

# Auto-Similaridade do Tráfego de Telecomunicações

Carlos Marcelo Pedroso

Mestrado em Engenharia Elétrica - PPGEE/UFPR

25 de novembro de 2013

## Auto-Similaridade e Long Range Dependence

Estacionariedade

LRD e SRD

Auto-similaridade e a geometria fractal

## Identificação de um processo auto-similar

Estatística R/S

Método da Variância

Método da Transformada de Wavelet

## Modelos de tráfego Auto Similares

Modelo ON-OFF

Modelo FARIMA

Movimento Browniano Fracionário

Modelo SURGE

## Estacionariedade

- ▶ A observação de uma série temporal discreta realizada em instantes de tempo  $\tau_1, \tau_2, \dots, \tau_t, \dots, \tau_n$  pode ser denotada por  $X(\tau_1), X(\tau_2), \dots, X(\tau_t), \dots, X(\tau_N)$ . Normalmente são consideradas séries temporais discretas onde as observações são realizadas em um intervalo fixo  $h$ . Quando  $N$  valores sucessivos da série forem analisados, será escrito  $X_1, X_2, \dots, X_t, \dots, X_N$  para denotar observações realizadas a intervalos de tempo equidistantes  $\tau_0 + h, \tau_0 + 2h, \dots, \tau_0 + th, \dots, \tau_0 + Nh$ . A esperança será denotada  $E(X) = \mu$ ; a variância será denotada por  $V[X] = E[(X - \mu)^2] = \sigma^2$ ; a auto-covariância com defasagem  $k$  será denotada por  $E[(X_t - \mu)(X_{t-k} - \mu)] = \gamma_k$  e a auto-correlação com defasagem  $k$  será denotada por  $\rho_k = \gamma_k / \gamma_0$ .

# Estacionariedade

- ▶ Estacionariedade estrita:  $X_t$  é estritamente estacionário se  $[X_{t_1}, X_{t_2}, \dots, X_{t_n}]$  e  $[X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}]$  possuem a mesma distribuição conjunta para todo  $n \in \mathbb{N}$ .
- ▶ Estacionariedade fraca ou de segunda ordem: A função de auto-covariância  $\gamma(r, s) = E [(X_r - \mu)(X_s - \mu)]$  deve satisfazer à relação de invariância 
$$\gamma(r, s) = \gamma(r + k, s + k) \quad \forall r, s, k \in \mathbb{Z}.$$

## Processo com Dependência de Curta Duração

- ▶ Processo com memória de curta duração ou Short Range Dependence (SRD) indica o fenômeno onde a observação corrente não é correlacionada com observações muito antigas. Para um processo *SRD*, a função de auto-correlação decai rapidamente para zero.
- ▶ O valor  $H = \frac{1}{2}$  é o limite entre a *LRD* e a *SRD*. Processos com  $\frac{1}{2} < H < 1$  são *LRD*, enquanto processos com  $0 < H < \frac{1}{2}$  são *SRD*. O efeito da *LRD* será maior quando  $H$  estiver próximo a 1.

# Processo com Dependência de Longa Duração

- ▶ É o fenômeno<sup>1</sup> ocorre quando observações de um evento são significativamente correlacionadas com observações muito distantes no tempo. Este fenômeno é de particular interesse na modelagem de tráfego, uma vez que as pesquisas comprovaram que o tráfego em redes modernas de comunicação frequentemente apresenta *LRD*.
- ▶ Em um processo *LRD* a função de auto-correlação  $\rho_k$  decai para zero muito lentamente em função do deslocamento  $k$ . Processos auto-similares apresentam *LRD*.

---

<sup>1</sup>teste123

## Processo com Dependência de Longa Duração

- ▶ Seja  $X_t$ ,  $t = 1, 2, \dots$ , um processo estacionário de segunda ordem. Se existe uma constante  $c$  tal que  $\gamma_k \sim c/k^\alpha$ ,  $0 < \alpha < 1$ , então o processo possui *LRD*.
- ▶ Hurst foi um famoso hidrólogo interessado no Egito e no rio Nilo que mostrou em 1951 que o nível do rio Nilo é uma sequência *LRD*. Por razões históricas, o parâmetro de Hurst  $H = 1 - \alpha/2$  é utilizado para representar a auto-similaridade.

# Fractais

- ▶ O modelo auto-similar foi estudado pelo matemático George Cantor no século XIX, em um estudo que o levou a teoria sobre fractais aprimorada por Benoit B. Mandelbrot.
- ▶ O valor  $H = \frac{1}{2}$  é o limite entre a *LRD* e a *SRD*. Processos com  $\frac{1}{2} < H < 1$  são *LRD*, enquanto processos com  $0 < H < \frac{1}{2}$  são *SRD*. O efeito da *LRD* será maior quando  $H$  estiver próximo a 1.

# Fractais

- ▶ A montagem do Conjunto de Cantor apresentado na abaixo, onde a sequência de iterações é representada usando a dimensão vertical, produz um objeto auto-similar. Para a montagem do conjunto, um dado segmento de reta é dividido em 3 partes iguais, sendo a parte central do segmento descartada, e assim recursivamente. O modelo geométrico referido é atingido no limite de infinitas iterações do processo descrito.



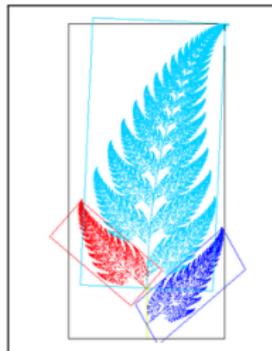
# Fractais

- ▶ Um objeto auto-similar em 2 dimensões é mostrado abaixo. Este objeto é conhecido como floco de neve Koch, formado pela divisão do segmento de reta em três partes iguais, removendo-se o segmento central e incluindo-se dois segmentos do mesmo tamanho do segmento que foi removido conectados em suas extremidades aos segmentos restantes e entre si. A regra recursiva é executada sobre todos os segmentos de reta restantes. Imagine que a montagem da figura prossiga indefinidamente. Qualquer ampliação em escala de uma porção do objeto resultante possui semelhanças entre si.



## Fractais - Exemplos

- ▶ Alguns fractais
- ▶ Fractais na natureza
- ▶ Mandelbrot set
- ▶ Ver apresentação Ashish Gupta



# Fractais

- ▶ Um dos conceitos fundamentais para a compreensão da auto-similaridade é o conceito de dimensão, que será abordado a seguir. Na geometria Euclidiana, um ponto possui dimensão zero, uma linha possui dimensão 1, um plano possui dimensão 2 e assim por diante. A dimensão fractal de um objeto é definida pela Equação 1: um objeto auto-similar é caracterizado pelo número  $N$  de cópias de si próprio criado com uma alteração de escala, enquanto  $\varepsilon$  indica o tamanho relativo da cópia em relação ao objeto original.

$$d = \frac{\log N}{\log \varepsilon} \quad (1)$$

# Fractais

- ▶ Suponha, por exemplo, o conjunto de Cantor. Cada divisão da figura original produz uma mudança de escala onde são criadas duas figuras iguais à original ( $N = 2$ ) e o tamanho da figura será 3 vezes menor ( $\varepsilon = 3$ ). O processo se repete de modo que a dimensão  $d = 0,631$  permanece invariável. Para o fractal floco de neve, a quantidade de novos segmentos criados a partir da figura original será  $N = 4$ , com seu tamanho reduzido em  $\varepsilon = 3$ . A dimensão fractal deste objeto será  $d = 1,262$ :
- ▶ É possível determinar a dimensão fractal de um objeto observado fazendo uso da definição. O procedimento, chamado *contagem de caixas*, é o seguinte:

## Dimensão Fractal

- ▶ 1- Coloque um *grid* sobre o objeto com quadrados de dimensão  $\varepsilon_0$ . Conte quantos quadrados contêm pontos da figura e chame o resultado de  $N(\varepsilon_0)$ ;
- ▶ 2- Repita o passo 1 utilizando quadrados com dimensão  $\varepsilon_1 = \varepsilon_0/2, \varepsilon_2 = \varepsilon_1/2, \dots$  e anote o resultado em  $N(\varepsilon_0), N(\varepsilon_1), \dots$ ;
- ▶ 3- A dimensão  $d$  é o número tal que  $\lim_{\varepsilon \rightarrow 0} a \cdot \varepsilon^{-d}$  (derivado da equação anterior), onde  $a$  é uma constante. Na prática,  $d$  pode se estimado como

$$d = \frac{\log N(\varepsilon_{i+1}) / \log N(\varepsilon_i)}{\log \varepsilon_{i+1} / \log \varepsilon_i} \quad (2)$$

## Dimensão Fractal

- ▶ A grande dificuldade do procedimento apresentado é selecionar o valor apropriado para  $i$ . Objetos reais não possuem um nível de granularidade infinitamente grande ou infinitamente pequeno. Além disso, a auto-similaridade pode manifestar-se apenas em determinadas escalas.
- ▶ O parâmetro de Hurst está relacionado com a dimensão fractal através da relação  $d = 2 - H$ ;

## Auto-Similaridade

- ▶ Assuma uma variável aleatória com  $E[X] = 0$ . Pela estacionariedade,  $\gamma(r, s) = \gamma(r - s, 0)$  e a auto-covariância será dada por  $\gamma_k$ . Para formular a invariância de escala, primeiro define-se o processo agregado  $X^{(m)}$  de  $X$  em um nível de agregação  $m$ .

$$X_k^{(m)} = \frac{1}{m} \sum_{t=m(k-1)+1}^{mk} X_t; \quad (3)$$

- ▶ Isso significa que  $X_t$  é particionado em blocos não sobrepostos de tamanho  $m$ , seus valores são anotados e  $k$  é utilizado como índice nestes blocos. O valor de  $X_t$  representa a quantidade de informação observada.

## Definição de Auto-Similaridade

- ▶ Seja  $\gamma^{(m)}(k)$  a função de auto-covariância de  $X^{(m)}$ .  
Assumindo que o processo é estacionário de segunda ordem, pode-se colocar a seguinte definição de auto-similaridade.

$$\gamma_k = \frac{\sigma^2}{2} \left[ (k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right], \quad \forall k \geq 1 \quad (4)$$

- ▶ implica em  $\gamma_k = \gamma_k^{(m)}$ ,  $\forall m \geq 1$ . A auto-similaridade estacionária tem sido o modelo dominante para a modelagem de tráfego em redes de comunicação modernas. A função de auto-correlação de um processo auto-similar apresenta um lento decaimento em função do deslocamento  $k$  e  $1/2 < H < 1$ . Processos SRD apresentam  $H \leq 1/2$ .

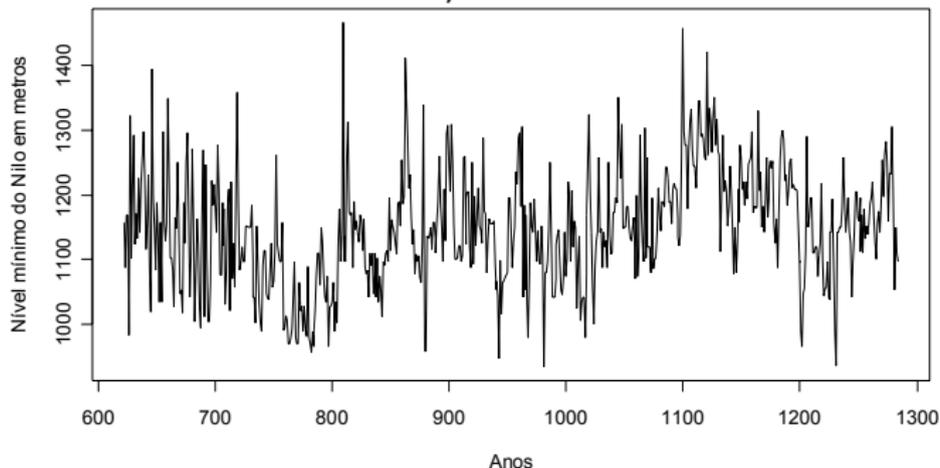
## Definição de Auto-Similaridade

- ▶ Um processo será *auto-similar do ponto de vista estocástico* se o processo mantém características estocásticas sobre uma certa faixa de escalas, de modo a satisfazer a relação dada pela equação abaixo. A interpretação desta equação é a seguinte: o processo estocástico  $X_t$  deslocado em escala por um fator constante  $m$  possui aproximadamente as mesmas características de  $X_t$  em termos de sua distribuição de probabilidade. A amplitude do processo  $X_{mt}$  resultante será o de  $X_t$  multiplicado por  $m^H$ , com  $H$  representando o parâmetro de Hurst.

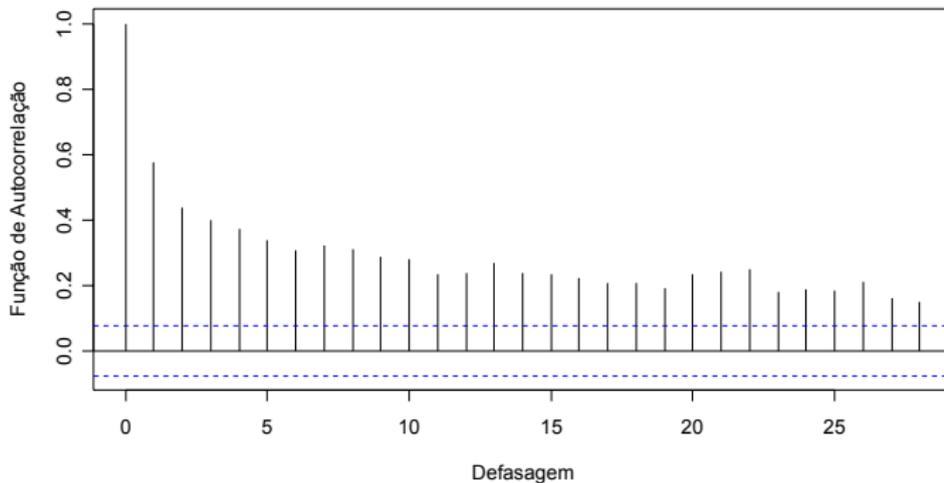
$$X_{mt} \cong_d m^H X_t, m > 0 \quad (5)$$

## Identificação de um processo auto-similar

- ▶ Um dos fatores que indicam a presença da auto-similaridade é o lento decaimento da função de auto-covariância.



# Identificação de um processo auto-similar



- ▶
- ▶ No entanto, o lento decaimento da ACF por si só não é suficiente para indicar a presença da auto-similaridade.

## Estatística R/S

- ▶ A auto-similaridade manifesta-se através de uma linha reta no gráfico *log-log* na estatística R/S. Para um dado conjunto de números  $\{X_1, \dots, X_n\}$  com média amostral  $\hat{\mu}$  e variância amostral  $\hat{S}^2(n)$ , o parâmetro de Hurst  $H$  é dado pela razão  $\hat{R}(n)/\hat{S}(n)$  (ou estatística R/S), onde  $\hat{R}$  é calculado através da Equação 6.

$$\hat{R}(n) = \max \left\{ \sum_{i=1}^k (X_i - \hat{\mu}), 1 \leq k \leq n \right\} - \min \left\{ \sum_{i=1}^k (X_i - \hat{\mu}), 1 \leq k \leq n \right\} \quad (6)$$

- ▶ O valor esperado de  $R(n)/S(n)$  assintoticamente satisfaz a relação  $E[R(n)/S(n)] \rightarrow a n^H$  conforme  $n \rightarrow \infty$  com  $\frac{1}{2} < H < 1$ .

## Método da Variância

- ▶ Este é um método simples e apresenta resultados aproximados. Consiste em verificar a auto-similaridade assintótica através do gráfico do estimador  $\sigma^{2(m)}$  da variância do processo agregado, dado pela Equação 7. Para valores suficientemente grandes de  $m$ ,  $\sigma^{2(m)} \sim c m^{-2(1-H)}$  para uma constante  $c$ . Isto pode ser verificado em uma escala logarítmica.

$$\sigma^{2(m)} = \frac{m}{N} \sum_{t=1}^{N/m} (X_t^{(m)} - \mu)^2 \quad (7)$$

- ▶ Neste método, a variância  $\sigma^{2(m)} = \text{Var}\{X_t^{(m)}\}$  em um nível de agregação  $m$  deve ser colocado em um gráfico em escala logarítmica ( $\sigma^{2(m)} \times m$ ). A inclinação assintótica da reta formada será  $\alpha = -2(1 - H)$ .

## Método da Transformada de Wavelet

- ▶ Wavelets são funções matemáticas que seccionam os dados em diferentes componentes de frequência e então estudam cada componente com uma resolução ajustada à sua escala. Elas possuem vantagens sobre os métodos tradicionais de Fourier para a análise de situações onde o sinal contém descontinuidades e picos. Wavelets foram desenvolvidas independentemente nos campos da matemática, física quântica, engenharia elétrica e geologia sísmica.
- ▶ O procedimento básico de análise de uma série temporal consiste em um algoritmo recursivo que constrói a função, desloca-a e troca sua escala. A função original é decomposta em um conjunto de escalas e representada como parâmetros da função padrão, o que torna a análise menos sensível ao ruído porque mede as flutuações médias do sinal em diferentes escalas.

## Automação da análise

- ▶ O software estatístico R possui o pacote FARma que disponibiliza funções para análise de séries auto-similares, incluindo diversos métodos para estimar o parâmetro de Hurst.



## Modelo ON-OFF

- ▶ O modelo *IPP* (*Interrupted Poisson Process*) modifica o modelo *ON-OFF* de tal maneira que, quando o processo está no estado *ON*, a chegada dos pacotes será realizada de acordo com a distribuição de Poisson com taxa  $\lambda$ .
- ▶ No entanto, estes dois casos constituem modelos que levam a séries com dependência de curta duração.

## Modelo ON-OFF

- ▶ Um dos resultados mais interessantes obtidos foi quando a distribuição do tempo ON ou do tempo OFF é modelado por uma distribuição de cauda pesada, o tráfego resultante apresenta dependência de longa duração.
- ▶ Para este caso,  $H = (3 - \alpha)/2$ , onde  $\alpha$  é o parâmetro de forma (shape) da distribuição.

## Distribuições de Cauda Pesada

- ▶ Uma das explicações sobre as causas do tráfego de telecomunicações auto-similar é a natureza dos objetos transmitidos na Internet: tamanho de arquivos, tamanho de objetos na Web, ritmos de chegada de conexões ou mesmo tempo que o usuário processa informações são descritos por uma distribuição de cauda pesada, que torna o tráfego total observado uma série temporal com memória longa.
- ▶ Uma variável aleatória  $X$  possui distribuição de cauda pesada se:

$$\Pr\{X > x\} \sim x^{-\alpha}, \quad x \rightarrow \infty, \quad 0 < \alpha < 2 \quad (8)$$

onde  $\alpha$  é chamado parâmetro de forma (*shape parameter*).

## A distribuição de Pareto

- ▶ A distribuição de Pareto é uma distribuição de cauda pesada cuja função de distribuição de probabilidade é dada pela equação abaixo, onde  $0 < \alpha < 2$  é o parâmetro de forma e  $\beta$  é chamado de parâmetro de localização. A esperança é dada por  $E[X] = \alpha \beta / (\alpha - 1)$ ,  $\alpha > 1$ .

$$\Pr\{X \leq x\} = 1 - \left(\frac{\beta}{x}\right)^\alpha, \quad \beta \leq x \quad (9)$$

- ▶ A distribuição de Pareto é uma das distribuições de cauda pesada mais utilizadas na modelagem de tráfego.

## Fractional ARIMA (FARIMA)

- ▶ O modelo *FARIMA* proposto por Hosking em 1981 é uma extensão natural do processo *ARIMA* que permite valores reais para o grau de diferenciação (parâmetro  $d$  do modelo *ARIMA*).
- ▶ A diferenciação utiliza a Equação 10 alterada com  $d$  representando um número real,  $-1/2 < d < 1/2$ . Neste modelo,  $X_t$  é um processo *LRD* se  $0 < d < 1/2$  e *SRD* se  $d = 0$ .

## Fractional ARIMA (FARIMA)

- ▶ Em um modelo *FARIMA* com *LRD*, o parâmetro de Hurst está relacionado com o parâmetro  $d$  através da relação  $H = d + \frac{1}{2}$ .
- ▶ O operador de diferenças é redefinido como

$$\nabla^d X_t = \sum_{i=0}^{\infty} \binom{d}{i} (-1)^i X_{t-i}, \quad -1/2 < d < 1/2 \quad (10)$$

- ▶ e o coeficiente binomial pode ser interpretado como

$$\binom{d}{i} (-1)^i = \frac{\Gamma(-d + i)}{\Gamma(-d)\Gamma(i + 1)} \quad (11)$$

# Movimento Browniano

- ▶ O termo movimento Browniano<sup>2</sup> refere-se a:
  1. O fenômeno físico em que minúsculas partículas imersas em um fluido movem-se randomicamente; ou
  2. Os modelos matemáticos utilizados para descrever tais movimentos randômicos.
  
- ▶ Ver:
  1. <http://www.youtube.com/watch?v=6VdMp46ZIL8>
  2. <http://www.youtube.com/watch?v=BaX7DOSvK1k>

---

<sup>2</sup>em homenagem ao botânico Robert Brown

# Movimento Browniano

- ▶ Matematicamente, o movimento Browniano pode ser caracterizado por um processo de *Wiener*, que é um processo contínuo no tempo com incrementos independentes. Para cada valor positivo de  $t$ , denote o valor do processo por  $W_t$ . O processo será caracterizado pelas duas condições seguintes:
  1. Para  $0 < s < t$ ,  $W_t - W_s \sim N(0, t - s)$ .  $N(\mu, \sigma^2)$  denota a distribuição normal com média  $\mu$  e variância  $\sigma^2$ ;
  2. Para  $0 \leq s < t \leq u < v$ , ou seja, dois intervalos  $[s, t]$  e  $[u, v]$  não sobrepostos,  $W_t - W_s$  e  $W_v - W_u$  são variáveis aleatórias independentes.

# Movimento Browniano Fracionário

- ▶ No movimento Browniano fracionário ( $fBm$ ), os incrementos também são normalmente distribuídos – mas não independentes. Para um número  $H$  com  $0 < H < 1$  em um intervalo de tempo  $[s, t]$ ,  $0 < s < t$ , o valor do processo  $W_t - W_s \sim N(0, (t - s)^H)$ . Seja o valor  $\delta = t - s$  a escala de observação. O incremento do processo estará relacionado em todas as escalas. Este modelo leva a um processo auto-similar, com dependência de longo prazo.

# Movimento Browniano Fracionário

- ▶ O modelo  $fBm$  tem sido utilizado para estudar o tráfego em redes de comunicação digital. Uma variação proposta é o movimento Browniano multi-fracionário ( $mBm$ ), onde ao invés de utilizar-se um número  $H$  se utiliza uma função contínua no tempo  $H(t)$ , chamada função de Hölder, e tem merecido atenção dos pesquisadores.
- ▶ No entanto, Darryl Veich, et al., questionam a qualidade das ferramentas matemáticas disponíveis atualmente para a análise das propriedades multi-fractais do tráfego.

## Scalable URL Reference Generator

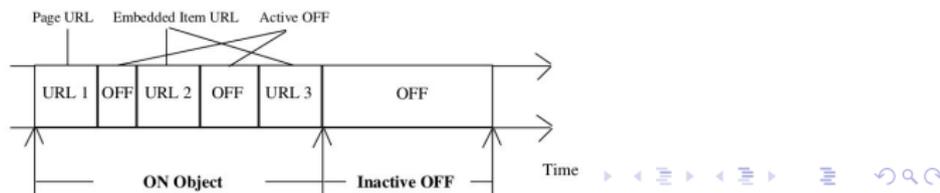
- ▶ Modelos de desempenho para servidores Web podem utilizar ou não características de comportamento do usuário. Modelos baseados na teoria de filas, como o modelo  $M/G/1/K*PS$  não consideram o comportamento do usuário nem o conteúdo do servidor. Por outro lado, no modelo *SURGE* o conhecimento do comportamento do usuário e do conteúdo do servidor são utilizados para atingir uma melhor aproximação do tráfego real.

## Scalable URL Reference Generator

- ▶ O modelo *SURGE* para geração de tráfego é baseado em um autômato *ON-OFF* que captura o comportamento do usuário. Quando o sistema está no estado *ON*, a sessão está ativa enviando os objetos requisitados na sessão. O intervalo de tempo entre os arquivos enviados durante a sessão é denominado de tempo *active-off*. O tamanho dos arquivos e o número de referências em uma sessão de usuário também é utilizada. As principais variáveis do modelo são:

## Scalable URL Reference Generator

- ▶ *Tempo OFF*: É o tempo que o usuário permanece pensando. Normalmente modelado por uma distribuição de Pareto;
- ▶ *Tamanho dos arquivos*: É o tamanho dos objetos transmitidos. Normalmente modelado por uma distribuição de Pareto;
- ▶ *Número de referências*: Número de arquivos transmitidos em uma sessão de usuário. Também modelado normalmente por uma distribuição de Pareto;
- ▶ *Tempo active-off*: É o intervalo de tempo entre os arquivos transmitidos em uma sessão de usuário. Modelada pela distribuição de Weibull;



# Scalable URL Reference Generator

- ▶ *Popularidade*: É o número relativo de acessos realizados a um arquivo individual. A popularidade de arquivos em servidores Web segue, via de regra, a lei de Zipf. A lei de Zipf argumenta que se os arquivos forem ordenados do mais popular para o menos popular, então o número de referências a um arquivo  $P$  tende a ser inversamente proporcional a sua posição na classificação  $r$ , ou  $P = kr^{-1}$  para uma constante positiva qualquer  $k$ ;

## Scalable URL Reference Generator

- ▶ *Localidade temporal*: A localidade temporal assume que, uma vez tendo sido requisitado um arquivo, a probabilidade de que ele seja novamente requisitado no futuro aumenta. Para o estudo desta variável os acessos são armazenados em uma estrutura de pilha. A distância entre os acessos nesta pilha são estudados e modelados comumente com uma distribuição lognormal.
- ▶ Tráfego agregado resultante: auto-similar.

## Exercício

- ▶ Escolher um método para geração de uma série temporal autosimilar.
- ▶ Avaliar a qualidade através de:
  1. Inspeção visual da série.
  2. Verificação da função de auto-correlação.
  3. Estimação do parâmetro de Hurst utilizando os métodos disponíveis no pacote fArma (R).