Tráfego de Serviços Orientados a Conexões

- Caracterização do tráfego
- Análise de sistemas de estados
- Análise de tráfego em sistemas de perda
- Análise de tráfego em sistemas de atraso
- Bloqueio em sistemas de andares múltiplos



Intensidade de tráfego

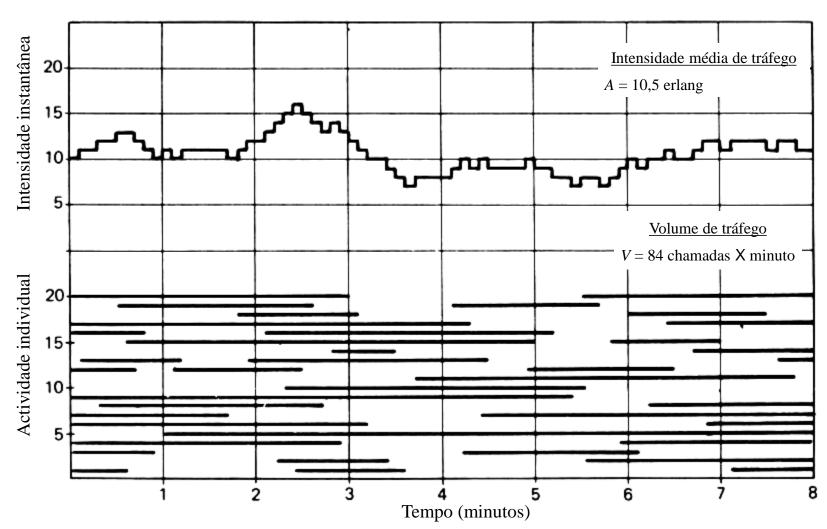
A utilização de um nó ou de uma ligação de uma rede que suporta um serviço orientado a conexões é determinada por **dois fatores**:

- chegada de chamadas (chegada de pacotes);
- duração de chamadas (tempo de transmissão de pacote).

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 exprimirse 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:

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



Atividade de tráfego de um serviço orientado a conexões num nó ou numa ligação

Por outro lado, a taxa média de chegada de chamadas (pacotes) λ e a duração média das chamadas (tempo de transmissão médio dos pacotes) d_m são dadas por:

$$\lambda = \frac{n}{T}$$

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

Logo, a intensidade de tráfego virá:

$$A = \lambda d_m$$

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

Nos serviços orientados a conexões, 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.

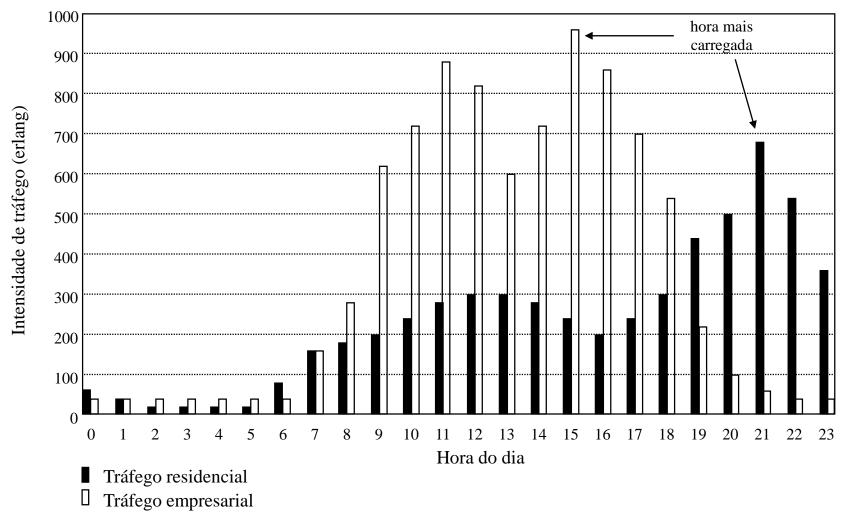
Variação do tráfego – exemplo do tráfego de voz

O dimensionamento do tráfego de serviços de voz é geralmente efetuado em termos da atividade durante a **hora mais carregada** do dia, procurando-se um compromisso entre dois extremos:

- projetar para a utilização média total, que inclui tempos noturnos virtualmente sem tráfego;
- projetar 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 acesso de 0,05 a 0,1 erlang na hora mais carregada, e tempos médios de duração de chamadas de 3 a 4 minutos; a hora mais carregada ocorre ao fim do dia/noite;
- empresariais, com utilizações superiores em horas correspondentes ao fim da manhã/meio da tarde.



Intensidade de tráfego de voz em função da hora do dia

Sistemas de perda e sistemas de espera

Há dois grandes tipos de sistemas usados nos serviços orientados a conexões:

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

Sistemas de perda – tráfego transportado é sempre menor ou igual ao tráfego oferecido ao sistema pelas fontes.

Sistemas de espera – tráfego transportado é igual ao tráfego oferecido, 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.

Os **sistemas telefónicos clássicos são sistemas de perda**. Contudo, são cada vez mais comuns sistemas de espera, sendo exemplos significativos os centros de atendimento (*call centres*) e certos tipos de redes de comunicação móvel.

Chegada de chamadas

Processo de Poisson

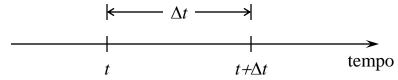
O processo de Poisson é o mais usado para modelizar o tráfego em sistemas de comunicações (sistemas telefónicos e redes de dados).

Assume o princípio básico de **independência de chegada de chamadas (pacotes)** entre fontes, podendo formular-se através da **probabilidade de chegada num intervalo infinitesimal** $[t, t+\Delta t]$, introduzindo uma constante de proporcionalidade λ :

P (uma chegada no intervalo $[t, t+\Delta t]$) = $\lambda \Delta t$

P (nenhuma chegada no intervalo $[t, t+\Delta t]$) = $1 - \lambda \Delta t$

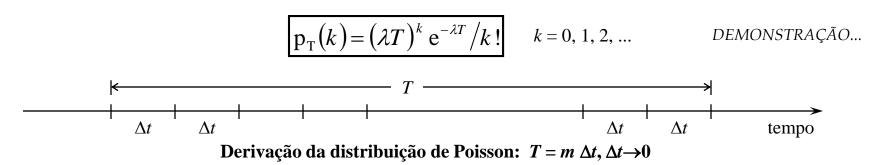
Note-se que um **processo de Poisson não tem memória**: um acontecimento num intervalo Δt é independente de acontecimentos nos intervalos anteriores ou futuros.



Intervalo temporal usado para definir um processo de Poisson

Probabilidades de chegada de chamadas

Para um intervalo finito T, a probabilidade $p_T(k)$ de k chegadas é dada por:

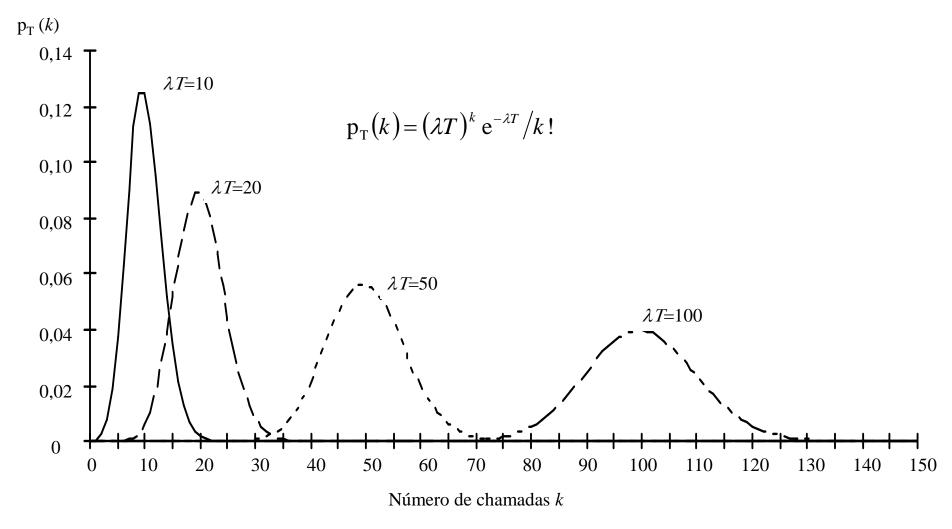


Esta é a chamada distribuição de Poisson, caracterizada por:

Média:
$$E(k) = \lambda T$$
 Variância: $\sigma_k^2 = \lambda T$ Desvio padrão: $\sigma_k = \sqrt{\lambda T}$ DEMONSTRAÇÃO...

Algumas conclusões:

- o parâmetro λ é afinal a **taxa média de chegada de chamadas** ($\lambda = E(k)/T$);
- como a média cresce mais rapidamente do que o desvio padrão, para valores grandes de λT a distribuição torna-se mais compacta em torno da média logo o número n (aleatório) de chamadas chegadas num intervalo grande T conduz a uma boa estimativa de λ através de $\lambda \approx n/T$.



Distribuição de Poisson

Combinação de fontes independentes

Na análise de tráfego em sistemas de comunicação é muitas vezes necessário combinar o tráfego gerado por fontes independentes.

Assumindo n fontes de tráfego de Poisson com taxas arbitrárias λ_i (i=1, 2, ... n), a definição de processo de Poisson conduz directamente a que o **agregado das fontes é também um processo de Poisson** com taxa:

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

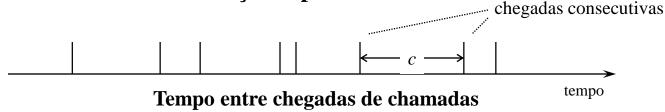
Exprime-se este facto dizendo que a soma de processos de Poisson é conservativa relativamente à distribuição: isto é, a soma retém a propriedade de Poisson.

Note-se que a probabilidade de ocorrência de uma chegada no intervalo $[t, t+\Delta t]$ é a soma das probabilidades de chegada de cada uma das n fontes independentes (assume-se que Δt é tão pequeno que não ocorrem múltiplas chamadas). Ou seja:

P (uma chegada no intervalo
$$[t, t + \Delta t]$$
) = $\sum_{i=1}^{n} (\lambda_i \Delta t) = \left(\sum_{i=1}^{n} \lambda_i\right) \Delta t = \lambda \Delta t$

Tempos entre chegadas de chamadas

Num **processo de Poisson**, o tempo c entre acontecimentos consecutivos é uma variável aleatória com **distribuição exponencial**.



De facto, a probabilidade de o tempo entre chamadas c ser $\leq t$ é:

$$P(c \le t) = 1 - P(c > t) = 1 - p_t(0) = 1 - e^{-\lambda t}$$

A função de densidade de probabilidade virá então, por derivação:

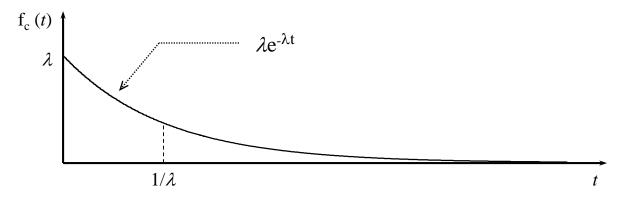
$$f_{c}(t) = \lambda e^{-\lambda t}$$

O valor médio e a variância desta distribuição exponencial são:

Média:
$$E(c) = \int_0^\infty t f_c(t) dt = 1/\lambda$$

Variância:
$$\sigma_c^2 = 1/\lambda^2$$

O valor médio do tempo entre chegadas consecutivas atrás obtido é o esperado, uma vez que, sendo a taxa de chegadas do processo de Poisson λ , então o tempo entre chegadas será $1/\lambda$.



Distribuição exponencial de tempos de chegada entre chamadas

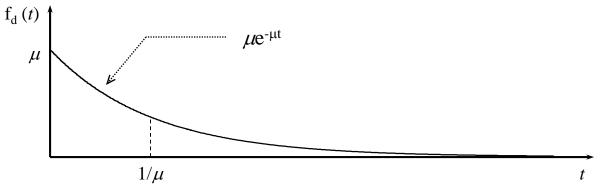
Note-se que uma variável aleatória exponencial é a única variável aleatória que tem a propriedade já referida de não memória: em termos de chegada de chamadas, a distribuição temporal dos próximos eventos é a mesma independentemente do tempo presente.

Duração de chamadas

Distribuição de durações

Para as chamadas entradas no sistema, é habitual assumir uma **distribuição** exponencial da duração d, com valor médio $E(d) = 1/\mu$:

Distribuição cumulativa de probabilidades $P(d \le t) = 1 - e^{-\mu t}$ Função de densidade de probabilidade $f_{d}(t) = \mu e^{-\mu t}$

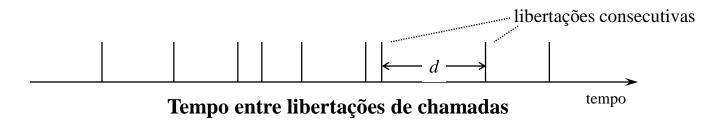


Distribuição exponencial de duração de chamadas

Esta hipótese tem sido considerada adequada para modelizar o tráfego, e tem a vantagem de apresentar a característica já referida de "sem memória": a probabilidade de uma chamada terminar num intervalo Δt é independente da sua duração anterior.

Probabilidade de libertação de chamadas

Considere-se um sistema constituído por uma fila de espera com um certo número de chamadas à espera de serem atendidas e uma única saída (uma única chamada de cada vez no sistema com possibilidade de ser libertada). Marquem-se à saída os pontos correspondentes à libertação de cada chamada.



Uma vez que os intervalos entre libertação de chamadas têm distribuição exponencial, o processo de libertação é análogo ao de chegada de chamadas, podendo concluir-se que os tempos de libertação de chamadas representam eles próprios um processo de Poisson:

P (uma libertação no intervalo $[t, t+\Delta t]$) = $\mu \Delta t$

P (nenhuma libertação no intervalo $[t, t+\Delta t]$) = $1 - \mu \Delta t$

De igual modo, para um intervalo finito T, a probabilidade $p_T(k)$ de k libertações virá:

$$p_{T}(k) = (\mu T)^{k} e^{-\mu T}/k!$$
 $k = 0, 1, 2, ...$

Logo:

Média: $E(k) = \mu T$

Variância: $\sigma_k^2 = \mu T$

Desvio padrão: $\sigma_k = \sqrt{\mu T}$

O parâmetro μ é, portanto:

$$\mu$$
 = taxa média de libertação de chamadas = $\frac{1}{\text{duração média de chamadas}}$

No caso mais geral de existirem no sistema i chamadas activas com possibilidade de serem libertadas de forma independente entre si, o processo de libertação resultante continuará a ser de Poisson (propriedade conservativa), sendo:

Ou seja:

$$\mu_i = i\mu$$

$$\mu$$
 = taxa média de libertação de chamadas = $\frac{\text{número de chamadas activas}}{\text{duração média de chamadas}}$

