REDES DE TELECOMUNICAÇÕES

Tráfego

Engª de Sistemas e Informática

UALG/FCT/ADEEC 2004/2005

Sumário

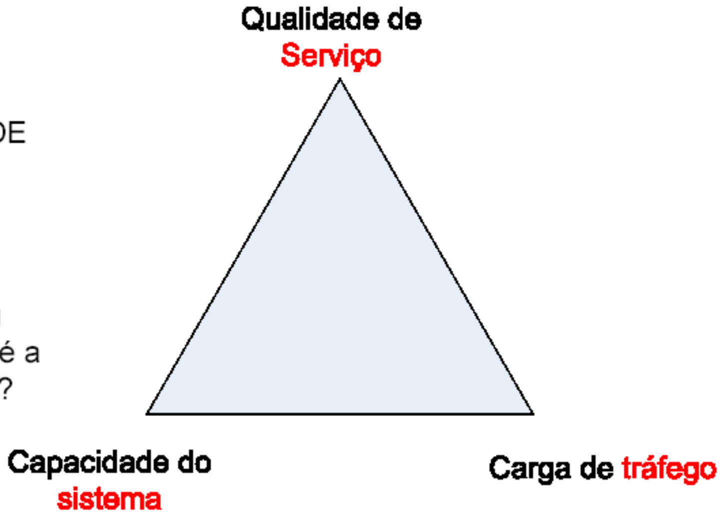
- **Objectivos da teoria do tráfego**
- **Modelos de Teletráfego**

Objectivos da teoria do tráfego

Sendo dados o SISTEMA e o TRÁFEGO DE ENTRADA qual é a **qualidade de serviço** sentida pelo utilizador ?

Sendo dados o TRÁFEGO DE ENTRADA e um grau de QUALIDADE DE SERVIÇO qual deve ser a **capacidade do sistema** ?

Sendo dados o SISTEMA e um grau de QUALIDADE DE SERVIÇO qual é a carga máxima de **tráfego** permitida ?



Objectivos da teoria do tráfego

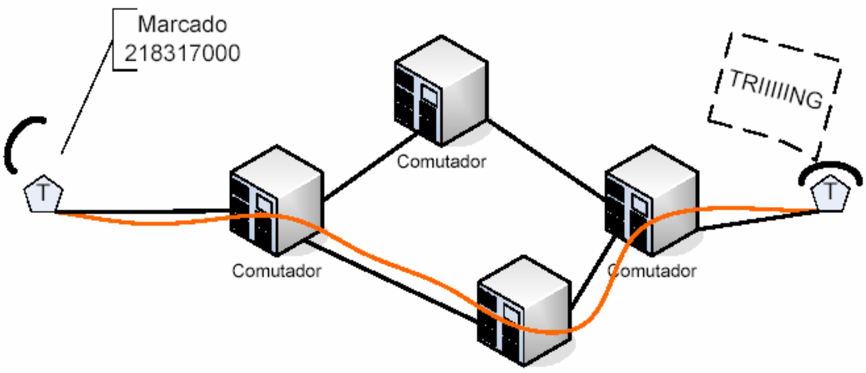
Exemplo 1

Chamada telefónica

– Tráfego = Chamadas telefónicas

– Sistema = Rede Telefónica

– Qualidade de serviço = Probabilidade do telefone de destino tocar.



Intensidade de tráfego

A utilização de um nó de comutação de circuitos (por exemplo, uma central telefónica) ou de um grupo de circuitos de transmissão é determinada por dois factores:

– chegada de chamadas;

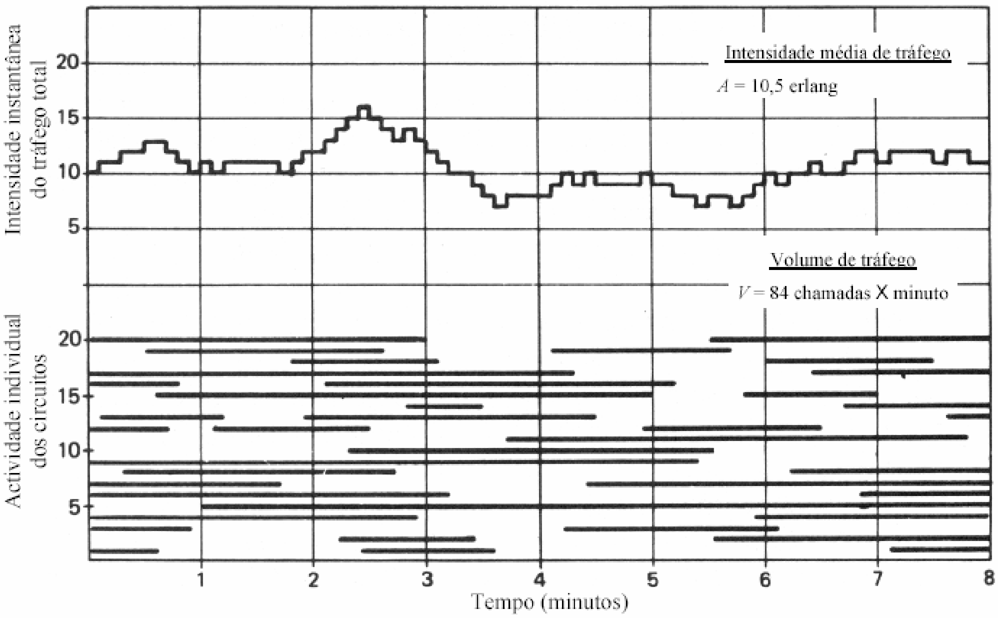
– duração de chamadas.

O tráfego toma em consideração estas componentes, podendo quantificar-se através da intensidade em unidades de erlang (segundo A.K. Erlang, pioneiro dinamarquês da teoria do teletráfego): se um sistema suporta A chamadas, diz-se que transporta A erlangs de tráfego (note-se que esta unidade é adimensional).

De um modo geral, a intensidade de tráfego num intervalo de tempo pode exprimir-se através da razão entre o volume de tráfego V - somatório das durações d_i de todas as chamadas n ocorridas no intervalo - e o período de tempo T:

$$\boxed{A = \frac{V}{T} = \frac{\sum_{i=1}^{n} d_i}{T}}$$

Caracterização do tráfego



Actividade de tráfego num conjunto de circuitos

Por outro lado, a taxa média de chegada de chamadas λ e a duração média das chamadas d_m são dadas por:

$$\boxed{\lambda = \frac{n}{T}} \qquad \boxed{d_m = \frac{\sum_{i=1}^{n} d_i}{n}}$$

Logo, a intensidade de tráfego virá:

$$\boxed{A = \lambda\, d_m}$$

A intensidade de tráfego é apenas uma medida da utilização média durante um período de tempo e não reflecte a relação entre chegada e duração de chamadas: muitas chamadas de curta duração podem produzir a mesma intensidade de tráfego que poucas chamadas longas.

Nas redes de comutação de circuitos, é geralmente suficiente caracterizar o tráfego apenas em termos de intensidade de tráfego, podendo por vezes ser necessário considerar os padrões de chegada de chamadas ou as distribuições de durações.

Variação do tráfego

As redes de comutação de circuitos, como a rede telefónica pública, são tipicamente dimensionadas em termos da actividade durante a **hora mais carregada** do dia, procurando-se um compromisso entre dois extremos:

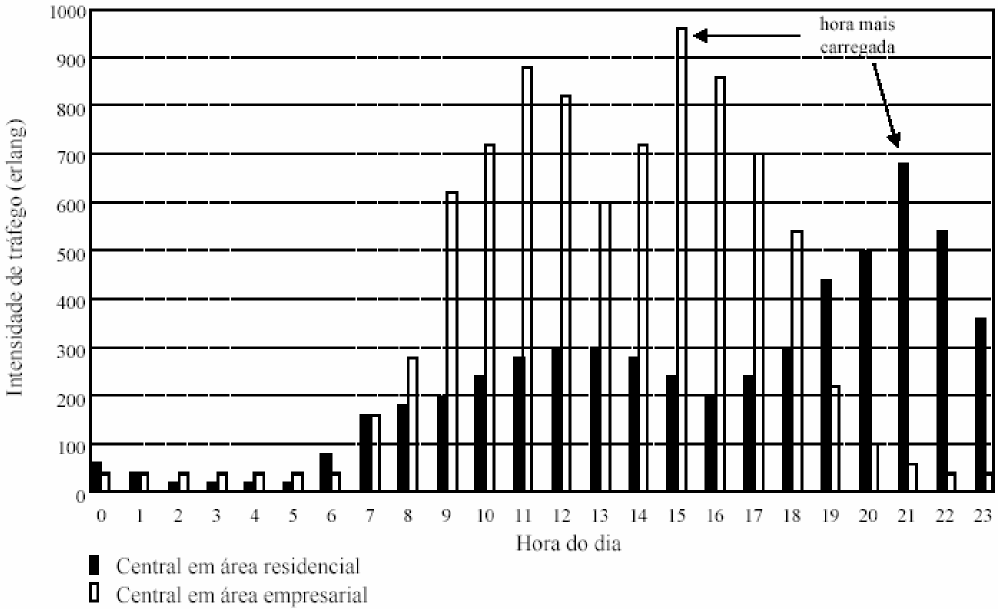
- projectar para a utilização média total, que inclui tempos nocturnos virtualmente sem tráfego;
- projectar para picos de curta duração resultantes de acontecimentos esporádicos (fim de programas de TV, concursos por telefone, etc.).

É habitual considerar dois tipos de utilizadores no que respeita ao tráfego telefónico:

- **residenciais**, com utilização por linha de 0,05 a 0,1 erlang na hora mais carregada, e tempos médios de duração de chamadas de 3 a 4 minutos; as centrais de zonas residenciais apresentam maiores cargas ao fim do dia/noite;
- **empresariais**, com utilizações superiores em horas correspondentes ao fim da manhã/meio da tarde.

O tráfego de acesso à Internet, caracterizado por chamadas de mais longa duração, introduz uma carga adicional significativa que altera os padrões tradicionais.

Caracterização do tráfego



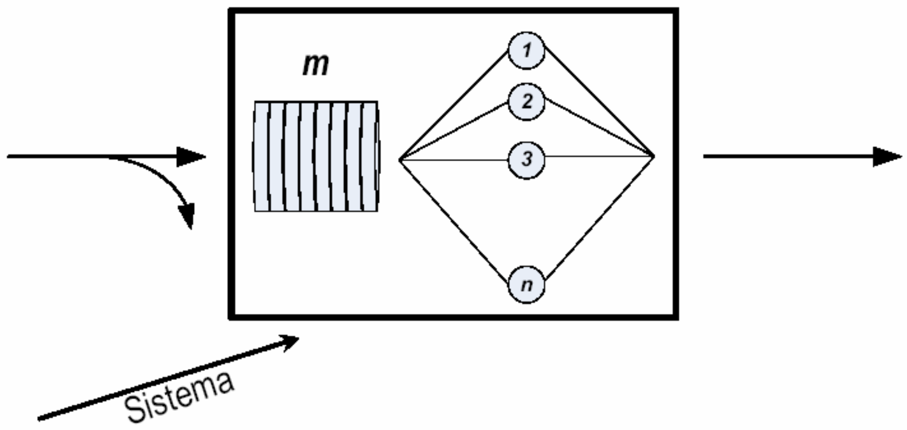
Sistemas de perda e sistemas de espera

Há dois grandes tipos de sistemas usados em redes de comutação de circuitos:

- **sistema de perda**, em que as chamadas que excedem em qualquer instante a capacidade máxima do sistema são rejeitadas, podendo ou não regressar em novas tentativas;
- **sistemas de espera**, em que as chamadas em excesso são colocadas em filas de espera até estarem disponíveis os recursos necessários para as suportar (excepcionalmente, se a fila encher, o sistema passa a comportar-se como na situação de perda).

Nos sistemas de perda, o tráfego transportado é sempre menor ou igual ao tráfego oferecido ao sistema pelas fontes, enquanto nos sistemas de espera estas duas quantidades são iguais, desde que a média a longo prazo do tráfego oferecido seja inferior à capacidade máxima do sistema e as filas de espera tenham profundidade suficiente para absorver os picos.

- Chamadas chegam ao sistema a um ritmo λ (chamadas por unidade de tempo)
  - $1/\lambda$ = tempo médio entre chamadas
- As chamadas são distribuídas por n servidores em paralelo
- Um servidor presta o serviço a um ritmo μ (chamadas por unidade de tempo)
  - $1/\mu$ = tempo médio de serviço a uma chamada
- Há m buffers de espera
- Assume-se que chamadas bloqueadas (chegadas com o sistema cheio) são perdidas.



- Sem filas de espera (m = 0)
  - Quando a chamada chega, se o sistema estiver cheio (os n servidores ocupados) ela não é servida e perde-se
  - Há algumas chamadas perdidas
- Na perspectiva da chamada de entrada
  - Qual é a probabilidade do sistema estar cheio quando a chamada chega ?
- Na perspectiva do sistema
  - Qual é a taxa de utilização dos servidores ?

- Fila de espera de dimensão infinita ($m = \infty$)
  - Uma chamada ocupa um lugar na fila de espera se todos os servidores estiverem ocupados
  - Não há chamadas perdidas mas algumas terão de esperar em fila de espera

- Na perspectiva da chamada de entrada
  - Qual é a probabilidade da chamada ficar em fila de espera ?

- Na perspectiva do sistema
  - Qual é a taxa de utilização dos servidores ?

- Fila de espera de dimensão finita ($0< m <\infty$)

- À chegada de uma chamada se os *n* servidores estiverem ocupados mas houver vagas na fila de espera, a chamada ocupa um dos lugares vagos na lista de espera.

- À chegada de uma chamada se estiverem ocupados os *n* servidores e as *m* posições na fila de espera então a chamada perde-se

- Algumas chamadas são perdidas outras ficam em fila de espera antes de serem servidas.

- Num sistema com perdas algumas chamadas perdem-se por:
  - Encontrarem todos os n canais ocupados
- Há 2 tipos de bloqueio:
  - Bloqueio de chamada B_c = Probabilidade de uma chamada de entrada encontrar todos o n canais ocupados = fracção de chamadas que são perdidas
  - Bloqueio no tempo B_t = Probabilidade de todos os n canais estarem ocupados num intervalo de tempo t = Fracção do tempo em que todos os n canais estão ocupados
- Se a chegada das chamadas fôr de acordo com o processo de Poisson ➔ $B_c = B_t$
- Bloqueio da chamada é a medida mais próxima do QoS percebida pelo assinante

Num sistema de perdas cada chamada é transportada ou perdida

$\lambda_{oferecido}$ = taxa de chegadas de todas as tentativas de chamadas

$\lambda_{transportado}$ = taxa de chegadas de chamadas transportadas

$\lambda_{perdido}$ = taxa de chegadas de chamadas perdidas

$$\lambda_{oferecido} = \lambda_{transportado} + \lambda_{perdido} = \lambda$$

$$\lambda_{transportado} = \lambda\,(1-\boldsymbol{B}_c) \qquad B_c = \frac{\lambda_{\text{perdida}}}{\lambda}$$

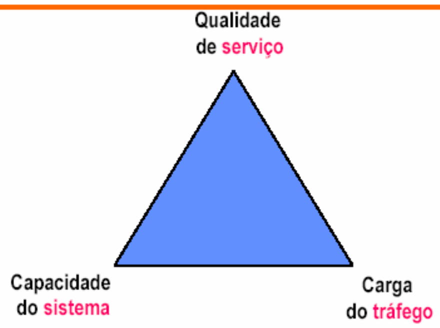
$$\rho_{oferecido} = \rho_{transportado} + \rho_{perdido} = \rho$$

$$\rho_{transportado} = \rho\,(1-\boldsymbol{B}_c) \qquad B_c = \frac{\rho_{\text{perdido}}}{\rho}$$

$\rho_{transportado}$= *Numero médio de canais ocupados no feixe*

Formula de Erlang B

$$B_c = \frac{\dfrac{\rho^N}{N!}}{\displaystyle\sum_{i=0}^{N} \frac{\rho^i}{i!}}$$



ρ = Tráfego oferecido (intensidade de trafego)

= $\lambda \times h$

λ = média de novas chamadas no tempo h

h = duração média da chamada

N = Numero de canais para escoar o trafego

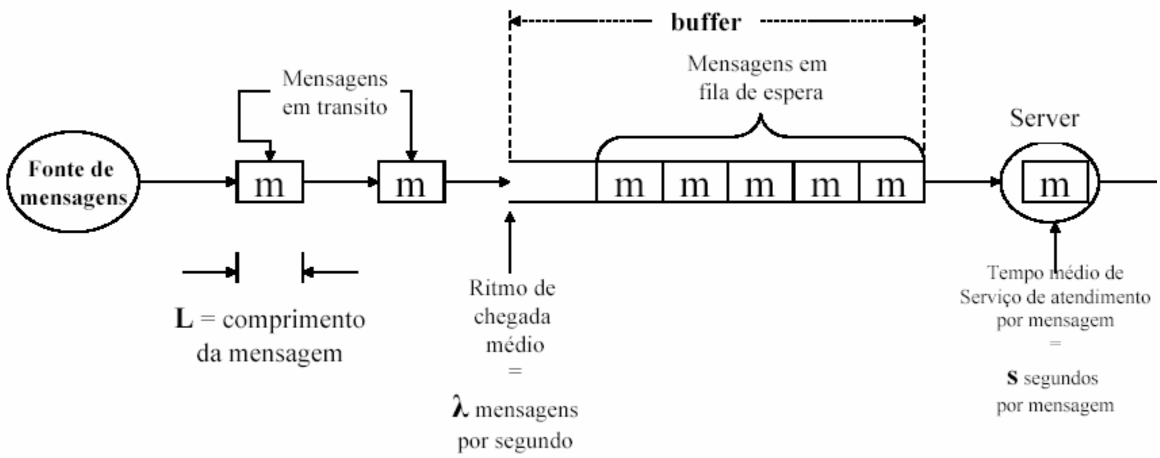
B_c = Probabilidade de bloqueio

Trafego telefónico Ex. da tabela de Erlang B

Valores de A função de Bc e N

B	.005	.01	.02	.03	.05	.10
N						
1	.005	.011	.021	.031	.053	.113
2	.106	.153	.224	.282	.382	.602
3	.349	.456	.603	.716	.900	1.275
4	.702	.870	1.093	1.259	1.525	2.052
5	1.132	1.361	1.658	1.876	2.219	2.896
6	1.662	1.909	2.276	2.543	2.961	3.751
7	2.158	2.501	2.936	3.250	3.738	4.678
8	2.703	3.128	3.637	3.987	4.543	5.613
9	3.333	3.783	4.345	4.748	5.371	6.560
10	3.961	4.462	5.084	5.530	6.216	7.502
11	4.611	5.160	5.842	6.328	7.077	8.504
12	5.279	5.876	6.615	7.141	7.950	9.472

Modelo simples de uma fila de espera



(Comutação de pacotes/mensagens)

O trafego de dados num link pode ser modulado como um sistema de "filas de espera puro" (1 servidor e uma memória de dimensão $m = \infty$)

- pacote
  - λ = taxa de chegada de pacotes
  - L = Comprimento médio dos pacotes (unidade de dados)

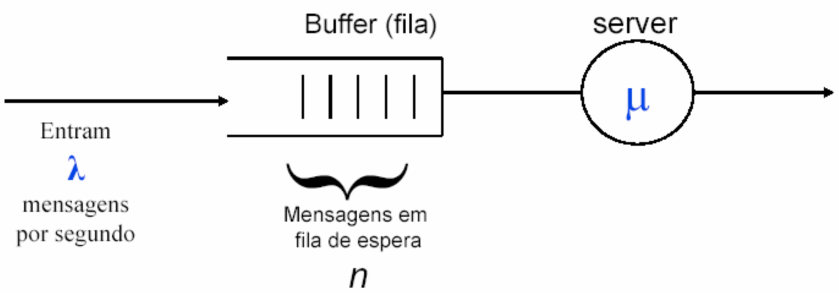
- Servidor (link), buffer
  - C = débito no link (dados por unidade de tempo)

Tempo de serviço = tempo de transmissão do pacote

- $\frac{1}{\mu} = \frac{L}{C}$ tempo de transmissão médio dos pacotes

Tempo de espera

Um servidor e uma fila de espera



T (tempo de espera na fila) é no geral aleatório pois o estado n do buffer varia aleatoriamente com o tempo

λ e μ são variáveis aleatórias – (distribuição de probabilidade).

Modelo simples de uma fila de espera

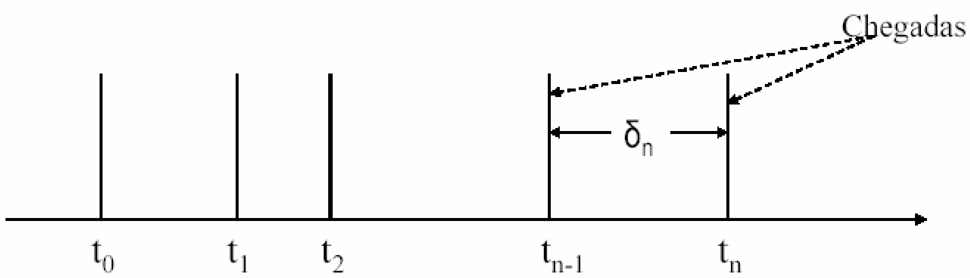
Tempo total em espera de uma mensagem

T = tempo de espera em fila + tempo de serviço

$$T = \frac{nL}{C} + \frac{L}{C} = \frac{n+1}{\mu}$$

C =	capacidade de processamento do sistema em unidades de dados por segundo;
L =	comprimento de uma mensagem em unidade de dados
n =	numero de mensagens em fila de espera (buffer) à sua frente
$\mu = C/L$	mensagens por segundo (service rate)

Processo de chegada das mensagens



- $\delta_n = t_n - t_{n-1}$ - tempo entre chegadas n
- δ_n é uma sequência de variáveis aleatórias independentes e identicamente distribuídas.
- a(t)=densidade de probabilidade de δ

Processo de chegada de mensagens

Formula de Poisson

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

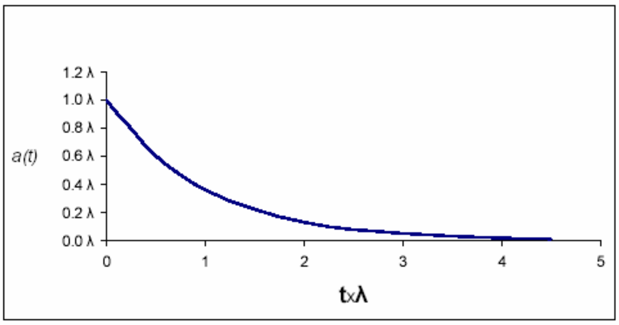
λ = ritmo médio de chegada de mensagens

n = 0, 1, 2, 3, ...

$P_n(t)$ – Probabilidade de no intervalo t chegarem n mensagens

Processo de chegada das mensagens

$$a(t)dt = \lambda e^{-\lambda t} dt$$



O processo de chegadas com distribuição de Poisson gera uma probabilidade de tempo entre chegadas de densidade exponencial.

Notação Kendall

A/B/c/K/N/Z

A – Distribuição dos tempos entre chegadas
B – A característica de serviço de atendimento
c – Numero de servidores
K – A capacidade do sistema (dimensão do buffer)
N – Numero potencial de customers na população fonte
Z – A disciplina de fila de espera

Notação Kendall

- Símbolos para A e B

G - Distribuição de tempo de serviço de atendimento geral (não especificada)

M - Distribuição de tempo de serviço de atendimento ou de tempo entre chegadas exponencial

D - Distribuição de tempo de serviço de atendimento ou de tempo entre chegadas determinístico/constante

Notação Kendall - exemplo

- Uma fila de espera M/G/1
  - Chegadas com distribuição de Poisson
  - Tempo de serviço de atendimento com distribuição geral (i.e., poucas especificações da distribuição de tempo de serviço de atendimento)
  - 1 servidor

- Uma fila de espera M/D/1
  - Tempo de serviço de atendimento com distribuição constante

Modelo de fila de espera M/M/1

- M - As mensagens chegam com uma distribuição de probabilidade de Poisson
- M - As mensagens entram numa fila de espera de dimensão infinita com um tempo de serviço de distribuição exponencial
- 1 - As mensagens são encaminhadas por um único server na base de FCFS

Os padrões de chegada de mensagens/pacotes em muitas aplicações (redes de computadores) seguem uma distribuição de probabilidade de Poisson.