An improvement of the Energy function*

Konstantinos Drakakis† University of Edinburgh

Dragan Radulović‡ Florida Atlantic University

July 29, 2004

Abstract

We propose an alternative definition/method of computation of the Energy function of a stationary time series, which yields more reliable results than the traditional one. We also propose the more intuitively appealing Averaging function, a simple renormalization of the Energy function.

1 Introduction

The Energy function of a time series is, in brief, a function yielding, for every scale, the energy of the Haar wavelet transform of the series contained in this particular scale (a more precise definition will be given below). Ever since its introduction in [2, 3], it has proved a very useful tool in Internet traffic research. Unfortunately, the way it has been defined and used so far makes its results quite unreliable for coarse scales, and thus inadequate for some particular directions of research (see [5, 6]), for which otherwise it would have been ideal.

In what follows, we will specify exactly what the Energy function is and in which way it is unreliable, we will modify its definition in order to make it more reliable, and we will also derive some alternative representations for it, linking it to other important statistical functions of the series, such as the autocorrelation. Moreover, we will derive a variant of the Energy function, called the Averaging function, which is, in our opinion, more intuitive, and we will provide the exact conversion formula from the one to the other.

We will demonstrate the improvement induced by the modified definition by using examples of real Internet traffic data sets, details for which can be found in [6]. Until the very end of the paper, we will keep the discussion in a fully general context, and no specific mention to Internet traffic applications will be made. At the end, however, we will dedicate a small section to Internet traffic applications.

2 Correlation and Covariance

Let $\{X_k\}_{k=1}^n$ and $\{Y_k\}_{k=1}^n$ represent stationary time series (it is possible that $n = \infty$). Their *correlation* is defined to be:

$$\mathbf{Corr}(k) = \frac{\mathbf{E}[(X_1 - \overline{X_1})(Y_{1+k} - \overline{Y_1})]}{\sqrt{\mathbf{Var}(X_1)}\sqrt{\mathbf{Var}(Y_1)}}$$

and their *covariance*:

$$\mathbf{Cov}(k) = \mathbf{E}[(X_1 - \overline{X_1})(Y_{1+k} - \overline{Y_1})]$$

*A preliminary version of this paper appeared in Proceedings of SPIE Vol. 4868 (2002) under the title: "Some results on the multiresolution structure of Internet traffic traces".

†The author was a scholar of the Lilian Boudouris Foundation. This work was completed while the author was affiliated to Princeton University.

‡The largest portion of the research for this paper was conducted while the author was affiliated to Princeton University, and was partially supported by AT&T Research Center.

where $\overline{X}, \mathbf{Var}(X)$ denote the mean value and the variance of the variable X, respectively, while the *autocorrelation* and *autocovariance* are defined analogously for the special case where $Y = X$[1].

3 First (standard) definition of the Energy and Averaging functions

The first definition of the Energy function we provide is completely equivalent to the (standard) definition given in [2, 3] and very similar to the Normalized Variance appearing in [1]. We restate it here in order to a) establish our notation, b) provide a basis of comparison with the more complicated equations of the second and main definition following shortly, and c) introduce the Averaging function in a simpler setting.

Given a stationary sequence $\{X_i\}_{i=1}^{M}$, for some $M = 2^m$, define

$$
\begin{aligned}
X_{j,k} &= 2^{-j} \sum_{i=(k-1)2^j+1}^{k2^j} X_i & j &= 0, ..., m & k &= 1, ..., 2^{m-j} \\
m_{j,k} &= |X_{j,2k} - X_{j,2k-1}| & j &= 0, ..., m-1 & k &= 1, ..., 2^{m-j-1} \\
\widetilde{X}_{j,k} &= 2^{-j/2} \sum_{i=(k-1)2^j+1}^{k2^j} X_i & j &= 0, ..., m & k &= 1, ..., 2^{m-j} \\
d_{j,k} &= \frac{1}{\sqrt{2}} \left| \widetilde{X}_{j,2k} - \widetilde{X}_{j,2k-1} \right| & j &= 0, ..., m-1 & k &= 1, ..., 2^{m-j-1}
\end{aligned}
$$

For $M_j = M/2^{j+1}$ and $p > 0$, we *preliminarily* define the **p-Averaging** $A_{j+1}^{[p]} = \left(\frac{1}{M_j} \sum_{k=1}^{M_j} (m_{j,k})^p \right)^{1/p}$ and the **Energy** $E_{j+1} = \frac{1}{M_j} \sum_{k=1}^{M_j} (d_{j,k})^2, \; j = 0, ..., m-1$ (and in fact this is the standard definition of the Energy function appearing in [2, 3], in connection to the Haar wavelet coefficients). For $p = 2$: $\log_2(E_j) = j - 2 + 2\log_2 A_j^{[2]}, \; j = 1, ..., m$.

4 Criticism and modification

According to the standard definition, the computation at the scale 2^j, for either the Energy or the Averaging functions, involves a mean value over 2^{m-j-1} numbers. Consequently, for j close to m, the sum will comprise an insufficient number of terms to produce a reliable mean, assuming that such a mean is indeed the ultimate goal of the computation, and undesirable oscillations may be observed (see Fig. 1(a)). Intuitively, it is expected that averaging over more numbers will lead to smoother curves. Thus, we will modify the method of computation of the Energy and p-Averaging function, so that we allow for averaging over the maximum number of terms possible, i.e. 2^{m-1}, by using a standard technique of statistics, known as *overlapping blocks*.

In the context of wavelets, the definition modification amounts to averaging the Energy function computations over all possible choices of the "origin" for the Haar basis underlying the definition of the $d_{j,k}$; this type of averaging is now commonly done in situations where the use of a wavelet basis introduces a translation non-invariance that was not present earlier.

---

[1] In practice the series we work with are of finite length and we use the estimators

$$\mathbf{Corr}(k) = \frac{\sum_{i=1}^{n-k} (X_i - \overline{X}_n)(Y_{i+k} - \overline{Y}_n)}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2}}$$

$$\mathbf{Cov}(k) = \sum_{i=1}^{n-k} (X_i - \overline{X}_n)(Y_{i+k} - \overline{Y}_n)$$

where $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ and similarly for Y.

5 The second definition

We will substitute the operator $\frac{1}{M_j}\sum_{\cdot=1}^{M_j}\cdot$, appearing in the formulas, by $\mathbf{E}$, the expectation operator. Moreover, for reasons of clarity, we will shift left the domain of A by 1, i.e. it will start with 0. The same name and notation will be kept for the new functions emerging out of these definitions, which, explicitly, take the form: $A_j^{[p]} = (\mathbf{E}(m_{j,1})^p)^{\frac{1}{p}}$ and $E_j = \mathbf{E}(d_{j,1})^2$. Notice here that stationarity implies that $m_{j,k} = m_{j,1}$, and $d_{j,k} = d_{j,1}$, the equality holding in distribution; and the probability measure with respect to which $\mathbf{E}$ is computed is obtained from the finite dimensional distributions of the original time series, transformed according to the defining formula for $d_{j,1}$ in the case of the Energy function, or $m_{j,1}$ in the case of the Averaging function.

How different is the new method from the old? Stationarity can now be fully exploited, in the sense that there is no "beginning of time" for the sequences, as if they were ordered circularly (on a ring), not linearly (on an interval). Thus, for example, computing $A_0^{[2]}$, in addition to considering the differences used before, i.e. $X_{0,1} - X_{0,2}, X_{0,3} - X_{0,4}, ..., X_{0,2^m-1} - X_{0,2^m}$, one will also take into account $X_{0,2} - X_{0,3}, X_{0,4} - X_{0,5}, ..., X_{0,2^m-2} - X_{0,2^m-1}, X_{0,2^m} - X_{0,1}$.

In practice, this means that the computations will be done according to the formula:

$$
\begin{aligned}
A_j^{[p]} &\approx \left(\frac{1}{2^{j+1}}\sum_{l=0}^{2^{j+1}-1}\left(\frac{1}{2^{m-j-1}}\sum_{k=1}^{2^{m-j-1}}\left|F_{kl}^1(j) - F_{kl}^2(j)\right|^p\right)\right)^{1/p} \\
F_{kl}^1(j) &= \frac{1}{2^j}\sum_{i=1}^{2^j} X_{1+((k-1)2^{j+1}+l+i-1) \bmod 2^m} \\
F_{kl}^2(j) &= \frac{1}{2^j}\sum_{i=1}^{2^j} X_{1+((k-1)2^{j+1}+2^j+l+i-1) \bmod 2^m}
\end{aligned}
$$

which can be simplified into:

$$
\begin{aligned}
A_j^{[p]} &\approx \left(\frac{1}{2^{m+jp}}\sum_{l=0}^{2^{j+1}-1}\sum_{k=1}^{2^{m-j-1}}\left|\sum_{i=1}^{2^j}\left(G_{ikl}^1(j) - G_{ikl}^2(j)\right)\right|^p\right)^{1/p} \\
G_{ikl}^1(j) &= X_{1+((k-1)2^{j+1}+l+i-1) \bmod 2^m} \\
G_{ikl}^2(j) &= X_{1+((k-1)2^{j+1}+2^j+l+i-1) \bmod 2^m}
\end{aligned}
$$

and similarly for the Energy function. Observe that the inner mean value (on k) is the standard definition. The outer (on l) is the improvement: now, for every j, the calculation involves a mean value taken on 2^m points.

6 Connection between the Averaging function and the autocorrelation

In the case $p = 2$, an alternative representation exists, which can be derived directly by the definition, after a simple computation. The defining formulas yield:

$$\left(A_j^{[2]}\right)^2 = \mathbf{E}(m_{j,1})^2 \overset{*}{=} \mathbf{E}\left(X_{j,2} - X_{j,1}\right)^2 = 2^{-2j}\mathbf{E}\left(\sum_{i=2^j+1}^{2^{j+1}} X_i - \sum_{i=1}^{2^j} X_i\right)^2 =$$

$$= 2^{-2j}\mathbf{E}\left(\left(\sum_{i=2^j+1}^{2^{j+1}} X_i\right)^2 + \left(\sum_{i=1}^{2^j} X_i\right)^2 - 2\sum_{i=1}^{2^j}\sum_{k=2^j+1}^{2^{j+1}} X_i X_k\right) =$$

$$= 2^{-2j}\mathbf{E}\left(\sum_{i=1}^{2^j}\sum_{k=1}^{2^j} X_{i+2^j}X_{k+2^j} + \sum_{i=1}^{2^j}\sum_{k=1}^{2^j} X_i X_k - 2\sum_{i=1}^{2^j}\sum_{k=1}^{2^j} X_i X_{k+2^j}\right) \overset{**}{=}$$

$$= 2^{-2j+1}\mathbf{Var}\,(X)\left(\sum_{i=1}^{2^j}\sum_{k=1}^{2^j}\mathbf{R}(k-i) - \sum_{i=1}^{2^j}\sum_{k=1}^{2^j}\mathbf{R}(k+2^j-i)\right) \overset{***}{=}$$

$$= 2^{-2j+1}\mathbf{Var}\,(X)\left(2^j + 2\sum_{k=1}^{2^j}\sum_{i=k+1}^{2^j}\mathbf{R}(i-k) - 2^j\mathbf{R}(2^j) + \sum_{k=1}^{2^j}\sum_{i=k+1}^{2^j}\big(\mathbf{R}(2^j+(i-k)) + \mathbf{R}(2^j-(i-k))\big)\right) =$$

$$= 2^{-2j+1}\mathbf{Var}\,(X)\left(2^j + 2\sum_{k=1}^{2^j}\sum_{i=1}^{2^j-k}\mathbf{R}(i) - 2^j\mathbf{R}(2^j) + \sum_{k=1}^{2^j}\sum_{i=1}^{2^j-k}\big(\mathbf{R}(2^j+i) + \mathbf{R}(2^j-i)\big)\right)$$

In $(*)$ we see that, without loss of generality, we can take the time series to have zero mean; in $(**)$ we use the stationarity of the time series to introduce the autocorrelation $\mathbf{R}$; in $(***)$ we use the symmetry of the autocorrelation ($\mathbf{R}(i) = \mathbf{R}(-i)$). If we also change the order of the sums according to $\sum_{k=1}^{2^j}\sum_{i=1}^{2^j-k}\cdot = \sum_{i=1}^{2^j-1}\sum_{k=1}^{2^j-i}\cdot$, we obtain the following result:

$$\left(A_j^{[2]}\right)^2 = \frac{\mathbf{Var}\,(X)}{2^{j-1}}\left\{1 - \mathbf{R}\left(2^j\right) + 2\sum_{i=1}^{2^j-1}\left[\left(1 - \frac{i}{2^j}\right)\left[2\mathbf{R}\,(i) - \mathbf{R}\left(2^j+i\right) - \mathbf{R}\left(2^j-i\right)\right]\right]\right\} \tag{1}$$

This representation proves that (the square of) A_j is really an affine transformation of the autocorrelation. Note that the $A_j^{[p]}$ can be computed via a fast algorithm of complexity $\Theta\left(m2^m\right)$; for A_j this can be seen directly from (1), since $\mathbf{R}$ as well as the convolution in the last term of (1) can be computed via the FFT. For general p, one can use the multiresolution character of what is essentially a computation with Haar wavelet coefficients to obtain a $\Theta\left(m2^m\right)$ algorithm (this also works for $p = 2$).

For all ordinary applications, $A_j^{[2]}$ can be used almost exclusively, because it is easier to analyze, and simply connected to the commonly adopted (at least at present) E_j; we will indicate it just by A_j and refer to it simply as the *Averaging function*.

7 Motivation for the introduction of the Averaging function

The introduction of the Averaging function has been motivated by the following heuristic: let X_i represent a time series value at the ith (discrete) time instant, and assume the sequence is ergodic, i.e. that $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbf{E}\left(X_1\right)$. What is the rate of this convergence? The way to determine the rate is to find a function $\psi(n)$ such that

$\psi(n)\sqrt{\mathbf{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mathbf{E}\left(X_1\right)\right|^2\right)}\underset{n\to\infty}{\longrightarrow}1$. Typically, the function ψ is increasing in n, at least for n sufficiently large, expressing the fact that the average $\frac{1}{n}\sum_{i=1}^{n}X_i$ gets closer to $\mathbf{E}(X_1)$ when n increases. In practice, because $\mathbf{E}(X_1)$ is unknown, one can still get a handle on ψ by comparing two different averages:

$$\mathbf{E}\left(\psi^2(n)\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}X_{i+n}\right|^2\right)=\psi^2(n)\mathbf{E}\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mathbf{E}(X_1)\right)-\left(\frac{1}{n}\sum_{i=1}^{n}X_{i+n}-\mathbf{E}(X_1)\right)\right|^2\right)=$$
$$=\psi^2(n)\mathbf{E}\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n}Y_i\right|^2+\left|\frac{1}{n}\sum_{i=1}^{n}Y_{i+n}\right|^2-2\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}Y_{i+n}\right)\right\}\right)\overset{*}{=}$$
$$=\frac{\psi^2(n)}{n^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}(Y_iY_j)+\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}(Y_{i+n}Y_{j+n})-2\mathbf{E}\left(\left(\sum_{i=1}^{n}Y_i\right)\left(\sum_{i=1}^{n}Y_{i+n}\right)\right)\right\}\overset{**}{=}$$
$$=2\frac{\psi^2(n)}{n^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}(Y_iY_j)-\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}(Y_iY_{j+n})\right\}=2\frac{\psi^2(n)}{n^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}[\mathbf{E}(Y_iY_j)-\mathbf{E}(Y_iY_{j+n})]\right\}\quad(2)$$

where $Y_i=X_i-\mathbf{E}(X_1)$ was set. In $(**)$ we used stationarity. Notice also that $(*)$ proves that the limit of (2) is positive and not greater than 2.

We claim that, if the X_i are *asymptotically independent*, i.e. $\forall i, \mathbf{E}(X_iX_{i+j})-\mathbf{E}(X_1)^2)=\mathbf{E}(Y_iY_{i+j})\to 0$, as $j\to\infty$ (irrespectively of the rate), then (2) has a *strictly* positive limit, and, therefore, that it can be used to estimate ψ (within a multiplicative constant). Because asymptotic independence is satisfied by most processes arising in applications (it is extremely unlikely that two random variables arbitrarily long apart will exhibit strong dependence), this method can be applied to a wide range of processes, and is, for this reason, very powerful.

Indeed, as $n\to\infty$, for fixed i,j, $\mathbf{E}(Y_iY_{n+j})\to 0$, and therefore it becomes infinitely smaller than $\mathbf{E}(Y_iY_j)$. The RHS of (2) shows then that the limit has to be strictly positive.

Thus, if a random variable $\epsilon_{j,k}$ is defined by: $\epsilon_{j,k}=\psi(2^j)\frac{1}{2^j}\left|\sum_{k2^j+1}^{(k+1)2^j}(X_i-\mathbf{E}(X_1))\right|=\psi(2^j)|m_{j,k}|$ then $A_{j+1}=\left(\frac{1}{M_j}\sum_{k=1}^{M_j}|m_{j,k}|^2\right)^{1/2}=\frac{1}{\psi(2^j)}\left(\frac{1}{M_j}\sum_{k=1}^{M_j}|\epsilon_{j,k}|^2\right)^{1/2}\approx\frac{C}{\psi(2^j)}$.

If M_j is large, A_{j+1} gives us a way to estimate $\psi(2^j)$. In particular, if $\psi(n)=n^\alpha$, then:

$$\log_2 A_{j+1}^{[p]}=K-\alpha j \text{ and } \log_2 E_{j+1}=2K-1+j(1-2\alpha)$$

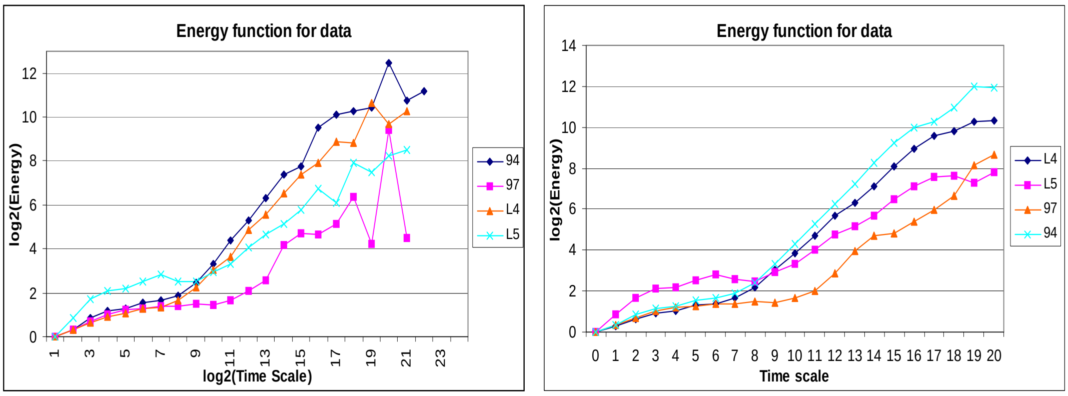
Therefore, the binary logarithms of the Averaging resp. Energy functions of the processes that "average" with a rate of n^{-a} will be straight lines with a slope of $-\alpha$ resp. $1-2\alpha$. Under our modified definitions, the Energy functions assume a slightly different, smoother form (see Fig. 1).

We conclude this section with some remarks:

Remark 1:. In the case of processes with marginals possessing finite moments only up to the nth order, the Energy function only makes sense if $n\geq 2$, just as $A_j^{[p]}$ makes sense only if $p\leq n$. The reason is that the $\epsilon_{j,k}$ possess, by definition, as many finite moments as the process itself.

Remark 2:. A_j has another interpretation that can prove essential for applications. Namely:

$$A_j=\sqrt{\mathbf{E}\left(m_{j,1}\right)^2}=\frac{1}{2^j}\sqrt{\mathbf{E}\left(X_{j,2}-\mathbf{E}X_{j,2}+\mathbf{E}X_{j,1}-X_{j,1}\right)^2}=\frac{\sqrt{2}}{2^j}\sqrt{\mathbf{Var}\left(X_{j,1}\right)-\mathbf{Cov}\left(X_{j,1},X_{j,2}\right)}\quad(3)$$



(a) (b)

Figure 1: Energy functions of some time series (corresponding to Internet traffic data sets described in [6]), according to the old and to the new definition. Notice how much smoother curves the new definition produces.

8 An application to Internet traffic

Suppose we are studying a network link through which n users send data; the data each user sends is modeled as some function of time $W_i(t),\ i = 1, \ldots, n$ (with units of e.g. bits/sec). The total traffic per time unit through the link will be in this case $W(t) = \sum_{l=1}^{n} W_l(t)$. We can now discretize each user's traffic by dividing the time axis into adjacent time intervals of length $\Delta > 0$, and compute the total traffic the ith user generates within the kth interval ($j\Delta \leq t < (j+1)\Delta$) as $X_k^i = \int_{k\Delta}^{(k+1)\Delta} W_i(t)dt$. We can discretize in the same way the (normalized) aggregate traffic $X_k := \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \frac{1}{\sqrt{n}} \sum_{l=1}^{n} W_l(t)dt$, thus creating a discrete time series. Let us assume that the statistics of user behavior are the same for all users. Applying (3), we find the Averaging function as

$$A_j \approx K \frac{\sqrt{\mathbf{Var}\left(\int_0^{2^j\Delta} W_1(t)dt\right) - \mathbf{Cov}\left(\int_0^{2^j\Delta} W_1(t)dt, \int_{2^j\Delta}^{2^{j+1}\Delta} W_1(t)dt\right)}}{2^j\Delta}.$$

for some positive constant K, which also absorbs n, as it is now fixed. This formula connects the Energy/Averaging function of the total traffic to the behavior of one (typical) user.

9 Discussion

We demonstrated that, by fully exploiting the stationarity of a time series, we can alter the traditional definition of the Energy function so as to render it much more stable numerically. We also proposed a renormalization of the Energy function, which we called the Averaging function, and which is perhaps more appealing, as it demonstrates directly how fast the convergence of the mean value estimators of the series to the true mean is. Moreover, we showed that the (square of the) Averaging function is an affine transformation of the autocorrelation of the time series. Finally, we gave a simple example of an application in Internet traffic study, where we linked the Averaging function to the statistical behavior of a user.

This application may seem to be rather a "toy" example; let us add, though, that the Averaging function plays a key role in the exploration of the multifractal structure of a time series, and that characteristic time scales of the series can be immediately "read" off the Averaging function (see [6, 4] for more details).

Despite the usefulness of the Energy function, no attempt has —to the best of our knowledge— been undertaken so far to implement the Energy function using wavelets other than the Haar. Such an implementation, especially with wavelets possessing more vanishing moments, might enhance the numerical accuracy of the method even further.

10 Acknowledgements

The authors would like to thank Prof. Ingrid Daubechies of the Department of Mathematics, Princeton University, for improving decisively this paper both æsthetically and scientifically, working as hard as the authors themselves did. K. Drakakis would also like to thank the Lilian Boudouris Foundation for the scholarship it granted him. Finally, the authors would like to thank NSF, AFOSR, ONR, DARPA, and AT&T for their financial support, as well as the anonymous reviewers for their comments, which definitely resulted in a better paper.

References

[1] V. Paxson, S. Floyd. *Wide-Area Traffic: The Failure of Poisson Modelling* IEEE/ACM Transactions on Networking, Vol. 3 No. 3, pp. 226-244, June 1995

[2] A. Feldmann, A. Gilbert, W. Willinger. *Investigating the multifractal nature of Internet WAN traffic* Computer Communication Review, Vol. 28. No. 4, Proceedings of the ACM/SIGCOMM'98, September 1998, Vancouver, Canada pp. 12–55, 1998

[3] A. Feldmann, A. Gilbert, P. Huang, W. Willinger. *Dynamics of IP traffic: A study of the role of variability and the impact of control* Proceedings of the ACM/SIGCOMM'99, August 29–September 1, 1999, Cambridge, MA

[4] R. Riedi. *Multifractal Processes* Long range dependence: theory and applications, eds. Doukhan, Oppenheim and Taqqu, Birkhauser 2002

[5] K. Drakakis, D. Radulović. *Some results on the multiresolution structure of Internet traffic traces* Proceedings of SPIE Vol. 4868 (2002)

[6] K. Drakakis. *A detailed mathematical study of several aspects of the Internet* Ph.D. Thesis, Princeton University, June 2003