MŰEGYETEM 1782

Statistical analysis of peer-to-peer live streaming traffic

Levente Bodrog[1] Ákos Horváth[1] Miklós Telek[1]

[1] Technical University of Budapest

Probability and Statistics with Applications, 2009

Outline

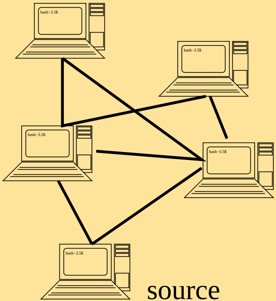
Motivation

- Nowadays observations on high speed network traffic
  - heavy tailed distributions,
  - long-range dependency,
  - multi-fractal behaviour.
- Emerging peer-to-peer (P2P) traffic includes
  - file sharing,
  - military, telecommunication, bioinformatics and other research,
  - live streaming.

A new kind of traffic, the P2P live streaming traffic, is analysed.

Technical introduction

P2P is for TV programme broadcasting.



- One source, *without high resources*
- spreads the *stream* via packets over
- the estabilished *overlay topology*.

Advantages:

- simple,
- no special infrastructure needed,
- robust.

Outline

Heavy tailed distributions

A distribution is heavy tailed if its complementary cumulative distribution function is

$$1 - F_Y(x) = x^{-\alpha} L(x),$$

where $\lim_{x \to \infty} L(ax)/L(x) = 1$ for $a > 0$ (e.g. Pareto family with cumulative distribution function $1 - (x_m/x)^{-k}, \quad x_m > 0, k > 0$).

α is the tail index of the distribution.

Tests for heavy tailed distributions

These tests estimate the tail index of a distribution based on its samples.

Hill estimator

$$\alpha_{n,k} = \left(\frac{1}{k} \sum_{i=0}^{k-1} \left(\log X_{(n-i)} - \log X_{(n-k)} \right) \right)^{-1},$$

where $X_{(1)} \leq \ldots \leq X_{(n)}$ denotes the order statistics of the dataset.

Dynamic quantile-quantile regression plot

The slope of the linear regression of

$$\left\{ \left(-\log \left(1 - \frac{j}{n+1} \right), \log X_{(j)} \right), n-k+1 \leq j \leq n \right\}$$

gives $\alpha_{n,k}$.

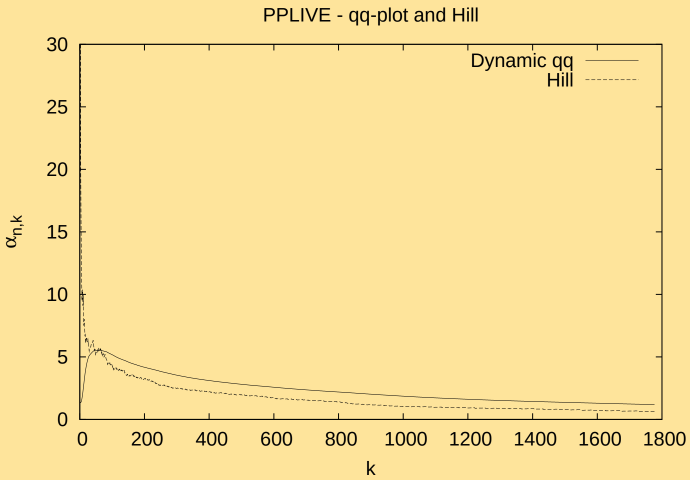


Figure: The Hill and the dynamic qq-plot of a P2P live streaming trace

Outline

Self-similarity

First we define self-similarity
A stochastic process $\mathcal{X} = \{X_i, i = 0, 1, 2, \ldots\}$ with aggregated process

$$\mathcal{X}^{(m)} = \left\{ X_k^{(m)} = \ldots, \frac{X_{km} + \ldots + X_{(k+1)m-1}}{m}, \ldots, \forall k \right\}$$

is exactly self-similar if $\mathcal{X} \stackrel{d}{=} m^{1-H} \mathcal{X}^{(m)}$, i.e., if $\mathcal{X}$ and $\mathcal{X}^{(m)}$ are identical within a scale factor in a finite dimensional distribution sense.
Here H is the Hurst, or the self-similarity, parameter.

Definition of LRD

A process is long-range dependent (LRD) if its autocorrelation coefficients (ϱ_k) are not summable, i.e.,

$$\lim_{N\to\infty} \sum_{k=0}^{N} \varrho_k = \infty.$$

- It is observed through the self-similarity.
- Self-similarity is determined based on the Hurst parameter,
- if $0.5 \leq H \leq 1$ then the trace is self-similar and it is also long-range dependent.

Variance time plot

For self-similar time series $\{X_i\}$ (with m-aggregated process $X^{(m)}$)

$$\operatorname{Var} X^{(m)} \sim m^{-\beta}, \text{ as } m \to \infty, 0 < \beta < 1.$$

The slope of the linear regression of the plot $\log \operatorname{Var} X^{(m)}$ versus $\log m$ gives β and $H = 1 - (\beta/2)$

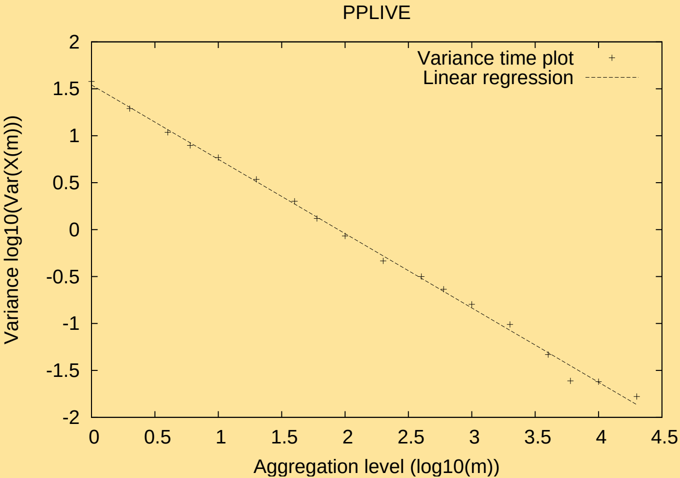


Figure: Variance time plot of a P2P live streaming trace

R/S plot

H is the slope of the linear regression of

$$R/S(n) = \frac{1}{S(n)}\left(\max_{0\le k\le n}\left(Y(k) - \frac{k}{n}Y(n)\right) - \min_{0\le k\le n}\left(Y(k) - \frac{k}{n}Y(n)\right)\right),$$

the scaled difference between the fastest and the slowest arrival period considering n arrivals.

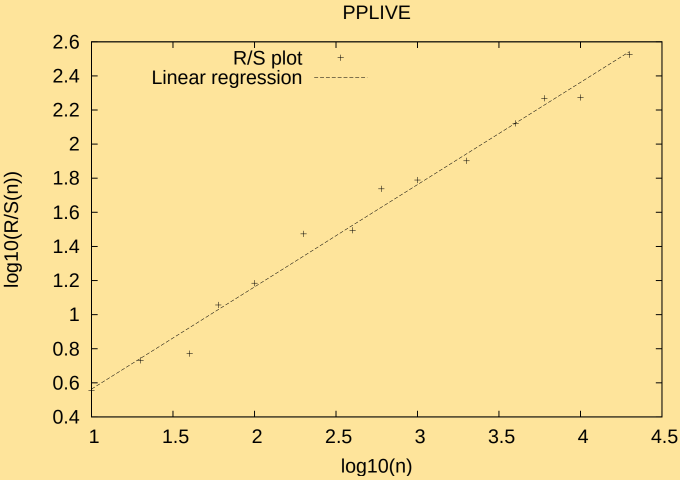


Figure: R/S plot of a P2P live streaming trace

Outline

Definition

Considering $\mathrm{E}\left(\left|X^{(m)}\right|^{q}\right)$, the absolute moments of the m-aggregated arrival process:

- self-similarity: one scaling parameter (the Hurst parameter) for all moments,
- multi-fractal behaviour:
  - different scaling for the qth absolute moment,
  - results in a spectrum depending on q.

Test – Legendre spectrum

Considering the rate process of a continuous time process:

- its local scaling exponent is $\alpha(t)$ at time t,
- the “number of” $\alpha(t)$s falling ε within α is the multi-fractal spectrum, $f_L(\alpha)$,
- the scaling of the absolute moments is $T(q)$, the partition function,
- the Legendre transform of the partition function is also the multi-fractal spectrum.

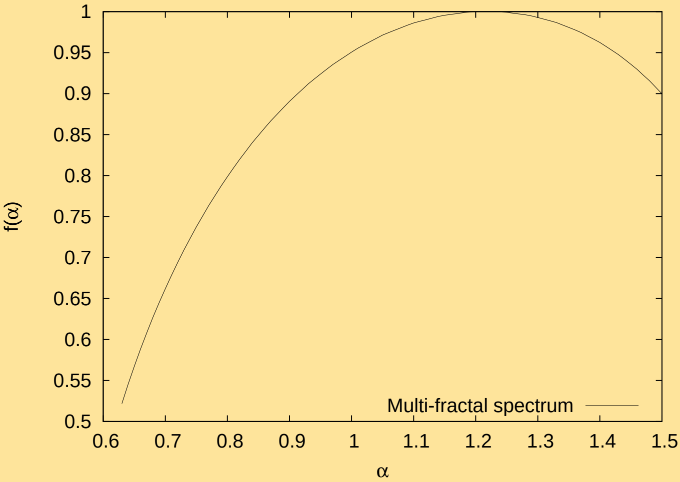
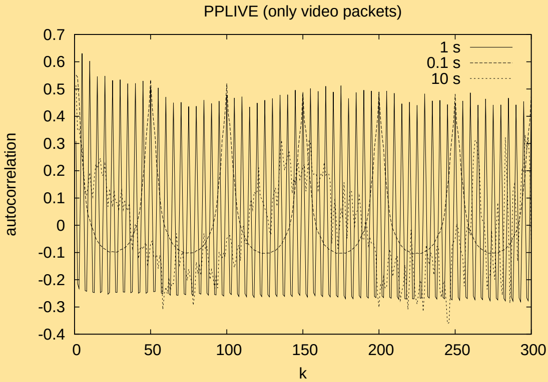


Figure: The multi-fractal spectrum of a P2P live streaming trace

Outline

Lag-k correlation

$$r(k) = \frac{\mathrm{E}\Big(\big(X_i - \mathrm{E}(X)\big)\big(X_{i+k} - \mathrm{E}(X)\big)\Big)}{\sigma_X \sigma_X}$$



The analysed P2P live streaming traffic traces:

- are heavy tailed,
- are long-range dependent considering several time scales,
- and have rich multi-fractal spectrum.

Consequences:

- the heavy tailed distributions degrades the quality of service parameters in the network,
- both the LRD and multi-fractal behaviour should taken into consideration when one considers P2P live streaming traffic.