:

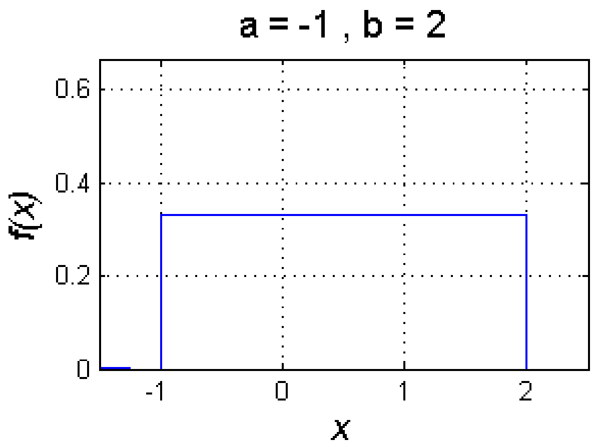
$$F(x) = P(X \in [-\infty, x]) = \int_{-\infty}^{x} f(y)dy$$

Exemplos de variáveis aleatórias contínuas

Variável aleatória com Distribuição Uniforme: uma v.a. diz-se uniformemente distribuída no intervalo [*a*,*b*] se a função densidade de probabilidade for dada por

$$f(x)=\begin{cases}\dfrac{1}{b-a} & , a<x<b \\ 0 & , cc\end{cases}$$

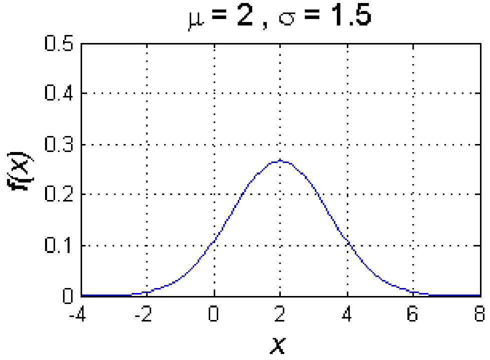




Exemplos de variáveis aleatórias contínuas

<u>Variável aleatória com Distribuição Gaussiana (ou Normal)</u>: Uma v.a. X tem uma distribuição Gaussiana com média μ e desvio padrão σ se a função densidade é dada por:

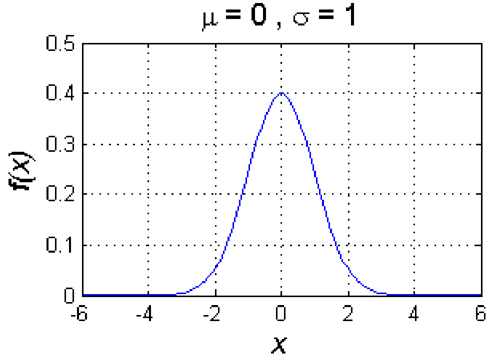
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$



Designa-se por distribuição Gaussiana (ou Normal) padrão à distribuição Gaussiana com média 0 e desvio padrão 1.

Neste caso:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



Média de uma variável aleatória

- *<u>Média</u>* ou *<u>valor esperado</u>* de uma v.a. *X*:

$$E[X] = \begin{cases} \sum_{j=1}^{\infty} x_j f_X(x_j) & se\ X\ discreta \\ \int_{-\infty}^{+\infty} x f_X(x) dx & se\ X\ continua \end{cases}$$

- Propriedades importantes:

$$E[cX] = cE[X]$$

$$E\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i E[X_i]$$

- Média da v.a. $Y = g(X)$:

$$E[g(X)] = \begin{cases} \sum_{j=1}^{\infty} g(x_j) f_X(x_j) & se\ X\ discreta \\ \int_{-\infty}^{+\infty} g(x) f_X(x) dx & se\ X\ continua \end{cases}$$

Variância e desvio padrão de uma variável aleatória

- *Variância* de uma v.a. *X*:

$$Var[X] = E\left[\left(X - E[X]\right)^2\right] = E\left[X^2\right] - E[X]^2$$

2º momento da v.a. X

- Propriedades importantes da variância:

$$Var[X] \geq 0$$

$$Var[cX] = c^2 \, Var[X]$$

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] \quad \text{se } X_i \text{ forem independentes}$$

- *Desvio padrão* de uma v.a. *X*:

$$\sigma[X] = \sqrt{Var[X]}$$

Exemplo 4

Uma ligação de dados de 10 Mbps suporta um fluxo de pacotes cujo tamanho é 100 Bytes com probabilidade 10%, 500 Bytes com probabilidade 50% e 1500 Bytes com probabilidade 40%. Considere a variável aleatória *X* representativa do tempo de transmissão dos pacotes.

Determine: *(i)* o tempo médio de transmissão dos pacotes *E*[*X*], *(ii)* o segundo momento do tempo de transmissão dos pacotes *E*[X^2] e *(iii)* a variância do tempo de transmissão dos pacotes *Var* [*X*].

(i)
$$E[X] = \sum_{j=1}^{\infty} x_j f_X(x_j) = \frac{100\times 8}{10^7}\times 0.1 + \frac{500\times 8}{10^7}\times 0.5 + \frac{1500\times 8}{10^7}\times 0.4$$
$$= 0.688\times 10^{-3}\ \text{seg} = 0.688\ \text{mseg}$$

(ii)
$$E[X^2] = \sum_{j=1}^{\infty} (x_j)^2 f_X(x_j) = \left(\frac{100\times 8}{10^7}\right)^2\times 0.1 + \left(\frac{500\times 8}{10^7}\right)^2\times 0.5 + \left(\frac{1500\times 8}{10^7}\right)^2\times 0.4$$
$$= 6.5664\times 10^{-7}\ \text{seg}^2$$

Exemplo 4 - continuação

Uma ligação de dados de 10 Mbps suporta um fluxo de pacotes cujo tamanho é 100 Bytes com probabilidade 10%, 500 Bytes com probabilidade 50% e 1500 Bytes com probabilidade 40%. Considere a variável aleatória *X* representativa do tempo de transmissão dos pacotes.

Determine: *(i)* o tempo médio de transmissão dos pacotes $E[X]$, *(ii)* o segundo momento do tempo de transmissão dos pacotes $E[X^2]$ e *(iii)* a variância do tempo de transmissão dos pacotes $Var[X]$.

(iii) **1ª alternativa:** $Var[X] = E\left[\left(X - E[X]\right)^2\right]$

$$Var[X] = \left(\frac{100 \times 8}{10^7} - E[X]\right)^2 \times 0.1 + \left(\frac{500 \times 8}{10^7} - E[X]\right)^2 \times 0.5 + \left(\frac{1500 \times 8}{10^7} - E[X]\right)^2 \times 0.4$$

$$= 1.833 \times 10^{-7}\ \text{seg}^2$$

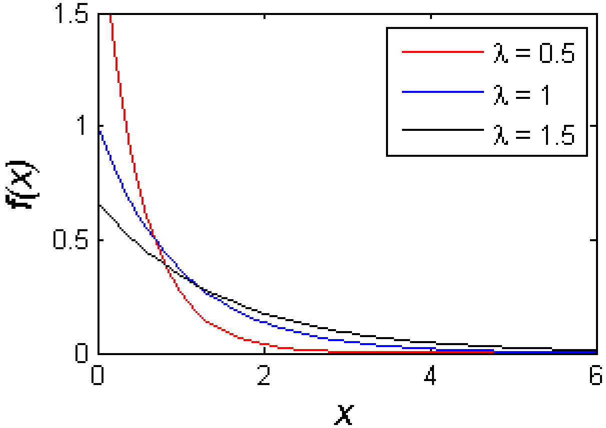
2ª alternativa: $Var[X] = E\left[X^2\right] - E[X]^2$

$$Var[X] = 6.5664 \times 10^{-7} - \left(0.688 \times 10^{-3}\right)^2 = 1.833 \times 10^{-7}\ \text{seg}^2$$

Distribuição exponencial

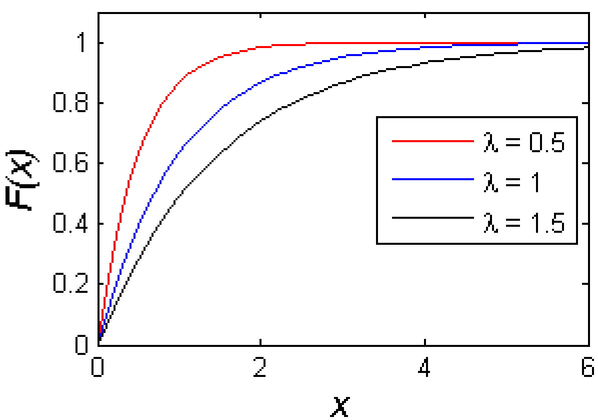
- Uma v. a. contínua *X* tem uma distribuição exponencial com parâmetro λ, $\lambda > 0$, se a sua função densidade de probabilidade for:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



- A função distribuição é dada por:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



Distribuição exponencial

- A média, a variância e o desvio padrão de uma distribuição exponencial são:

$$E[X] = \frac{1}{\lambda} \qquad Var[X] = \left(\frac{1}{\lambda}\right)^2 \qquad \sigma[X] = \frac{1}{\lambda}$$

- A distribuição exponencial não tem memória, isto é,

$$P\{X > s + t \mid X > t\} = P\{X > s\}$$

- Se X_1 e X_2 são v. a. independentes e exponencialmente distribuídas com médias $1/\lambda_1$ e $1/\lambda_2$ respetivamente, então

$$P\{X_1 < X_2\} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Distribuição exponencial – Exemplo 5

Uma ligação de dados com a capacidade de 10 Mbps suporta um fluxo de pacotes cujo tamanho é exponencialmente distribuído com média de 1000 Bytes. Considere a variável aleatória *X* representativa do tempo de transmissão dos pacotes.

Determine: *(i)* o tempo médio de transmissão dos pacotes *E*[*X*], *(ii)* a variância do tempo de transmissão dos pacotes *Var*[*X*] e *(iii)* o segundo momento do tempo de transmissão dos pacotes $E[X^2]$.

(i) $$E[X] = \frac{1000 \times 8}{10^7} = 8 \times 10^{-4} = 0.8 \text{ mseg}$$

Capacidade da ligação em pacotes por segundo

$$E[X] = \frac{1}{\mu} \quad \Leftrightarrow \quad \mu = \frac{1}{E[X]} = \frac{1}{8 \times 10^{-4}} = 1250 \text{ pacotes/s}$$

(ii) $$Var[X] = \left(\frac{1}{\mu}\right)^2 = \left(8 \times 10^{-4}\right)^2 = 6.4 \times 10^{-7} \text{ seg}^2$$

E[X^2]= 2 x (E[X])^2 = 2x(0.0008)^2 = 1.28e-6

(iii) $$\cancel{Var[X] = E[X^2] - E[X]^2 \quad \Leftrightarrow \quad E[X^2] = Var[X] + E[X]^2}$$

$$E[X^2] = 6.4 \times 10^{-7} + \left(8 \times 10^{-4}\right)^2 = 1.28 \times 10^{-6} \text{ seg}^2$$

Processos estocásticos

- Um *processo estocástico* $\{X(t),\ t \in T\}$ é um conjunto de variáveis aleatórias: para cada $t \in T$, $X(t)$ é uma variável aleatória.

- O índice t é frequentemente interpretado como tempo. Nesta interpretação, $X(t)$ é o *estado* do processo no instante t.

- O conjunto T é o *conjunto de índices* do processo.

  (1) se T é um conjunto contável, designa-se o processo estocástico como sendo *em tempo discreto*

  (2) se T é um intervalo da reta real, designa-se o processo estocástico como sendo *em tempo contínuo*

- O *espaço de estados* é o conjunto de todos os valores que as variáveis aleatórias $X(t)$ podem tomar.

Exemplos de processos estocásticos

Considere um sistema com uma fila de espera e um servidor. A este sistema chegam clientes para serem servidos.

Atrasos sofridos por cada cliente

na fila de espera

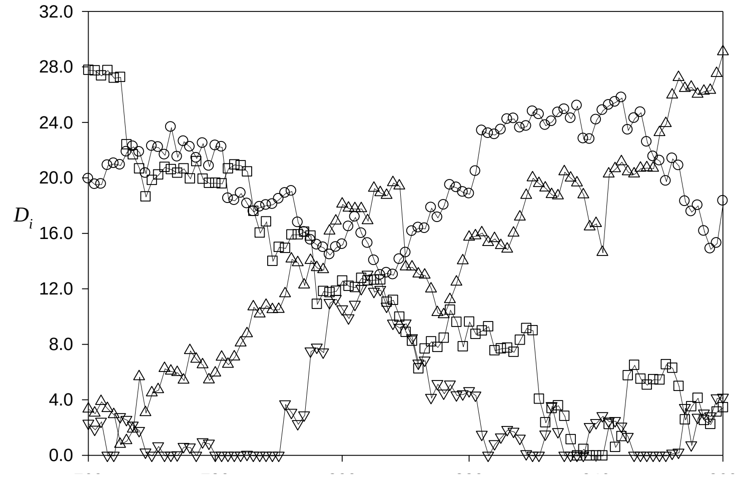
(1) é um processo estocástico em

tempo discreto (1º cliente,

2º cliente, etc.)

(2) o estado é uma variável contínua

(o tempo de espera é um valor real)



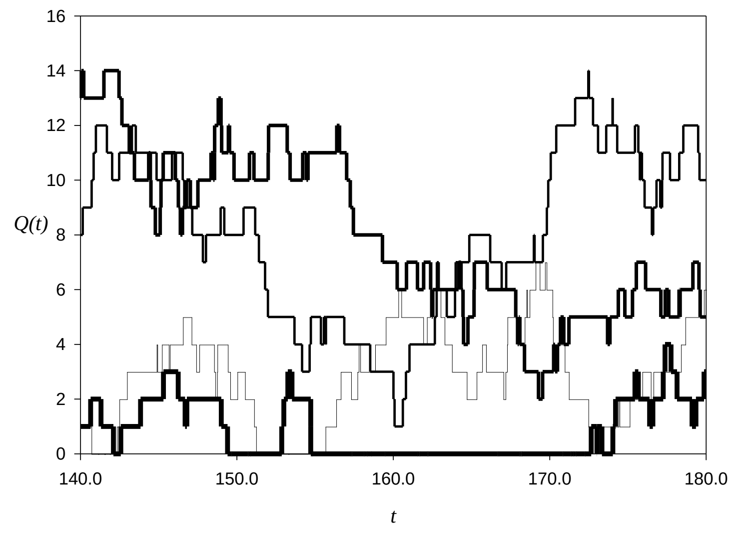
O número de clientes em espera

(1) é um processo estocástico

em tempo contínuo

(2) o estado é uma variável discreta

(0 clientes, 1 cliente, 2 clientes, etc.)



Processo de contagem

- Um processo estocástico $\{N(t), t \geq 0\}$ diz-se um <u>*processo de contagem*</u> se $N(t)$ representar o número total de eventos que ocorreram até ao instante t.
- Um processo de contagem satisfaz as seguintes condições:

  (1) $N(t) \geq 0$.

  (2) $N(t)$ toma valores inteiros apenas.

  (3) Se $s < t$, então $N(s) \leq N(t)$.

  (4) Se $s < t$, então $N(t) - N(s)$ é igual ao número de eventos ocorridos no intervalo de tempo $[s,t]$.
- Um processo de contagem tem <u>*incrementos independentes*</u> se o número de eventos em intervalos de tempo disjuntos for independente.
- Um processo de contagem tem <u>*incrementos estacionários*</u> se a distribuição do número de eventos que ocorre em qualquer intervalo de tempo depender apenas do comprimento do intervalo de tempo.

Processo de Poisson

- Um processo de contagem diz-se um *processo de Poisson* com taxa λ, $\lambda > 0$, se:

  (1) $N(0) = 0$;

  (2) o processo tem incrementos independentes; chegadas no futuro sao ind do passado

  (3) o número de eventos num intervalo de duração t tem uma distribuição de Poisson com média λt. Isto é, para todo s, $t \geq 0$

$$P\{N(s+t)-N(s)=n\}=e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

- Um processo de Poisson tem incrementos estacionários e média

$$E[N(t)] = \lambda t$$

razão pela qual λ é designada a *taxa* (i.e., o número médio de eventos por unidade de tempo) do processo de Poisson.

Propriedades de um processo de Poisson

- **Propriedade 1:** Num processo de Poisson com taxa λ considere-se:
  - T_1 o instante do primeiro evento
  - T_n, n > 1, o intervalo de tempo entre o (n-1)-ésimo evento e o n-ésimo evento
- Então, T_n, n = 1, 2, …, são variáveis aleatórias independentes e identicamente distribuídas com <u>*distribuição exponencial*</u> de média $1/\lambda$.
- **Propriedade 2:** Num processo de Poisson $\{N(t), t \geq 0\}$ com taxa λ considere-se que cada evento é classificado de forma independente em:
  - evento do tipo 1 com probabilidade p
  - evento do tipo 2 com probabilidade $1 – p$
- Ou seja, $\{N_1(t), t \geq 0\}$ e $\{N_2(t), t \geq 0\}$ são o número de eventos de cada tipo que ocorreram no intervalo $[0,t]$.
- Então, $N_1(t)$ e $N_2(t)$ são <u>*ambos processos de Poisson independentes*</u> com taxas λp e $\lambda(1–p)$.

Propriedades de um processo de Poisson

- **Propriedade 3:** Sejam $\{N_1(t), t \geq 0\}$ e $\{N_2(t), t \geq 0\}$ processos de Poisson independentes com taxas λ_1 e λ_2,

40% Homens + 20% Mulheres logo o processo poisson é 60 %

- Então, o processo $N(t) = N_1(t) + N_2(t)$ é também *<u>um processo de Poisson</u>* com taxa $\lambda = \lambda_1 + \lambda_2$.

- **Propriedade 4:** Sabendo-se que num processo de Poisson ocorreram exatamente n eventos até ao instante t,

- Então, os instantes de ocorrência dos eventos são *<u>distribuídos independentemente e uniformemente</u>* no intervalo $[0, t]$. Por esta razão diz-se que num processo de Poisson as chegadas são *<u>aleatórias</u>*.

3 formas de simular:

escolher intervalo de tempo e depois tiramos o N para esse intervalo com a distribuição

ou estima-se um evento de cada vez .....

Cadeias de Markov em tempo contínuo

- Considere-se um processo estocástico em tempo contínuo $\{X(t),\ t \geq 0\}$ com o espaço de estados definido pelo conjunto dos números inteiros não negativos (i.e., {0, 1, 2, …}).
- $X(t)$ *é uma cadeia de Markov* se para todo o s, $t \geq 0$ e inteiros não-negativos i, j, $x(u)$, $0 \leq u < s$:

$$P\{X(s+t) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s\} = P\{X(s+t) = j \mid X(s) = i\}$$

- Significa que a distribuição futura $X(s+t)$ condicionada ao presente $X(s)$ e ao passado $X(u)$, $0 \leq u < s$, depende apenas do presente e é independente do passado (propriedade *Markoviana*).

- Se $P\{X(s+t) = j \mid X(s) = i\}$ for independente de s então diz-se que a cadeia de Markov em tempo contínuo tem *probabilidades de transição estacionárias* ou *homogéneas*:

$$P\{X(s+t) = j \mid X(s) = i\} = P\{X(t) = j \mid X(0) = i\}$$

independentemente de tar no inicio ou mais à frente no tempo

Cadeias de Markov em tempo contínuo

- Uma cadeia de Markov em tempo contínuo tem como propriedades:

  (1) Quando o processo entra no estado i, o tempo de permanência nesse estado, antes de efetuar uma transição para um estado diferente, é exponencialmente distribuído (designamos a média por $1/q_i$);

  NOTA: Esta propriedade é equivalente a dizer que quando o processo está no estado i, ele transita para outro estado qualquer a uma taxa q_i.

  (2) Quando o processo deixa o estado i, entra de seguida no estado j com uma probabilidade P_{ij} que satisfaz as seguintes condições

$$P_{ii} = 0 \qquad 0 \leq P_{ij} \leq 1 \quad , j \neq i \qquad \sum_j P_{ij} = 1$$

- Numa cadeia de Markov em tempo contínuo, o tempo de permanência num estado e o próximo estado visitado são variáveis aleatórias independentes.

Taxas de transição instantâneas

qi = taxa da prob de saltar para os outros estados a partir de i

- Para qualquer par de estados i e j seja:

$$q_{ij} = q_i P_{ij}$$

q_i - a taxa à qual o processo faz uma transição quando está no estado i (definida no slide anterior)

P_{ij} - a probabilidade que a transição seja para o estado j quando está no estado i (definida no slide anterior)

q_{ij} - a taxa à qual o processo faz uma transição para o estado j quando está no estado i

- As q_{ij} designam-se por *<u>taxas de transição instantâneas</u>*. Estas são as grandezas habitualmente representadas nos diagramas de transição de estados.

- Como $\quad q_i = \sum_j q_i P_{ij} = \sum_j q_{ij} \qquad P_{ij} = \dfrac{q_{ij}}{q_i} = \dfrac{q_{ij}}{\sum_j q_{ij}}$

resulta que a especificação das taxas de transição instantâneas determina a cadeia de Markov em tempo contínuo.

Probabilidades limite

- Seja $$P_{ij}(t) = P\{X(s+t) = j \mid X(s) = i\}$$

  a probabilidade de um processo presentemente no estado *i* estar no estado *j* após um intervalo de tempo *t*.

- A probabilidade de uma cadeia de Markov em tempo contínuo estar no estado *j* no instante *t* converge para um valor limite independente do estado inicial:

$$\pi_j \equiv \lim_{t \to \infty} P_{ij}(t)$$

- Condição suficiente para a existência de probabilidades limite:
  (1) a cadeia é irredutível, isto é, começando no estado *i* existe uma probabilidade positiva de alguma vez se estar no estado *j*, para todo o par de estados *i*, *j*
  (2) a cadeia de Markov é recorrente positiva, isto é, começando em qualquer estado o tempo médio para voltar a esse estado é finito

Cálculo das probabilidades limite

- As probabilidades limite podem calcular-se resolvendo as equações:

$$q_j \pi_j = \sum_{k \neq j} q_{kj} \pi_k \,, \qquad \text{para todos os estados } j$$

$$\sum_j \pi_j = 1$$

- Estas equações são designadas por <u>*equações de balanço*</u>:

taxa à qual o sistema transita do estado *j*
=
taxa à qual o sistema transita para o estado *j*

- A probabilidade π_j pode ser interpretada como a proporção de tempo em que o processo está no estado *j*.

- As probabilidades π_j são designadas por <u>*probabilidades estacionárias*</u>: se o estado inicial for dado pela distribuição $\{\pi_j\}$, então a probabilidade de se estar no estado *j* no instante *t* é π_j, para todo o *t*.

Processos de nascimento e morte

- Considere um sistema cujo estado representa o número de clientes no sistema.

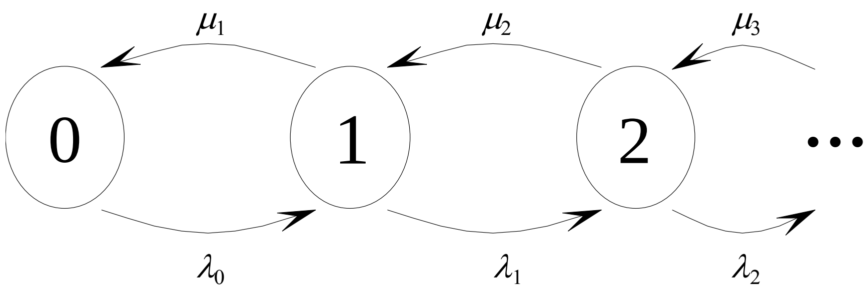
- Sempre que o sistema tem n clientes:

  (1) chegam novos clientes ao sistema a uma taxa exponencial λ_n

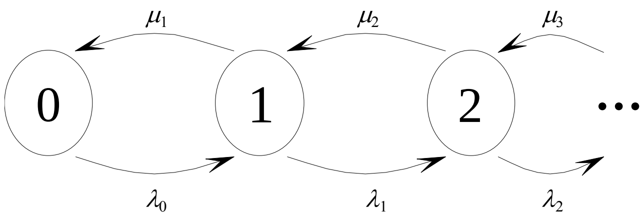
  (2) partem clientes do sistema a uma taxa exponencial μ_n

- Este sistema é designado por <u>*processo de nascimento e morte*</u>.

- Os parâmetros λ_n (n = 0, 1, …) e μ_n (n = 1, 2, …) são designados por <u>*taxas de chegada*</u> (ou de nascimento) e <u>*taxas de partida*</u> (ou de morte), respetivamente.



Equações de balanço de processos de nascimento e morte



Num processo de nascimento e morte, é possível calcular as probabilidade limite π_n de cada estado n (= 0, 1, 2, …) da seguinte forma.

Equações de balanço: $\boxed{q_j \pi_j = \sum_{k \neq j} q_{kj} \pi_k}$

Estado	*taxa de saída = taxa de entrada*
0	$\lambda_0 \pi_0 = \mu_1 \pi_1$
1	$(\lambda_1 + \mu_1)\pi_1 = \mu_2 \pi_2 + \lambda_0 \pi_0$
2	$(\lambda_2 + \mu_2)\pi_2 = \mu_3 \pi_3 + \lambda_1 \pi_1$
n, $n \geq 1$	$(\lambda_n + \mu_n)\pi_n = \mu_{n+1} \pi_{n+1} + \lambda_{n-1} \pi_{n-1}$

Ou, de forma equivalente (por manipulação das equações anteriores):

$$\lambda_n \pi_n = \mu_{n+1} \pi_{n+1}, \qquad n \geq 0$$

Probabilidades limite de processos de nascimento e morte

Probabilidade limite de cada estado:

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

$$\pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n \left(1+\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}\right)} = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \cdot \pi_0, \;\; n \geq 1$$

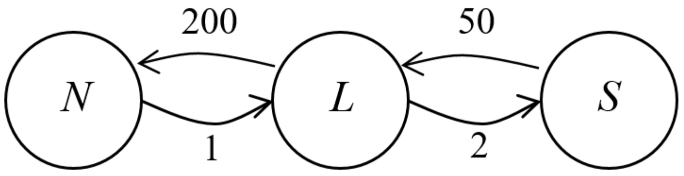
NOTA: Se o processo tiver um número finito *N* de estados (i.e., *n* = 0, 1, …, *N*), o somatório das expressões é de 0 até ao número de estados *N*.

Condição necessária para a existência de probabilidades limite (no caso de um número infinito de estados):

$$\sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} < \infty$$

Exemplo 6

Considere que uma ligação sem fios para comunicação de pacotes pode estar num de 3 estados possíveis – Normal (*N*), Interferência Ligeira (*L*) ou Interferência Severa (*S*) – de acordo com a cadeia de Markov seguinte (taxas em número de transições por hora):



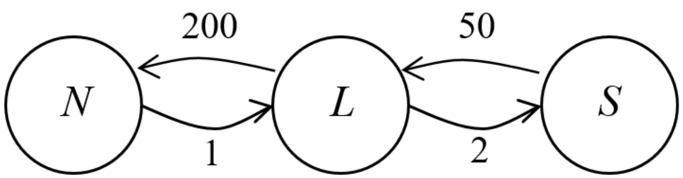
(a) Determine a probabilidade de cada um dos estados.

(b) Determine o tempo de permanência médio de cada estado (em minutos).

(c) Sabendo que a probabilidade de cada pacote ser recebido com erros é de 0.01% no estado *N*, 0.1% no estado *L* e 1% no estado *S*, qual a probabilidade da ligação estar no estado *N* quando um pacote é recebido com erros?

Exemplo 6 – Resolução (a)

Considere que uma ligação sem fios para comunicação de pacotes pode estar num de 3 estados possíveis – Normal (*N*), Interferência Ligeira (*L*) ou Interferência Severa (*S*) – de acordo com a cadeia de Markov seguinte (taxas em número de transições por hora):



(a) Determine a probabilidade de cada um dos estados.

$$P_N = \frac{1}{1 + \frac{1}{200} + \frac{1}{200} \times \frac{2}{50}} = 0.99483 = 99.483\%$$

99,48 % das vezes está no estado normal

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

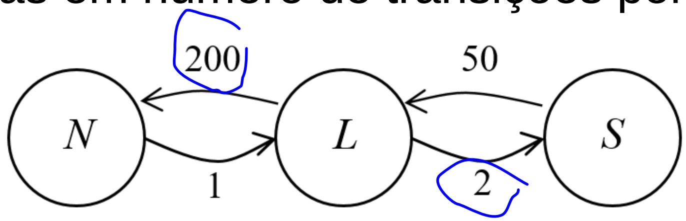
$$P_L = \frac{\frac{1}{200}}{1 + \frac{1}{200} + \frac{1}{200} \times \frac{2}{50}} = 0.00497 = 0.497\%$$

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

$$P_S = \frac{\frac{1}{200} \times \frac{2}{50}}{1 + \frac{1}{200} + \frac{1}{200} \times \frac{2}{50}} = 0.0002 = 0.02\%$$

Exemplo 6 – Resolução (b)

Considere que uma ligação sem fios para comunicação de pacotes pode estar num de 3 estados possíveis – Normal (*N*), Interferência Ligeira (*L*) ou Interferência Severa (*S*) – de acordo com a cadeia de Markov seguinte (taxas em número de transições por hora):



(b) Determine o tempo de permanência médio de cada estado (em minutos).

$$T_N = \frac{1}{1} = 1 \text{ hora} = 60 \text{ minutos}$$

$$T_L = \frac{1}{2+200} = 0.00495 \text{ horas} = 0.3 \text{ minutos}$$

$$T_S = \frac{1}{50} = 0.02 \text{ horas} = 1.2 \text{ minutos}$$

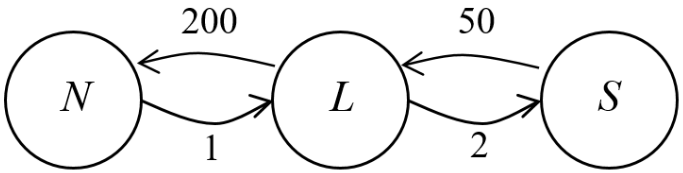
Tempo médio de permanência $T = 1/q_i$

$$q_i = \sum_j q_i P_{ij} = \sum_j q_{ij}$$

Exemplo 6 – Resolução (c)

Considere que uma ligação sem fios para comunicação de pacotes pode estar num de 3 estados possíveis – Normal (*N*), Interferência Ligeira (*L*) ou Interferência Severa (*S*) – de acordo com a cadeia de Markov seguinte (taxas em número de transições por hora):

0.99483 = 1-(1-0.01)



(c) Sabendo que a probabilidade de cada pacote ser recebido com erros é de 0.01% no estado *N*, 0.1% no estado *L* e 1% no estado *S*, qual a probabilidade da ligação estar no estado *N* quando um pacote é recebido com erros?

$$P(N|E) = \frac{P(E|N) \times P(N)}{P(E|N) \times P(N) + P(E|L) \times P(L) + P(E|S) \times P(S)}$$

$$= \frac{0.0001 \times 0.99483}{0.0001 \times 0.99483 + 0.001 \times 0.00497 + 0.01 \times 0.00020}$$

$$= 0.9346 = 93.46\%$$

Definições do teorema de Little

- Admita-se que se observa um sistema desde o instante $t = 0$. Seja:

  $L(t)$ - o número de clientes no sistema no instante t,

  $N(t)$ - o número de clientes que chegaram no intervalo $[0,t]$,

  W_i - o tempo despendido no sistema pelo i-ésimo cliente.

- Média temporal do *número de clientes* observados até ao instante t:

$$L_t = \frac{1}{t}\int_0^t L(\tau)d\tau \qquad L = \lim_{t\to\infty} L_t$$

- Média temporal da *taxa de chegada* no intervalo $[0,t]$:

$$\lambda_t = N(t)/t \qquad \lambda = \lim_{t\to\infty} \lambda_t$$

- Média temporal do *atraso dos clientes* até ao instante t:

$$W_t = \frac{\sum_{i=0}^{N(t)} W_i}{N(t)} \qquad W = \lim_{t\to\infty} W_t$$

Teorema de Little

- O teorema de Little enuncia que

$$L = \lambda W$$

- O teorema de Little traduz a ideia intuitiva de que, para a mesma taxa de chegada de clientes, sistemas mais congestionados (L elevado) impõem maiores atrasos (W elevado).

- Exemplos:
  - Num dia de chuva, o mesmo tráfego (mesmo λ) é mais lento do que normalmente (W maior) e, consequentemente, as ruas estão mais congestionadas (L maior).
  - Um restaurante de refeições rápidas (W menor) precisa de uma sala menor (L menor) que um restaurante normal, para a mesma taxa de chegada de clientes (mesmo λ).

Se a sala é menor as refeiçoes têm de ser mais rapidas

Propriedade PASTA
(*Poisson Arrivals always See Time Averages*)

- Considere um sistema em que os clientes chegam um de cada vez e são servidos um de cada vez.
- Seja $L(t)$ o número de clientes no sistema no instante t e defina-se P_n, $n \geq 0$, como

$$P_n = \lim_{t \to \infty} P\{L(t) = n\}$$

P_n é a probabilidade em estado estacionário de existirem exatamente n clientes no sistema (ou a proporção de tempo em que o sistema contém exatamente n clientes).

- Considere a_n a proporção de clientes que ao chegar encontram n clientes no sistema.

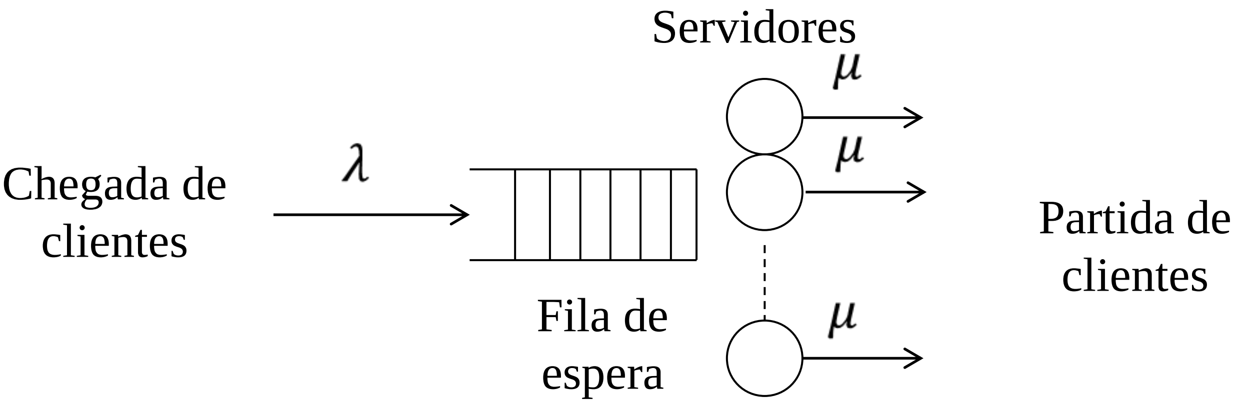
- **Propriedade PASTA:**

As <u>*chegadas de Poisson*</u> em que o <u>*tempo de serviço é estatisticamente independente dos instantes de chegada*</u>, vêem sempre médias temporais:

$$a_n = P_n$$

Sistema de fila de espera

- Um sistema de fila de espera é caracterizado por:
  - um conjunto de *c* servidores, cada um com capacidade para servir clientes a uma taxa μ
  - uma fila de espera com uma determinada capacidade (em nº de clientes)
- A este sistema chegam clientes a uma taxa λ
- Quando um cliente chega:
  - ele começa a ser servido por um servidor se houver algum disponível
  - ele é colocado da fila de espera se os servidores estiverem todos ocupados (ou é perdido se a fila de espera estiver cheia)
- Os clientes na fila de espera são atendidos segundo uma disciplina FIFO (*First-In-First-Out*)



Sistema de fila de espera

- Um sistema de fila de espera é representado por:

$$A/B/c/d$$

A – o processo de chegada de clientes:

M – Markoviano, *D* – Determinístico, *G* – Genérico

B – o processo de atendimento de clientes:

M – Markoviano, *D* – Determinístico, *G* – Genérico

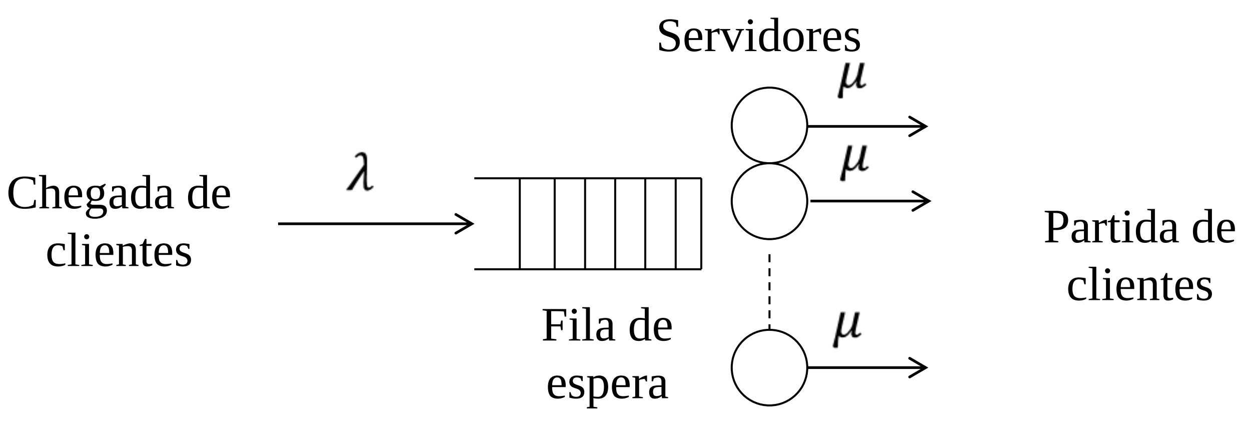
exponencial constante

c – o número de servidores

d – capacidade do sistema em nº de clientes:

número de servidores + capacidade da fila de espera

- Quando *d* é omisso, a fila de espera tem tamanho infinito.



Sistema *M*/*M*/1

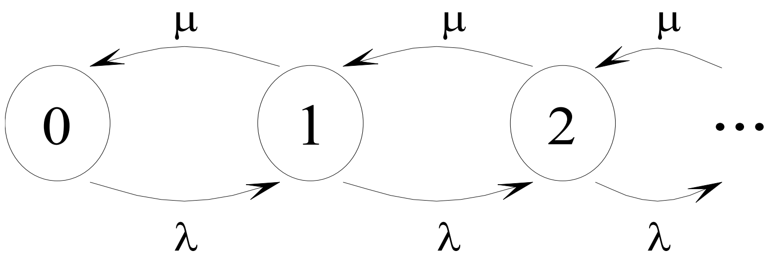
- Processo de nascimento e morte em que:

  (1) a chegada de clientes é um processo de Poisson com taxa λ

  (2) o tempo de atendimento de um servidor é exponencialmente distribuído com média $1/\mu$

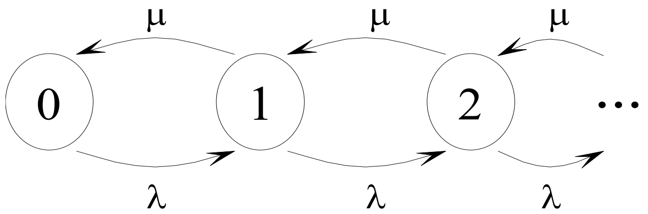
  (3) o sistema tem 1 servidor

  (4) o sistema acomoda um número infinito de clientes



- Uma ligação ponto-a-ponto com capacidade μ pacotes/s e uma fila de espera muito grande onde chegam pacotes a uma taxa de Poisson λ pacotes/s com comprimento exponencialmente distribuído de média $1/\mu$ é modelado por um sistema *M*/*M*/1

Sistema *M*/*M*/1



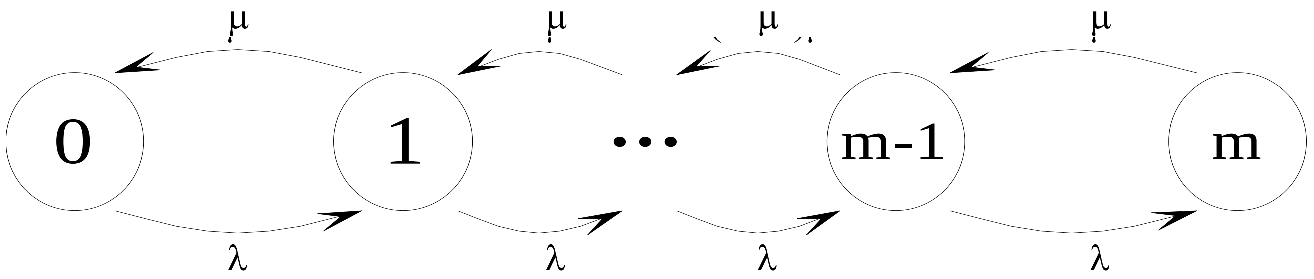
$$P_0 = \frac{1}{1+\sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}} = \frac{1}{1+\sum_{i=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^i}$$

$$P_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\cdot P_0 = \left(\frac{\lambda}{\mu}\right)^n \cdot P_0 = \frac{\left(\frac{\lambda}{\mu}\right)^n}{1+\sum_{i=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^i}$$

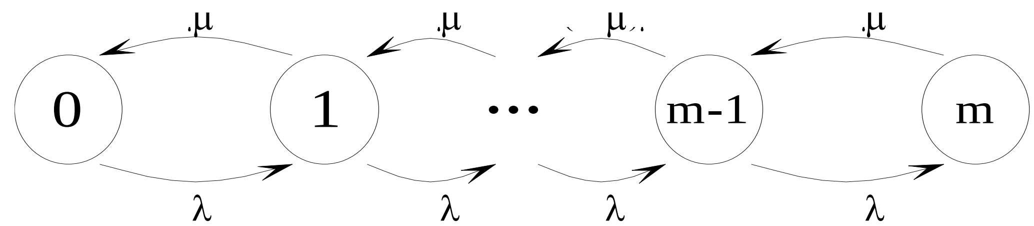
- Número médio de clientes no sistema: $L = \sum_{n=0}^{\infty} nP_n = \frac{\lambda}{\mu-\lambda}$
- Atraso médio no sistema: $W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda}$ pelo Teorema de Little
- Atraso médio na fila de espera: $W_Q = W - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}$
- Número médio de clientes na fila de espera: $L_Q = \lambda W_Q = \frac{\lambda^2}{\mu(\mu-\lambda)}$

Sistema *M*/*M*/1/*m*

- Processo de nascimento e morte em que:
  - (1) a chegada de clientes é um processo de Poisson com taxa λ
  - (2) o tempo de atendimento de um servidor é exponencialmente distribuído com média 1/μ
  - (3) o sistema tem 1 servidor
  - (4) o sistema acomoda no máximo *m* clientes (*i.e.*, a fila de espera tem capacidade para *m* – 1 clientes)



Sistema *M*/*M*/1/*m*



- Equações de balanço:

$$\lambda P_{n\text{-}1} = \mu P_n, \qquad n = 1, 2, \ldots, m$$

- Probabilidade de *n* clientes no sistema em estado estacionário:

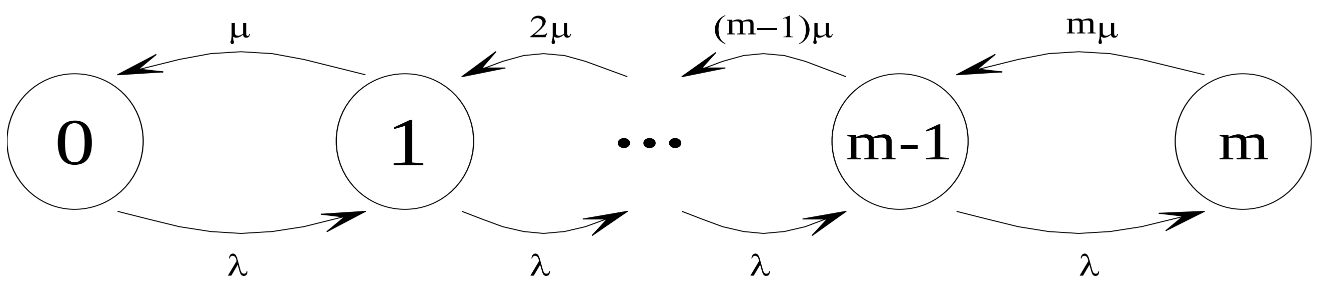
$$P_n = \frac{(\lambda/\mu)^n}{\sum_{i=0}^{m} (\lambda/\mu)^i} \qquad n = 0, 1, \ldots, m$$

- Pela propriedade PASTA, a probabilidade de um cliente chegar e encontrar o sistema cheio (*i.e.*, o servidor ocupado e a fila de espera cheia) é igual à probabilidade do estado *m*:

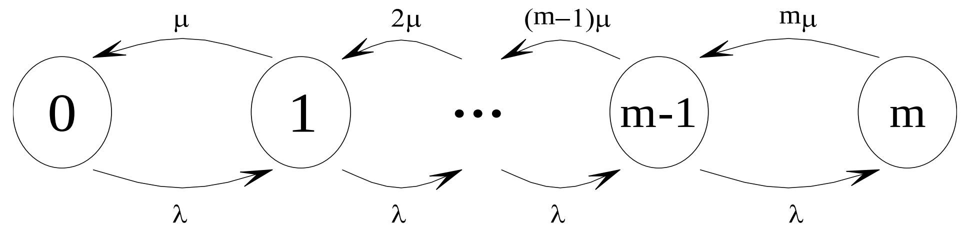
$$P_m = \frac{(\lambda/\mu)^m}{\sum_{i=0}^{m} (\lambda/\mu)^i}$$

Sistema *M/M/m/m*

- Processo de nascimento e morte em que:
  - (1) a chegada de clientes é um processo de Poisson com taxa λ
  - (2) o tempo de atendimento de um servidor é exponencialmente distribuído com média $1/\mu$
  - (3) o sistema tem *m* servidores
  - (4) o sistema acomoda no máximo *m* clientes (*i.e.*, não tem fila de espera)



Sistema *M*/*M*/*m*/*m*



- Equações de balanço:

$$\lambda P_{n\text{-}1} = n\mu P_n, \qquad n = 1, 2, ..., m$$

- Probabilidade de n clientes no sistema em estado estacionário:

$$P_n = \frac{(\lambda/\mu)^n / n!}{\sum_{i=0}^{m} (\lambda/\mu)^i / i!} \qquad n = 0, 1, \ldots, m$$

- Pela propriedade PASTA, a probabilidade de um cliente chegar e encontrar o sistema cheio é (fórmula de Erlang B):

$$P_m = \frac{(\lambda/\mu)^m / m!}{\sum_{i=0}^{m} (\lambda/\mu)^i / i!}$$

Sistema *M/G/1*

- Processo de nascimento e morte em que:
  - (1) a chegada de clientes é um processo de Poisson com taxa λ
  - (2) o tempo de atendimento S do servidor tem uma distribuição genérica e independente das chegadas dos clientes
  - (3) o sistema tem 1 servidor

so precisamos saber isto

  - (4) o sistema acomoda um número infinito de clientes
- Sendo conhecidos a média $E[S]$ e o segundo momento $E[S^2]$ do tempo de atendimento S, a fórmula de Pollaczek - Khintchine enuncia que o atraso médio de cada cliente na fila de espera é dado por:

$$W_Q = \frac{\lambda E[S^2]}{2(1-\lambda E[S])}$$

- O atraso médio de cada cliente no sistema é a soma do atraso médio na fila de espera mais o tempo médio de atendimento:

$$W = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} + E[S]$$

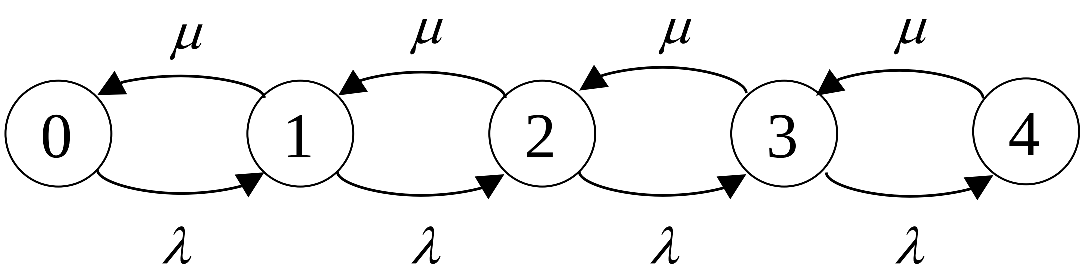
Exemplo 7

Considere um sistema de transmissão de pacotes com uma fila de espera seguida de uma linha de transmissão de 64 Kbps. O tráfego suportado é de Poisson com taxa de 15 pacotes/segundo. O comprimento dos pacotes é exponencialmente distribuído com média de 400 bytes. A fila de espera tem capacidade para 3 pacotes. Um pacote que chegue ao sistema é encaminhado pela linha de transmissão, se estiver livre, ou posto em fila de espera. Se a fila de espera se encontrar cheia, o pacote é perdido. Calcule:

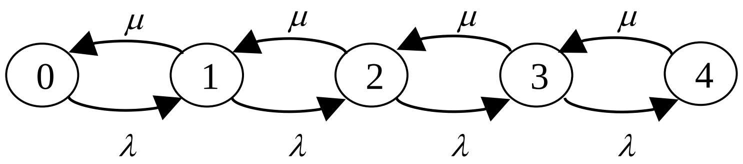
(a) A percentagem de pacotes perdidos.

(b) A percentagem de pacotes que não sofre atraso na fila de espera.

(c) A percentagem de utilização da linha de transmissão.



Exemplo 7 – Resolução (a)



Considere um sistema de transmissão de pacotes com uma fila de espera seguida de uma linha de transmissão de 64 Kbps. O tráfego suportado é de Poisson com taxa de 15 pacotes/segundo. O comprimento dos pacotes é exponencialmente distribuído com média de 400 bytes. A fila de espera tem capacidade para 3 pacotes. Um pacote que chegue ao sistema é encaminhado pela linha de transmissão, se estiver livre, ou posto em fila de espera. Se a fila de espera se encontrar cheia, o pacote é perdido. Calcule:

(a) A percentagem de pacotes perdidos.

$$P_n = \frac{(\lambda/\mu)^n}{\sum_{i=0}^{m} (\lambda/\mu)^i} \qquad n = 0, 1, \ldots, m$$

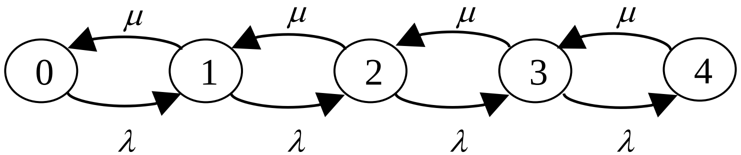
$$\mu = \frac{64000 \text{ bps}}{400 \times 8 \text{ bpp}} = 20 \text{ pps} \qquad \lambda = 15 \text{ pps}$$

$$P_4 = \frac{\left(\frac{\lambda}{\mu}\right)^4}{\left(\frac{\lambda}{\mu}\right)^0 + \left(\frac{\lambda}{\mu}\right)^1 + \left(\frac{\lambda}{\mu}\right)^2 + \left(\frac{\lambda}{\mu}\right)^3 + \left(\frac{\lambda}{\mu}\right)^4}$$

Pela propriedade PASTA

$$P_4 = \frac{\left(\frac{15}{20}\right)^4}{1 + \frac{15}{20} + \left(\frac{15}{20}\right)^2 + \left(\frac{15}{20}\right)^3 + \left(\frac{15}{20}\right)^4} = 0.104 = 10.4\%$$

Exemplo 7 – Resolução (b)



Considere um sistema de transmissão de pacotes com uma fila de espera seguida de uma linha de transmissão de 64 Kbps. O tráfego suportado é de Poisson com taxa de 15 pacotes/segundo. O comprimento dos pacotes é exponencialmente distribuído com média de 400 bytes. A fila de espera tem capacidade para 3 pacotes. Um pacote que chegue ao sistema é encaminhado pela linha de transmissão, se estiver livre, ou posto em fila de espera. Se a fila de espera se encontrar cheia, o pacote é perdido. Calcule:

(b) A percentagem de pacotes que não sofre atraso na fila de espera.

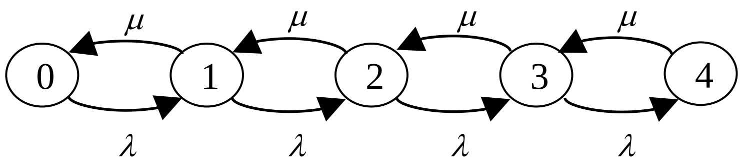
$$\mu = \frac{64000 \text{ bps}}{400 \times 8 \text{ bpp}} = 20 \text{ pps} \qquad \lambda = 15 \text{ pps} \qquad P_n = \frac{(\lambda/\mu)^n}{\sum_{i=0}^{m} (\lambda/\mu)^i} \qquad n = 0, 1, \ldots, m$$

$$P_0 = \frac{\left(\frac{\lambda}{\mu}\right)^0}{\left(\frac{\lambda}{\mu}\right)^0 + \left(\frac{\lambda}{\mu}\right)^1 + \left(\frac{\lambda}{\mu}\right)^2 + \left(\frac{\lambda}{\mu}\right)^3 + \left(\frac{\lambda}{\mu}\right)^4}$$

Pela propriedade PASTA

$$P_0 = \frac{1}{1 + \frac{15}{20} + \left(\frac{15}{20}\right)^2 + \left(\frac{15}{20}\right)^3 + \left(\frac{15}{20}\right)^4} = 0.328 = 32.8\%$$

Exemplo 7 – Resolução (c)



Considere um sistema de transmissão de pacotes com uma fila de espera seguida de uma linha de transmissão de 64 Kbps. O tráfego suportado é de Poisson com taxa de 15 pacotes/segundo. O comprimento dos pacotes é exponencialmente distribuído com média de 400 bytes. A fila de espera tem capacidade para 3 pacotes. Um pacote que chegue ao sistema é encaminhado pela linha de transmissão, se estiver livre, ou posto em fila de espera. Se a fila de espera se encontrar cheia, o pacote é perdido. Calcule:

(c) A percentagem de utilização da linha de transmissão.

$$U = 0 \times P_0 + 1 \times P_1 + 1 \times P_2 + 1 \times P_3 + 1 \times P_4$$

$$U = P_1 + P_2 + P_3 + P_4 = 1 - P_0$$

$$U = 1 - 0.328 = 0.672 = 67.2\%$$

Exemplo 8

soma de 2 processos de poisson é um M

Considere um sistema de transmissão ponto-a-ponto de 128 kbps que suporta dois fluxos de pacotes: no fluxo 1, os pacotes têm um tamanho constante de 128 Bytes e a chegada de pacotes é um processo de Poisson com taxa de 30 pacotes/segundo; no fluxo 2, os pacotes têm um tamanho constante de 512 Bytes e a chegada de pacotes é um processo de Poisson com taxa de 10 pacotes/segundo. Os pacotes dos dois fluxos partilham uma única fila de espera de tamanho muito grande.

(a) Indique justificando que tipo de fila de espera modela o desempenho deste sistema de transmissão.

(b) Calcule o atraso médio no sistema dos pacotes de cada fluxo.

Exemplo 8 – Resolução (a)

pode ser 512 com prob 10/40
ou
pode ser 128 com prob 30/40

Este sistema é modelado por uma fila de espera *M*/*G*/1:

- a soma de 2 processos de Poisson é um processo de Poisson e, assim, o processo de chegada de pacotes é um processo de Poisson ('*M*' em *M*/*G*/1) com taxa 30 + 10 = 40 pps;
- o tempo de transmissão dos pacotes do conjunto dos 2 fluxos é genérico ('*G*' em *M*/*G*/1) porque a distribuição dos tamanhos não segue uma distribuição comum: o tamanho é 128 Bytes com probabilidade 30/(30+10) = 0.75 = 75% ou 512 Bytes com probabilidade 10/(30+10) = 0.25 = 25%;
- o número de servidores é um ('1' em *M*/*G*/1) pois o canal de transmissão é usado para transmitir um pacote de cada vez;
- como o sistema tem uma fila de espera muito grande, a notação relativa à capacidade do sistema é omissa (considera-se que o sistema tem capacidade infinita).

Exemplo 8 – Resolução (b)

Atraso médio de cada pacote na fila de espera de um sistema *M*/*G*/1	$W_Q = \dfrac{\lambda E[S^2]}{2(1-\lambda E[S])}$

S_{128} = (128 × 8)/128000 = 8 × 10^{-3} seg

S_{512} = (512 × 8)/128000 = 32 × 10^{-3} seg

30/40 10/40

$E[S] = 0.75 \times S_{128} + 0.25 \times S_{512} =$
$= 0.75 \times 8 \times 10^{-3} + 0.25 \times 32 \times 10^{-3} = 14\text{e}{-3}$ seg

$E[S^2] = 0.75 \times (S_{128})^2 + 0.25 \times (S_{512})^2 =$
$= 0.75 \times (8 \times 10^{-3})^2 + 0.25 \times (32 \times 10^{-3})^2 = 3.04\text{e}{-4}\ \text{seg}^2$

$$W_Q = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} = \frac{40 \times 3.04\text{e}-4}{2(1-40 \times 14\text{e}-3)} = 0.0143 = 14.3 \text{ mseg}$$

$W_{128} = W_Q + S_{128} = 14.3 + 8 = 22.3$ mseg
$W_{512} = W_Q + S_{512} = 14.3 + 32 = 46.3$ mseg

Discrete Event Simulation

Modelação e Desempenho de Redes e Serviços

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

DETI-UA, 2023/2024

Discrete event simulation

A discrete event simulation models the operation of a system whose state changes with events that happen in discrete time instants:

- each event forces a change of either a system state value and/or of a variable value;
- between consecutive events, the system remains in the same state;
- thus, the simulation can directly jump in time from one event to the next event.

Elements of a discrete event simulator:

(1) State variables: describe the state of the system at any time instant

(2) Statistical counters: variables that store the appropriate statistical data related with the performance of the system

(3) Simulation clock: variable indicating the current simulated time instant (simulated time ≠ computation time)

(4) Events: types of occurrences that change either the system state and/or the statistical counters

(5) Event list: list of future events, their time instants and associated parameters

Besides these elements, additional supporting variables might be required.

Basic structure of a discrete event simulator

A discrete event simulator is mainly composed by the following steps:

(1) Initialization of the state variables, the statistical counters and the event list with the first event(s).

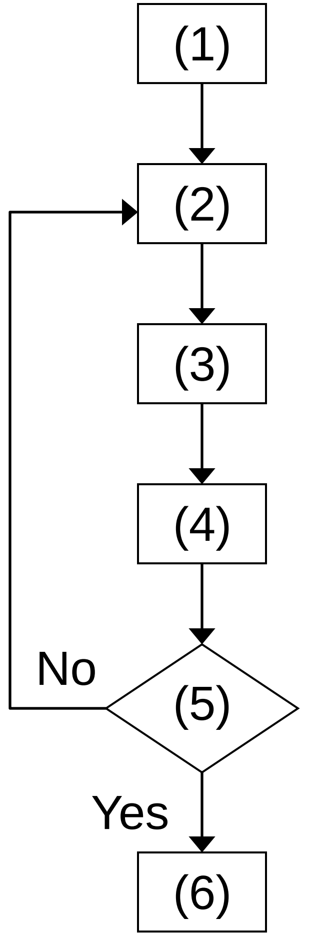
(2) Determination of the next event from the event list.

(3) Update of the simulation clock to the time instant of the next event and removal of the event from the event list.

(4) Execution of all actions associated to the event (generation of new events, update of state variables and/or statistical counters).

(5) Check if the simulation must end; if not, return to Step (2).

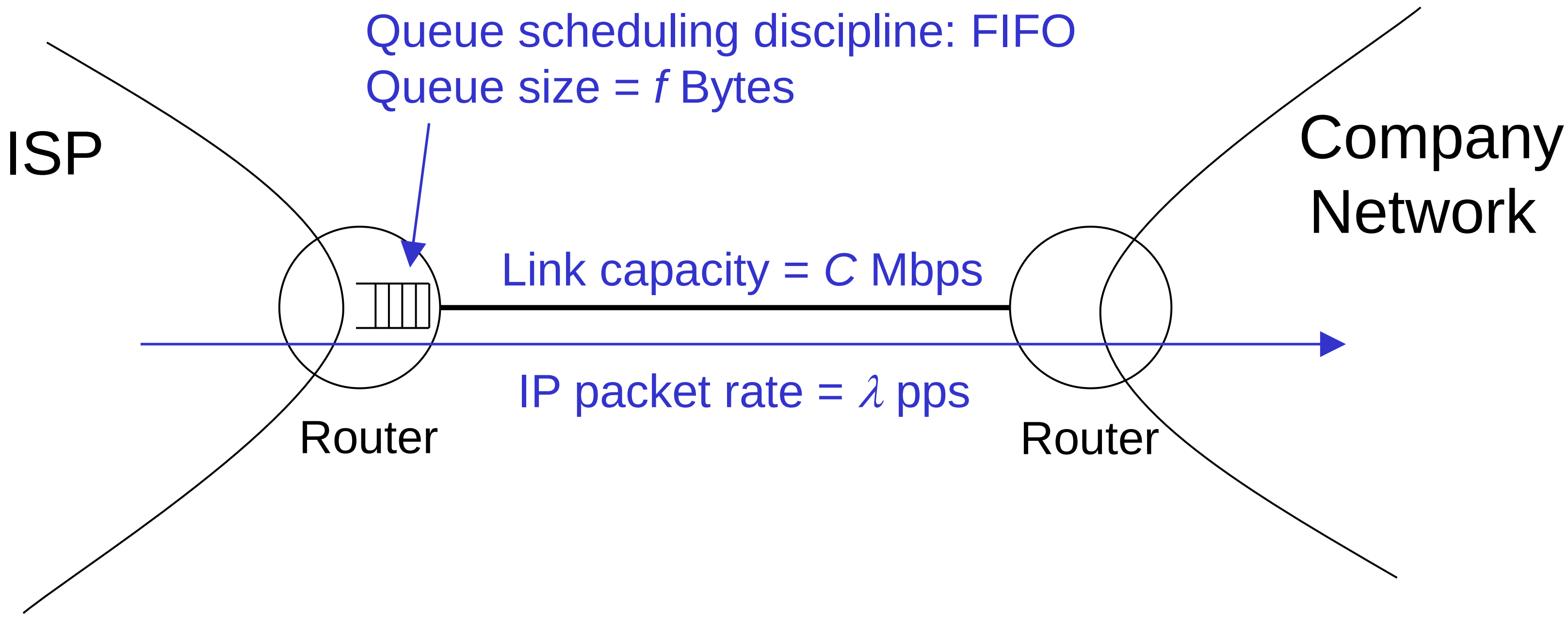
(6) Update of the statistical counters (if needed) and determination of the performance parameters.



Example: performance of a downstream link from an ISP to a client company

Consider the event driven simulation of a point-to-point IP link between a company router and its Internet Service Provider (ISP).

Let us consider the downstream direction, *i.e.*, from ISP to the company (usually, the direction with highest traffic load).



Example: performance of a downstream link from an ISP to a client company

Input parameters of simulation:

λ – packet rate, in packets per second (pps)

C – connection capacity, in Mbps

f – queue size, in Bytes

P – total number of transmitted packets of a simulation run

Stopping criteria of simulation:

Time instant when the link ends the transmission of the P^{th} packet

Performance parameters to be estimated by the simulation:

PL – Packet Loss (%)

APD – Average Packet Delay (milliseconds)

MPD – Maximum Packet Delay (milliseconds)

TT – Transmitted Throughput (Mbps)

Example: performance of a downstream link from an ISP to a client company

<u>Events</u>:

ARRIVAL – the arrival of a packet

DEPARTURE – the end of transmission of a packet

<u>State variables</u>:

STATE – binary variable indicating if the connection is free (i.e., not being used) or busy with the transmission of a packet

QUEUEOCCUPATION – occupation of the queue, in number of bytes, with the queued packets

QUEUE – matrix with (i) a number of rows equal to the number of queued packets and (ii) 2 columns where each column has the size and the arriving time instant of each packet in the queue

Example: performance of a downstream link from an ISP to a client company

Statistical Counters:

TOTALPACKETS – number of packets arrived to the system

LOSTPACKETS – number of packets dropped due to buffer overflow

TRANSMITTEDPACKETS – number of transmitted packets

TRANSMITTEDBYTES – sum of the bytes of the transmitted packets

DELAYS – sum of the delays of the transmitted packets

MAXDELAY – maximum delay among all transmitted packets

Performance parameters (at the end of the simulation):

PL = 100 × LOSTPACKETS / TOTALPACKETS

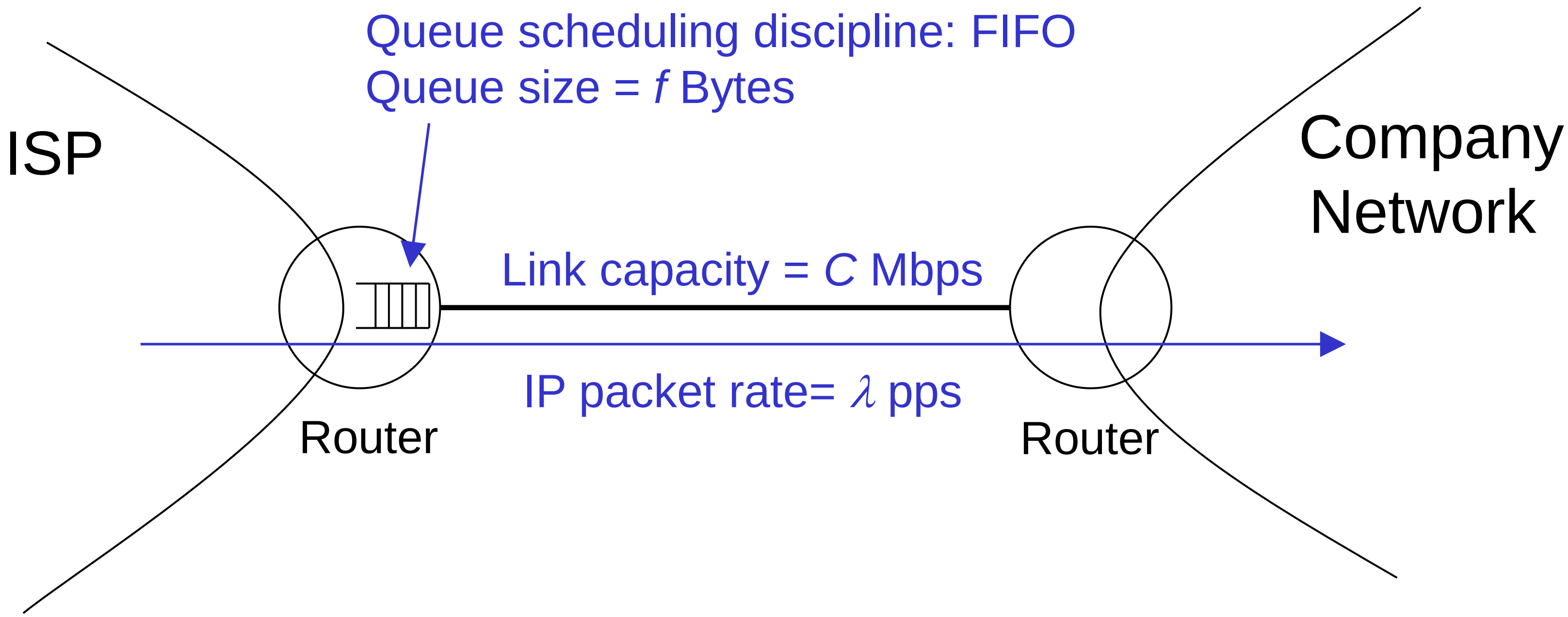
APD = 1000 × DELAYS / TRANSMITTEDPACKETS

MPD = 1000 × MAXDELAY

TT = 10^{-6} × TRANSMITTEDBYTES × 8 / total simulated time

Example: performance of a downstream link from an ISP to a client company

Illustration of a simulation run in MATLAB



Generation of random numbers with a uniform distribution between 0 and 1

A Linear Congruential Generator (LCG) is an algorithm that yields a sequence of randomized numbers calculated with a linear equation.

The method represents one of the oldest and best-known pseudorandom number generator algorithms.

Generation method:

(1) Generate integer values Z_1, Z_2, … with the following recursive expression:

$$Z_i = (aZ_{i-1} + c)(\mathrm{mod}\, m)$$

where m, a, c and Z_0 are non-negative integer parameters;

(2) Compute $U_i = Z_i / m$.

The values U_i seem to be real values uniformly distributed on the interval [0,1]

Example

Example: $Z_i = (5Z_{i-1} + 3)(\text{mod } 16)$

$Z_0 = 7$

i	Z_i	U_i	***i***	Z_i	U_i
0	7	----	10	9	0.563
1	6	0.375	11	0	0.000
2	1	0.063	12	3	0.188
3	8	0.500	13	2	0.125
4	11	0.688	14	13	0.813
5	10	0.625	15	4	0.250
6	5	0.313	16	7	0.438
7	12	0.750	17	6	0.375
8	15	0.938	18	1	0.063
9	14	0.875	19	8	0.500

- In this example, m = 16 and the algorithm repeats the generated numbers after 16 iterations (we say the generator has a period of 16).
- The random generator (function *rand*) of MATLAB has a period of 2^{31}-1.

Generation of random numbers with other distributions

Discrete variables:

Consider a random variable that can have the values X_1, X_2, …, X_n.

Consider the probability of value X_i as $P(X = X_i) = f_i$.

Method:

- Split the interval [0,1] in n intervals proportional to f_i, $i = 1 \ldots n$
- Generate a uniformly distributed random value U in [0,1]
- Return X_i if U falls into the i-th interval

For example, the Bernoulli variable X with $p(0) = 1/4$ and $p(1) = 3/4$ can be randomly generated as:

(1) Generate U~U(0,1)

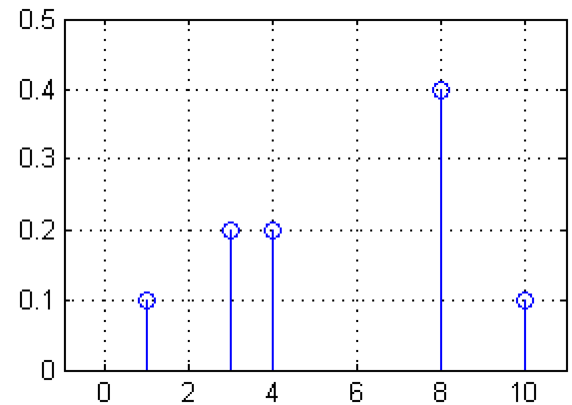
(2) If $U \leq 1/4$, return $X = 0$; otherwise, return $X = 1$

Generation of random numbers with other distributions

<u>Discrete variables</u> (MATLAB example)

```
x= [1 3 4 8 10];
f= [0.1 0.2 0.2 0.4 0.1];

figure(1)
stem(x,f)
axis([-1 11 0 0.5])
grid on
```



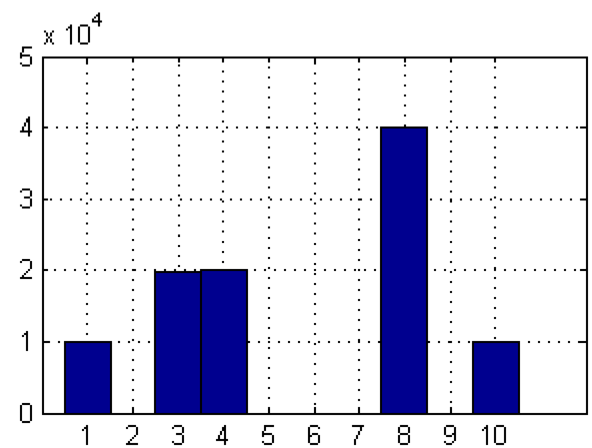

```
f_cum= [0 cumsum(f)]
```

```
f_cum =

    0.0   0.1   0.3   0.5   0.9   1.0
```

```
a= zeros(1,100000);
for it= 1:100000
    a(it)= x(sum(rand()>f_cum));
end

figure(2)
hist(a,1:10)
grid on
```



Generation of random numbers with other distributions

Generate a random packet size between 64 and 1518 bytes with the probabilities: 19% for 64 bytes, 23% for 110 bytes, 17% for 1518 bytes and an equal probability for all other values (i.e., from 65 to 109 and from 111 to 1517).

Custom MATLAB function:

```
function out= GeneratePacketSize()
    aux= rand();
    aux2= [65:109 111:1517];
    if aux <= 0.19
        out= 64;
    elseif aux <= 0.19 + 0.23
        out= 110;
    elseif aux <= 0.19 + 0.23 + 0.17
        out= 1518;
    else
        out = aux2(randi(length(aux2)));
    end
end
```

Generation of random numbers with other distributions

Continuous variables:

The most popular methods are based on the inverse of the cumulative distribution function (cdf).

Consider $F(X)$ as the cdf of a continuous random variable and $F^{-1}(U)$ as its inverse function.

Method:

(1) Generate U~U(0,1)

(2) Return $X = F^{-1}(U)$

For example, an exponential distributed random variable with average $1/\lambda$:

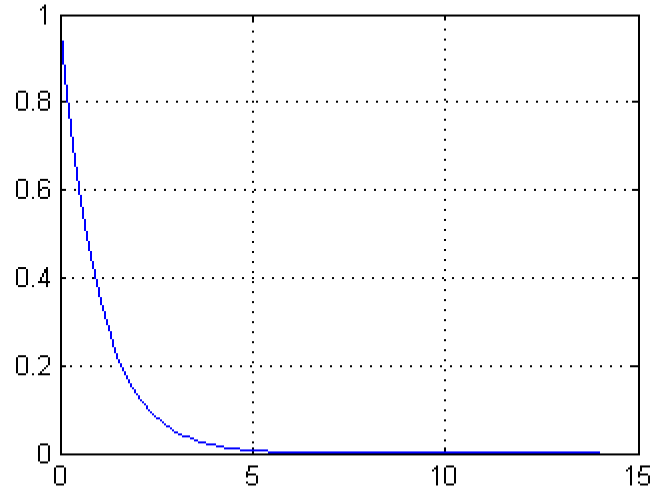
$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad F^{-1}(U) = -\frac{1}{\lambda}\ln(U)$$

In MATLAB, use function *exprnd.*

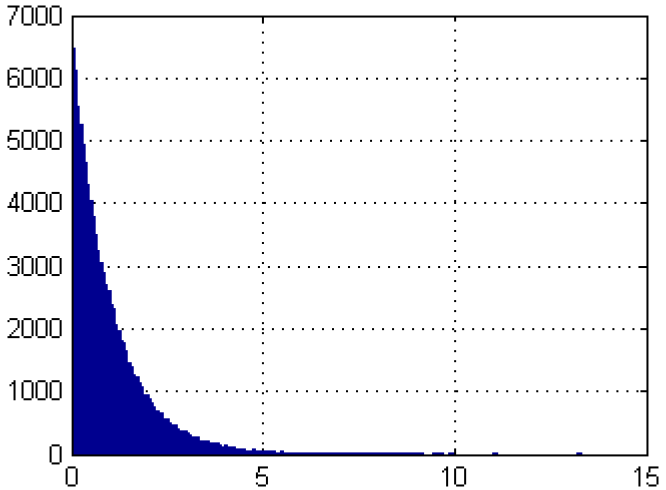
Generation of random numbers with other distributions

Exponential variable (MATLAB example)

```
lambda= 1
x= 0:0.1:14;
f=exppdf(x,1/lambda)
figure(1)
plot(x,f)
grid on
```




```
a=exprnd(1/lambda,1,100000);
figure(2)
hist(a,200)
grid on
```



Analysis of the results of a simulation

Consider X_1, X_2, …, X_n as the observations of independent and identically distributed (IID) random variables with average μ and finite variance σ^2 (for example, the results of different simulations of a given system).

The <u>sample mean</u> defined by

$$\overline{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

is an estimator for average μ.

The <u>sample variance</u> defined by

$$S^2(n) = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}(n)\right)^2}{n-1}$$

is an estimator for variance σ^2.

The analysis of the results of a simulation is, usually, based on the <u>Central Limit Theorem</u>.

Analysis of the results of a simulation

Consider Z_n as a random variable given by: $$Z_n = \frac{\overline{X}(n) - \mu}{\sqrt{\sigma^2 / n}}$$

Consider $F_n(z)$ the cumulative distribution function of Z_n for a sample of size n.

The ***Central Limit Theorem*** states that

$$\lim_{n \to +\infty} F_n(z) = \Phi(z)$$

where $\Phi(z)$ is the cumulative distribution function of a standard Gaussian random variable (i.e., a Gaussian distribution with mean 0 and variance 1).

Given that $\lim_{n \to +\infty} S^2(n) = \sigma^2$ than, the random variable $$\frac{\overline{X}(n) - \mu}{\sqrt{S^2(n)/n}}$$

has approximately a standard Gaussian distribution.

Analysis of the results of a simulation

For a sufficiently high value of n,

$$P\left(-z_{1-\alpha/2} \leq \frac{\overline{X}(n)-\mu}{\sqrt{S^2(n)/n}} \leq z_{1-\alpha/2} \right) =$$

$$P\left(\overline{X}(n)-z_{1-\alpha/2}\sqrt{S^2(n)/n} \leq \mu \leq \overline{X}(n)+z_{1-\alpha/2}\sqrt{S^2(n)/n} \right) \approx 1-\alpha$$

where $z_{1-\alpha/2}$ is the critical value of the standard Gaussian distribution ($z_{1-\alpha/2}$ is the value z such that $P(x \leq z) = 1 - \alpha/2$ where x is a random variable with a standard Gaussian distribution).

Therefore, the approximate confidence interval of 100(1-α)% for the average μ is given as

$$\overline{X}(n) \underbrace{\pm z_{1-\alpha/2}\sqrt{S^2(n)/n}}_{\text{INTERVALO DE CONFIANÇA}}$$

Analysis of the results of a simulation

The approximate confidence interval of 100(1-α)% for the average μ is

$$\overline{X}(n) \pm z_{1-\alpha/2}\sqrt{S^2(n)/n}$$

In MATLAB:

```
N = 20;               %number of simulations
per1= zeros(1,N); %vector with N simulation values
per2= zeros(1,N); %vector with N simulation values
for it= 1:N
     [per1(it),per2(it)]= simulator();
end

alfa= 0.1; %90% confidence interval%
media = mean(per1);
term = norminv(1-alfa/2)*sqrt(var(per1)/N);
fprintf('per1 = %.2e +- %.2e\n',media,term)
media = mean(per2);
term = norminv(1-alfa/2)*sqrt(var(per2)/N);
fprintf('per2 = %.2e +- %.2e\n',media,term)
```

Analysis of the results of a simulation

The central limit theorem requires variables X_1, X_2, …, X_n to be independent and identically distributed (IID).

- One way to guarantee the independence between the different values is to run different simulations guaranteeing that the random values are different on the different runs.
- This is done by using different seeds on the random generators.

In general, the stochastic processes have initial transient states (which are dependent on the initial conditions) before reaching the stationary state.

- In order to guarantee that the performance estimations are correct, the simulation must first warm-up to let the transient states vanish.
- If the simulated time is much higher than the warm-up time, statistical counters can be initialized at the beginning of the simulation.
- Otherwise, the statistical counters must be initialized only after the warm-up time (this time must be estimated though).