Multifractal Properties of TCP traffic: a numerical study.

Rudolf H. Riedi and **Jacques Lévy Véhel**
Projet Fractales, INRIA Rocquencourt,
B.P. 105, 78153 Le Chesnay Cedex, France
email: Jacques.Levy_Vehel@inria.fr, Rolf.Riedi@inria.fr

Technical Report 3129, INRIA Rocquencourt, March 1997
submitted for publication. to *IEEE/ACM Transactions on Networking*

Abstract

We analyze two traces of TCP–traffic recorded at the gateway of a LAN corresponding to two hours at Berkeley and to eight hours at CNET labs respectively. We are mainly interested in a multifractal approach, which we introduce with various examples. We elaborate on the difference to (mono)fractal statistical tests being used so far. Though we find statistical self-similarity, the data is clearly not monofractal. Consequently, previously proposed models based on fractional Brownian motion are correct only up to second order statistics. Multifractal analysis allows several conclusions which could not be made with the usual estimation of the Hurst exponent. Among these we mention the remarkable fact that incoming and outgoing traffic observed at the same gateway may show strikingly different multifractal behavior. Furthermore, our analysis suggests that heavy and moderate outgoing traffic at Berkeley are in some sense independent of each other. Finally, the outgoing and the incoming traffic at Berkeley and at CNET look entirely different from a multifractal point of view.

1 Introduction

"Fractal" analysis of computer network traffic has recently been the subject of various studies [LTWW, N1, N2, TTW1, TTW2]. Most of the effort has been focused on measuring and modeling a possible *long range dependence* in the data. This was motivated by a thorough experimental study [LTWW] providing strong evidence for the presence of long range dependence in real data traffic. Such a property would have important consequences in many area such as queuing theory [N1, N2] or network design.

However, long range dependence is only one feature of a "fractal" behavior. In this work, we study rather different properties which are conveniently described using *multifractal* *analysis*. Very roughly speaking, while previous studies tried to investigate the *low frequency* content of the signal, we perform here a statistical analysis of the *high frequency* part of it. The motivation for studying these high frequencies is twofold : first, in applications such as traffic control, attempting to understanding *large rapid variations* may prove more crucial than looking for long term correlations. Second, we try to find a possible relation between the local singularities of the signal, revealed by a multifractal analysis, and the findings of previous studies concerning the long range dependence. Indeed, for some simple fractal models such as *fractional Brownian motion* (proposed by several authors for traffic modeling [LTWW, N1, N2]), the exponent governing the local singular behavior is identically equal to the Hurst exponent H ruling the long range dependence. For general processes, however, the local singular behavior depends on time and is best described by the multifractal spectrum of singularity exponents. Our analysis may thus shed new light on the adequacy of previously proposed models by looking at very different properties of them, and help to clarify this rapidly evolving topic.

This paper is organized as follows. Section 2 gives an introduction to the basics of multifractal analysis through a simple example on which the various notions and techniques are explained. Then, the notion of the multifractal spectrum is thoroughly introduced along with the multifractal formalism which is essential in this context. Various pedagogical examples provide insight into the practical side. Section 3 makes precise the multifractal and statistical methods of moments, and explains in detail the more elaborate estimation procedure for the multifractal spectrum. Section 4, finally, presents the numerical analysis performed for two real traffic traces, addressing in particular quality and significance of the results in a comparison of the multifractal and the statistical approach. The multifractal spectra obtained allow to draw conclusions of a general kind.

2 Multifractal Analysis

It is well known that the geometrical complexity of a 'fractal' set may be described, at least in a global way, by giving its dimension. In order to provide more detailed information, multifractal analysis is concerned with describing the local singular behavior of measures, distributions, or functions in a geometrical and statistical fashion. It was first introduced in the context of turbulence [M2, M3, FP], and then studied as a mathematical tool [F2, BMP, HW, AP, O, R2, LV] as well as in many applications such as investigation of DLA pattern [ME, EM2], earth quake distribution analysis [HI], signal processing [L2, L3], and road traffic analysis [VL].

2.1 The Binomial Measure

Purpose and techniques of multifractal analysis are best explained in the most simple situation: the binomial measure on the unit interval.

The binomial measure is a probability measure μ which is defined conveniently via a recursive construction. Start by splitting $I := [0,1]$ into two subintervals I_0 and I_1 of equal length and assign the masses m_0 and $m_1 = 1 - m_0$ to them. With the two subintervals one proceeds in the same manner and so forth: at stage two, e.g. the four subintervals I_{00}, I_{01}, I_{10}, and I_{11} have masses m_0m_0, m_0m_1, m_1m_0, and m_1m_1 respectively. At stage n, the total mass 1 is distributed among the 2^n intervals $I_{\varepsilon_1\ldots\varepsilon_n}$ according to all possible products: $\mu(I_{\varepsilon_1\ldots\varepsilon_n}) = m_{\varepsilon_1} \cdot \ldots \cdot m_{\varepsilon_n}$. This defines a sequences of measures μ_n, all uniform by pieces, which converge weakly towards a probability measure, denoted by μ. By construction, the restrictions of μ to the intervals I_0 and I_1 have the same structure as μ itself. In fact, they are reduced copies of μ where the reductions in space and mass are by $1/2$ and m_i, respectively. In other words, μ is *self-similar* in a very strict way (see Fig. 2.1).

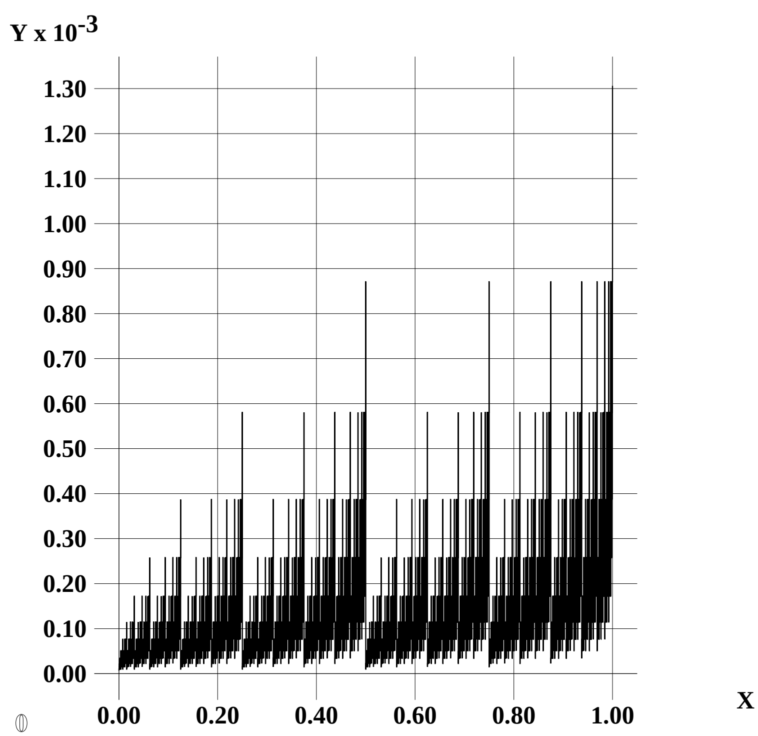


Figure 1: The Binomial measure with $m_0 = .4$ as obtained after 13 iterations.

Another way of defining μ is the following. Let $x = .\sigma_1\sigma_2\ldots$ be the dyadic representation of a point in $[0,1]$. Here, we don't have to care about points with multiple expansions since our results concern 'almost all points x'. Imagine that the digits σ_k are picked randomly such that $P[\sigma_k = i] = m_i$ independently of k. Then, μ is the law—or probability distribution—of the corresponding x on $[0,1]$.

This measure μ clearly has no density, unless $m_0 = m_1 = 1/2$. More precisely, $M(x) = \mu([0,x])$ has zero derivative almost everywhere. Nevertheless, any coarse graining (or sampling) of μ, e.g. through μ_n on the dyadic intervals $I_{\varepsilon_1\ldots\varepsilon_n}$ as above, will naturally result in a distribution with density. It is, therefore, essential to understand the limit behavior of such an approximation μ_n.

Let $I^{(n)}(x)$ denote the unique dyadic interval of order n containing x. Set

$$\alpha_n(x) := \frac{\log \mu(I^{(n)}(x))}{\log |I^{(n)}(x)|} = -\frac{1}{n}\log_2 \mu(I^{(n)}(x)).$$

For the binomial measure as introduced above, the Law of Large Numbers implies that

$$\alpha_n(x) = -\frac{1}{n}\sum_{k=1}^{n} \log_2 m_{\sigma_k} \to \mathbb{E}_\lambda[-\log_2 m_{\sigma_i}] = -\frac{1}{2}\log_2 m_0 m_1,$$

for (Lebesgue) almost all x. From $-\frac{1}{2}\log_2 m_0 m_1 > 1$ we conclude that $M'(x) = 0$ almost surely, as mentioned above. But the LLN implies also that for μ-almost all x

$$\alpha_n(x) \to \mathbb{E}_\mu[-\log_2 m_{\sigma_i}] = -m_0 \log_2(m_0) - m_1 \log_2(m_1). \tag{1}$$

More precise information on limits $\alpha(x) = \lim_{n\to\infty} \alpha_n(x)$ is provided by so-called large deviation theorems [Ell]. Let P_n be the uniform distribution on the set of all dyadic intervals $I_k^{(n)} \subset [0,1]$ of order n, i.e. $P_n[I_{K(\omega)}^{(n)} = I_k^{(n)}] = 1/2^n$ for all $k = 1, \ldots, 2^n$. Denote expectation w.r.t. P_n by $\mathbb{E}_n$ and consider the sequence of random variables $Y_n = \log \mu(I_K^{(n)})$. In order to apply Ellis' theorem [Ell, Thm 2] one has to calculate the asymptotic behavior of its moment generating functions:

$$\mathbb{E}_n[\exp(qY_n)] = 2^{-n}\sum_{k=1}^{2^n} \mu(I_k^{(n)})^q = 2^{-n}\left(m_0^q + m_1^q\right)^n.$$

Since

$$c(q) := \lim_{n\to\infty} \frac{-1}{n}\log_2 \mathbb{E}_n[\exp(qY_n)] = 1 - \log_2\left(m_0^q + m_1^q\right). \tag{2}$$

is a differentiable, concave function, we conclude with Ellis' theorem on Large Deviations that

$$\frac{1}{n}\log_2 P_n\Big[\frac{-1}{n\log 2}Y_n \in (\alpha - \varepsilon, \alpha + \varepsilon)\Big] \to c^*(\alpha) \qquad (n \to \infty, \varepsilon \to 0). \tag{3}$$

Hereby, c^* denotes the Legendre transform, i.e. $c^*(\alpha) = \inf_q(q\alpha - c(q))$. The distribution P_n used here is uniform, i.e. it reduces to counting. Noting that

$$\frac{-1}{n\log 2}Y_n = \frac{\log \mu(I_K^{(n)})}{\log |I_K^{(n)}|} =: \alpha(I_K^{(n)})$$

is in fact the *coarse Hölder exponent* of the dyadic intervals of order n, we may interpret (3) as giving the convergence of properly rescaled histograms of $\alpha(I_k^{(n)})$. This approach to the multiplicative structure of the measure μ and to its singularities has been called *coarse graining* and will be introduced momentarily with all rigor.

Before doing so, we suggest to have a more careful look into the Large Deviation result. Its proof involves a 'change of probability' meaning that the intervals $I_k^{(n)}$ are chosen randomly according to a law μ_q which insures the almost sure convergence of $\alpha(I_k^{(n)})$ towards some value α_q. This distribution μ_q is defined in the same way as μ but with probabilities $\overline{m}_0 := m_0^q 2^\beta$ and $\overline{m}_1 := m_1^q 2^\beta$ where β is such that $\overline{m}_0 + \overline{m}_1 = 1$. Hence,

$$\beta(q) = -\log_2\left(m_0^q + m_1^q\right) = c(q) - 1.$$

Choosing the digits σ_k of the dyadic expansion of a point x such that $P[\sigma_k = i] = m_i^q 2^\beta$ amounts to picking x randomly with law μ_q. Then, μ_q–almost surely

$$\alpha_n(x) \to \mathbb{E}_{\mu_q}[-\log_2(m_{\sigma_i})] = -\sum_{i=0}^{1} \overline{m}_i \log_2 m_i = \beta'(q)$$

by the LLN, whence the claimed almost sure convergence with $\alpha_q := \beta'(q)$.

Moreover, for the same points x we find that

$$\frac{\log \mu_q(I^{(n)}(x))}{\log |I^{(n)}(x)|} = -\frac{1}{n}\log_2 \mu_q(I^{(n)}(x)) \to q\alpha - \beta(q) = \beta^*(\alpha). \tag{4}$$

This result allows us to determine the Hausdorff dimension [F1] of K_α, the set of points x with $\alpha(x) = \alpha$. In fact, when fixing q and letting $\alpha = \beta'(q)$, (4) means that μ_q is equivalent to the $\beta^*(\alpha)$-dimensional Hausdorff measure [F1] restricted to K_α. Since K_α has full μ_q-measure, $\beta^*(\alpha)$ is a lower bound on the dimension of K_α. From the coarse graining approach it follows easily that this bound is in fact sharp [R2].

In summary, we verified that in this simple situation three approaches coincide: one through a 'partition function' c or β, one through 'coarse graining' and one using the concept of 'dimensions'. In a notion which we are about to introduce this reads as

$$f_{\mathrm{L}}(\alpha) = f_{\mathrm{G}}(\alpha) = f_{\mathrm{H}}(\alpha). \tag{5}$$

2.2 Multifractal spectra and the multifractal formalism

We introduce now rigorously what has been motivated in the preceding section.

Much effort has been made in order to obtain rigorous mathematical extensions of the aforementioned result (5) to more general cases [KP, CM, O, F2, AP, LV, R2]. The general setting is as follows.

Assume that a distribution of points in d-space is given in form of a measure μ: the probability for a point to fall in a set E is $\mu(E)$. If this distribution is singular one cannot describe it by means of a density and multifractal analysis proves useful in characterizing the complicated geometrical properties of μ. The basic idea is to classify the singularities of μ by strength. This strength is measured as a singularity exponent $\alpha(x)$, called Hölder exponent. Usually, points of equal strength lie on interwoven fractal sets K_α:

$$K_\alpha := \{x \in \mathbb{R}^d \,:\, \alpha(x) := \lim_{B\to\{x\}} \frac{\log \mu(B)}{\log |B|} = \alpha\}, \tag{6}$$

which explains the name 'multifractal'. Here, $B \to \{x\}$ means that B is a ball containing x, and that its diameter $|B|$ tends to zero. The geometry of the singular distribution μ can then be characterized by giving the 'size' of the sets K_α, more precisely their Hausdorff dimension [F1]:

$$f_{\mathrm{H}}(\alpha) := \dim(K_\alpha).$$

This definition is most useful in purely mathematical settings. It is not required, though, for the understanding of this paper. For the interested reader we refer to [F1, AP, R2, LV] for further details.

In applications, one assumes that μ has bounded support and considers a *coarse grained* version f_{G}, also called *large deviation spectrum*. As a matter of fact it has been introduced prior to f_{H} [M1, M3, Gr1, FP, HP, HJKPS, JKP]:

$$f_{\mathrm{G}}(\alpha) := \lim_{\varepsilon\to 0} \limsup_{\delta\to 0} \frac{\log N_\delta(\alpha,\varepsilon)}{\log 1/\delta}$$

with the convention $\log 0 := -\infty$. Here, N_δ denotes the number of cubes C of size δ with *coarse Hölder exponent* $\alpha(C)$ 'roughly equal to α'. More precisely, denote by G_δ the set of all cubes of the form $C = [l_1\delta, (l_1+1)\delta) \times \ldots \times [l_d\delta, (l_d+1)\delta)$ with integer $l_1, \ldots, l_d$ and with $\mu(C) \neq 0$. Then, we set

$$C^* = [(l_1-1)\delta, (l_1+2)\delta) \times \ldots \times [(l_d-1)\delta, (l_d+2)\delta),$$

$$\alpha(C) := \frac{\log \mu(C^*)}{\log \delta}, \tag{7}$$

and

$$N_\delta(\alpha,\varepsilon) = \#\{C \in G_\delta \,:\, \alpha(C) \in (\alpha-\varepsilon, \alpha+\varepsilon]\}.$$

As is pointed out in [R2, PR] using C^* instead of C greatly improves the theoretical properties as well as the numerical behavior of f_{G} since C provides a poor approximation of a ball centered in a point of the distribution μ, especially in points close to the boarder of the support of μ. Since singular measures are typically supported on fractals, these problems are present on all scales leading to wrong results. One of the advantages when using C^* is the fact that the spectrum $f_{\mathrm{G}}(\alpha)$ does not change when replacing the continuous limit $\delta \to 0$ by the discrete limit $\delta_n = c2^{-n}$ $(n \to \infty)$. Furthermore, for measures supported on an interval of length c this choice δ_n provokes no 'boarder effects' and C^* may again be replaced by C without changing the outcome. These properties have been used throughout our numerical analysis. For the ease of notation we will write $N_n^\varepsilon(\alpha) := N_{\delta_n}(\alpha, \varepsilon)$.

Though tempting it is *wrong* to interpret $f_{\rm G}$ as the box dimension of K_α. This function is better explained in statistical terms: Note first that the number N_δ of cubes in G_δ behaves roughly as $N_\delta \simeq \delta^{-D}$ where D denotes the box dimension of the support of μ. It follows that $f_{\rm G}(\alpha) \leq D$ for all α. Now suppose that one picks a cube C out of G_δ randomly and determines its coarse Hölder exponent $\alpha(C) := \log \mu(C^*)/\log \delta$. Then, the probability of finding $\alpha(C) \simeq \alpha$ behaves roughly like

$$N_\delta(\alpha, \varepsilon)/N_\delta = P_\delta[\alpha(C) \simeq \alpha] \simeq \delta^{D - f_G(\alpha)}. \tag{8}$$

Note in particular that in the limit $\delta \to 0$ the only Hölder exponent which is observed with non-vanishing probability is α_0, where $f_{\rm G}(\alpha_0) = D$.

What spectra $f_{\rm H}$ and $f_{\rm G}$ can we expect?[1] Could they be trivial functions ($f(\alpha) = -\infty$)? Before giving a proof (9) that this is not the case let us explain in simple terms why there is at least α_0 with $f_{\rm G}(\alpha_0) = D$. We give an argument for 'self-similar measures' with $D = 1$ using the Law of Large Numbers (LLN). Write

$$\alpha(x) = \lim_{n\to\infty} -\frac{1}{n}\log_2 \mu(C_n^*(x)) = \lim_{n\to\infty} -\frac{1}{n}\sum_{k=1}^{n} \log_2 \frac{\mu(C_k^*(x))}{\mu(C_{k-1}^*(x))}$$

where $C_n(x)$ is the unique cube in $G_{1/2^n}$ containing x. Then, the assumption of self-similarity means that the random variables $\log_2 \mu(C_k^*(x))/\mu(C_{k-1}^*(x))$ are i.i.d. (compare Subsection 2.1 and [R2, AP]). Denote the common expectation by α_0. The LLN implies that almost surely $\alpha(x) = \alpha_0$ when picking points x randomly with 'uniform' distribution, i.e. when picking C randomly in G_δ. This establishes the claim.

In special cases such as the binomial measure with $m_0 = m_1 = 1/2$ (uniform distribution) α_0 is the only Hölder exponent, i.e. $\alpha(x) = \alpha_0 = 1$, $f_{\rm H}(\alpha_0) = f_{\rm G}(\alpha_0) = 1$ and $f_{\rm H}(\alpha) = f_{\rm G}(\alpha) = -\infty$ for $\alpha \neq \alpha_0$ in this case. Such measures with only one Hölder exponent are called uniform or *monofractal*.

In general, other Hölder exponents occur. For the binomial, e.g. we find $\alpha(0) = -\log_2(m_0)$ $\alpha(1) = -\log_2(m_1)$ etc. Also, the coarse graining will show non-trivial spectra, i.e. on every finite level of approximation G_δ one will have a whole histogram of *coarse Hölder exponents* $\alpha(I_k^{(n)})$. For $\alpha \neq \alpha_0$, however, the probability of finding $\alpha(I_k^{(n)}) \simeq \alpha$ will decrease exponentially fast to 0 as $\delta \to 0$. A rigorous proof of this fact is most easily obtained — at least under certain conditions — by applying the Principle of Large Deviations (PLD) of Gärtner-Ellis (see [Ell]). Translated into our setting the PLD states, in simple terms, that

$$P_n[\frac{1}{n}\log_2 \mu(C_n^*(x)) \simeq \alpha] \simeq 2^{nc^*(\alpha)}$$

with some scaling function c^* with $c^*(\alpha) < 0$ unless $\alpha = \alpha_0$. A rigorous formulation is the following:

[1] Some answers of a general kind can be found in [LV, LT].

Theorem 1 ([Ell, R2]) *Assume that the 'moment generating function'*

$$c(q) := \lim_{n\to\infty} \frac{-1}{n} \log_2 \mathbb{E}[\exp(q \log \mu(C_n^*(x)))]$$

exists and is convex and differentiable for all $q \in \mathbb{R}$. Then,

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n \log 2} P_n\Big[\Big| -\frac{1}{n} \log_2 \mu(C_n^*(x)) - \alpha\Big| \le \varepsilon\Big] = c^*(\alpha)$$

where $c^*(\alpha) = \inf_q (q\alpha - c(q))$ is the Legendre transform of c.

The *partition function* $\tau(q)$ [M3, FP, HP, HJKPS, R2, LV]

$$\tau(q) := \lim_{\delta\to 0} \frac{\log S_\delta(q)}{\log \delta} \qquad \text{with} \qquad S_\delta(q) := \sum_{C \in G_\delta} \mu(C^*)^q$$

equals c up to a constant. Indeed, since $D = -\tau(0)$ by definition, we find

$$c(q) = \tau(q) - \tau(0) = \tau(q) + D.$$

For the binomial measure defined in Subsection 2.1 one finds with (2)

$$\tau(q) = -\log_2(m_0^q + m_1^q).$$

For other examples see Subsection 2.3 below.

Provided that Ellis' theorem applies, i.e. assuming that $\tau(q)$ exists and is differentiable, it follows that (8) holds with $c^* = f_G(\alpha) - D$, i.e.

$$f_G(\alpha) = \tau^*(\alpha). \tag{9}$$

This has been termed the *multifractal formalism*. The similarity to the well-known thermo-dynamical formalism [V, R3] is immediate.

Since $\tau(q)$ is obtained by averaging, it depends more regularly on the data than $f_G(\alpha)$ and is easier to compute. It is important to note, though, that it contains in general less information than f_G. Let us make this point more precise.

The partition function is always convex since $S_\delta(q)$ is convex for all δ. But it is not necessarily differentiable in every q and the multifractal formalism may not hold for all α. For some simple and convincing counterexamples see Fig. 4 and [LV, R2, MR, RM2, RM3]. It is natural, thus, to introduce the *Legendre spectrum*

$$f_L(\alpha) := \tau^*(\alpha).$$

This spectrum is sometimes referred to as obtained by the *method of moments*.

While (9) may be wrong for certain α, the opposite relation

$$\tau(q) = f^*_{\mathrm{G}}(q) = \inf_{\alpha \in \mathbb{R}} \left(q\alpha - f_{\mathrm{G}}(\alpha) \right)$$

holds for all q [R2].

As a first consequence, D is indeed the maximal value of f_{G} in general. Secondly, $f_{\mathrm{L}} = \tau^* = {f_{\mathrm{G}}}^{**}$ is the concave hull of f_{G}. Thus,

$$f_{\mathrm{G}}(\alpha) \leq f_{\mathrm{L}}(\alpha).$$

Thirdly, it follows that even a not everywhere differentiable $\tau(q)$ determines $f_{\mathrm{G}}(\alpha)$ at least in its concave points. To be more precise let $\alpha^+ := \tau'(q+)$ for $q > 0$ and $\alpha^- := \tau'(q-)$ for $q < 0$ denote the one-sided derivatives of $\tau(q)$ which must exist since $\tau(q)$ is convex. Then [R3, R2],

$$\begin{array}{rclr} f_{\mathrm{G}}(\alpha^+) & = & q\alpha^+ - \tau(q) = \tau^*(\alpha^+) & (q > 0) \\ f_{\mathrm{G}}(\alpha^-) & = & q\alpha^- - \tau(q) = \tau^*(\alpha^-) & (q < 0). \end{array} \tag{10}$$

An alternative way of displaying the scaling of moments is through the so-called *generalized dimensions*

$$D(q) := \frac{\tau(q)}{q-1}.$$

Besides $D = D_0$, a notable D_q is $\alpha_1 = \tau'(1) = D_1$. It has been termed information dimension [Gr2, GP1, GP2, OWY]: With respect to the given distribution μ we have $\alpha(x) = \alpha_1 = D_1$ almost surely. For a binomial measure α_1 is given by (1).

2.3 Reading a multifractal spectrum

Before listing some properties of the shape of a multifractal spectrum we give an intuitive interpretation in loose terms.

The multifractal spectra provide a global description of the singularities of the observed measure μ. The parameter α quantifies the degree of regularity in a point x: loosely speaking, the measure of an interval $[x, x + \Delta x]$ — in applications usually the number of events occurring in this interval — behaves as $(\Delta x)^\alpha$ (6,7).

For a uniform distribution one finds $\alpha(x) = 1$ for all x. More generally, for any $a > 0$ the distribution with density x^{a-1} on $[0, 1]$ has $\alpha(0) = a$ and $\alpha(x) = 1$ for all $x \in (0, 1]$. Values $\alpha(x) < 1$ indicate, thus, a burst of events around x 'on all levels' (bursts of bursts), while $\alpha(x) > 1$ is found in regions where events occur sparsely.

The spectrum $f_{\mathrm{G}}(\alpha)$ captures how 'frequently' a value $\alpha(x) = \alpha$ is found: the number of dyadic intervals C of size Δx with $\alpha(C) = \alpha$ behaves as $(\Delta x)^{-f_{\mathrm{G}}(\alpha)}$. For the 'almost sure' value α_0 one has $f_{\mathrm{G}}(\alpha_0) = D$ which is necessarily the maximal value of f_{G}. A more precise characterization is given by (8).

The spectrum $f_{\mathrm{H}}(\alpha)$ gives the size of the 'set of Hölder exponent α' K_α in terms of its dimension.

In order to give an idea of how to extract information from the shape of a spectrum f we proceed by giving examples for which the spectra are known explicitly. The plots are obtained by solving implicit equations for $\tau(q)$ similar as the one for the Binomial measure of Subsection 2.1:

$$m_0^q 2^\tau + m_1^q 2^\tau = 1,$$

and plotting its Legendre transform by varying the parameter q:

$$\alpha(q) = \tau'(q) \qquad\qquad f_{\mathrm{L}}(\alpha) = \tau^*(\alpha) = q\tau'(q) - \tau(q).$$

The shape of a 'typical' spectrum f includes various features. First of all, the Legendre spectrum f_{L} possesses an overall shape like the symbol $\cap$. This translates to f_{G} *provided* the multifractal formalism holds. Next, we have seen that $f_{\mathrm{G}}(\alpha) \leq \tau(0) = D = \sup_\alpha f_{\mathrm{G}}(\alpha)$. Finally, since we are considering measures, we have necessarily $S_\delta(1) = 1$ for all δ. This implies that $\tau(1) = 0$. In other words, f_{L} touches the internal bisector of the axis, whence we obtain a further bound of interest: $f_{\mathrm{G}}(\alpha) \leq f_{\mathrm{L}}(\alpha) \leq \alpha$ (see Fig. 2)..

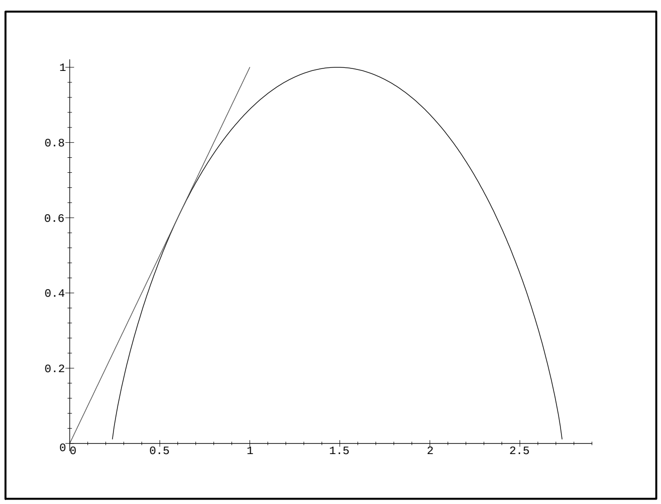


Figure 2: A typical spectrum will be strictly concave, it will touch the bisector (dashed) and it will reach the maximum D, the dimension of the support of the measure. Here, we display the spectrum of the binomial measure with $m_0 = .15$, $m_1 = .85$, and $r_0 = r_1 = 1/2$. It is supported on the unit interval which has dimension $D = -\tau(0) = 1$. The extremal Hölder exponents are $\alpha_{\min} = \log(.85)/\log(.5) \simeq .234$ and $\alpha_{\max} = \log(.15)/\log(.5) \simeq 2.737$.

Whenever a spectrum f fails to show the typical concave $\cap$-shape we have evidence that μ is not purely multiplicative. A search of models with similar features in their spectra may reveal telling details on the structure of the distribution μ. A most prominent example is DLA where the spectra estimated numerically showed clear evidence for an infinite range $[\alpha_{\min}, \infty]$ of Hölder exponents with a spectrum f increasing over this whole interval. Searching for multifractal measures displaying similar spectra Mandelbrot et al.

were led to suggest that the underlying self-similarity of DLA was in fact of an infinite type [M4, EM2, ME, MEH, RM1].

In the sequel we provide various examples with atypical spectra and explain the particular appearances. Let us start by generalizing step by step the construction of the binomial measure from Subsection 2.1, with the multifractal formalism remaining valid all through:

1. Take arbitrary contraction ratios r_0 and r_1 instead of $1/2$. More precisely, for each interval $I_{\varepsilon_1\ldots\varepsilon_n}$ in the iterative construction choose two subintervals $I_{\varepsilon_1\ldots\varepsilon_n k}$ with disjoint interiors such that the ratio of lengths is $r_k = |I_{\varepsilon_1\ldots\varepsilon_n k}|/|I_{\varepsilon_1\ldots\varepsilon_n}|$. Note that necessarily $r_0 + r_1 \leq 1$. The formula for $\tau(q)$ generalizes then to:
$$m_0^q {r_0}^{-\tau} + m_1^q {r_1}^{-\tau} = 1.$$
In particular, the support of the measure obtained in this way is a Cantor set of dimension $D = -\tau(0)$, which is in general strictly less than 1. As a matter of fact, $D = 1$ iff $r_0 + r_1 = 1$. In this paper, we consider only this case, i.e. measures supported on a whole interval (see Fig. 2).

2. Take a fixed, but arbitrary number N of disjoint subintervals in each step, with length ratios $r_0 + \ldots + r_{N-1} \leq 1$ and mass ratios $m_0 + \ldots + m_{N-1} = 1$. Then, $\tau(q)$ is uniquely determined by
$$\sum_{i=0}^{N-1} m_i^q {r_i}^{-\tau} = 1.$$
Such a measure is called *self-similar*.

3. Fix the distributions of $2N$ random variables R_i and M_i $(i = 0, \ldots, N-1)$ such that $\sum_{i=0}^{N-1} R_i \leq 1$ almost surely and $\mathbb{E} \sum_{i=0}^{N-1} M_i = 1$. Then, in each step of the construction for each of the intervals $I_{\varepsilon_1\ldots\varepsilon_n}$ choose disjoint subintervals $I_{\varepsilon_1\ldots\varepsilon_n k}$ randomly in the following way: When fixing k and letting $\varepsilon_1 \ldots \varepsilon_n$ run through all finite sequences of digits $\{0, \ldots, N-1\}$, the random variables
$$R_{\varepsilon_1\ldots\varepsilon_n k} := |I_{\varepsilon_1\ldots\varepsilon_n k}|/I_{\varepsilon_1\ldots\varepsilon_n}$$
are i.i.d. with same distribution as R_k, and the mass ratios $\mu_{n+1}(I_{\varepsilon_1\ldots\varepsilon_n k})/\mu_n(I_{\varepsilon_1\ldots\varepsilon_n})$ are i.i.d. with law M_k. This way of redistributing mass independently and identically in each step of the construction could fairly be called *multifractal stationarity*. As a sum of i.i.d. random variables, $\log \mu(I_{\varepsilon_1\ldots\varepsilon_n})$ is approximately Gaussian, if properly normalized. There is evidence that log-normal marginal distributions, though not heavy-tailed, may well characterize real data traffic [P].

   Under the given conditions, there is a unique τ such that
$$\mathbb{E} \sum_i M_i^q R_i^{-\tau} = 1,$$
and $\tau(q)$ is differentiable and convex on $\mathbb{R}$ [F2, AP]. Its Legendre transform $f_{\mathrm{L}}(\alpha) = \tau^*(\alpha)$ equals the dimension of the set K_α as well as the large deviation spectrum

almost surely where it is positive, i.e. if $f_{\rm L}(\alpha) \geq 0$ then $f_{\rm H}(\alpha) = f_{\rm G}(\alpha) = f_{\rm L}(\alpha)$ with probability one.

For the Hölder exponents α with $f_{\rm L}(\alpha) < 0$ no rigorous results are available in general. These values find a natural interpretation, though, as so-called **negative dimensions** [M5]: the probability of observing α decreases too fast with the grid size δ, i.e. approximately as $\delta^{1-f_L(\alpha)}$ (compare (3)). These Hölder exponents are, thus, only observed when oversampling the process at least $\delta^{f_L(\alpha)}$ times. Then, at least in the average, one of the $\delta^{f_L(\alpha)-1}$ intervals of size δ now available should show the Hölder exponent α.

Such a *self-similar random measure* is shown in Fig. 3 where it is also compared with a trace of real data traffic. The striking resemblance makes multifractals a natural candidate for data traffic modeling. This point of view is further supported by a thorough analysis in Section 4.

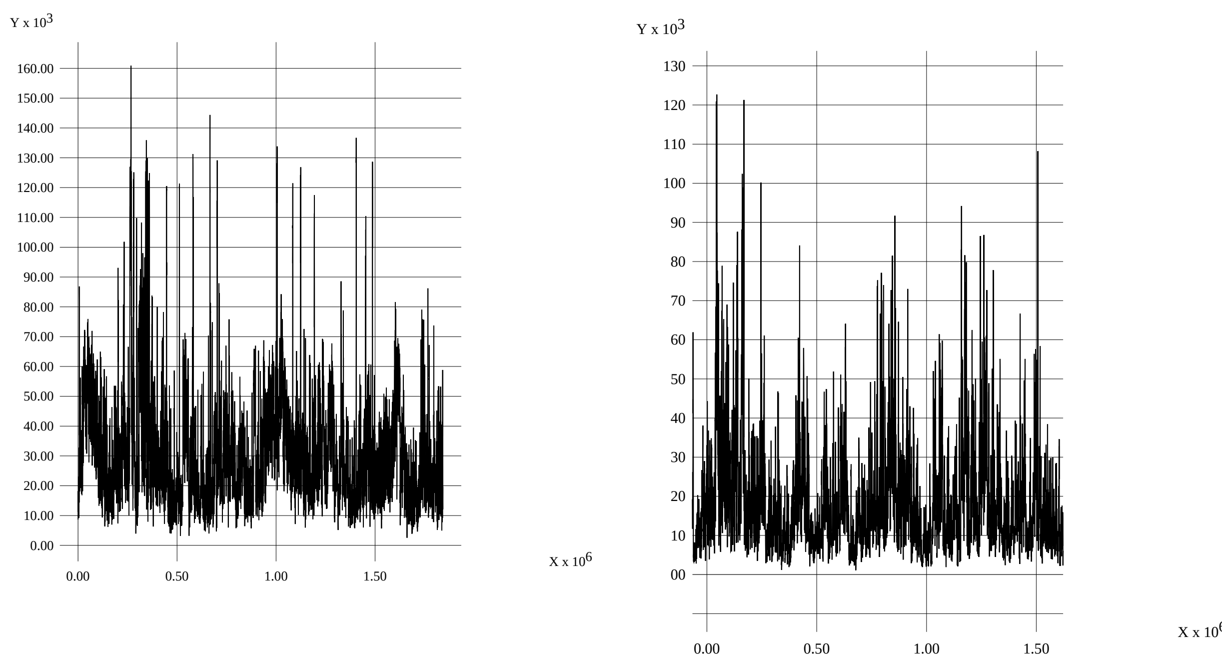


Figure 3: Traffic observed at the gateway of Berkeley and the measure obtained by a randomized version of the Binomial measure.

Let us now give examples of measures with atypical spectrum. The easiest way of breaking the property of being concave is by considering sums of binomial measures $\mu = \mu_1 + \mu_2$. The spectra are then known in two extreme cases.

In the first case, called 'lumping' of measures, the supports of μ_1 and μ_2 are disjoint. Then the formula

$$f(\alpha) = \max(f^1(\alpha), f^2(\alpha))$$

is valid for both, $f_{\rm G}$ and $f_{\rm H}$. For $f_{\rm L}$, on the other hand, it cannot hold in general, since the maximum of concave functions is in general not concave. We have, though, that $\tau(q) = \max(\tau^1(q), \tau^2(q))$. Provided ${f_{\rm G}}^k = {f_{\rm L}}^k$, $f_{\rm L}$ is the concave hull of $f_{\rm G}$. Such is the case for self-similar measures (see Fig. 4 and Fig. 6).

The other extreme case is the 'superposition' of two measures which means that the geometrical part of the construction of two self-similar measures μ_1 and μ_2 are identical and the only difference between them lies in the choice of the weights m_i. Then, a spectrum as in Fig. 5 results.

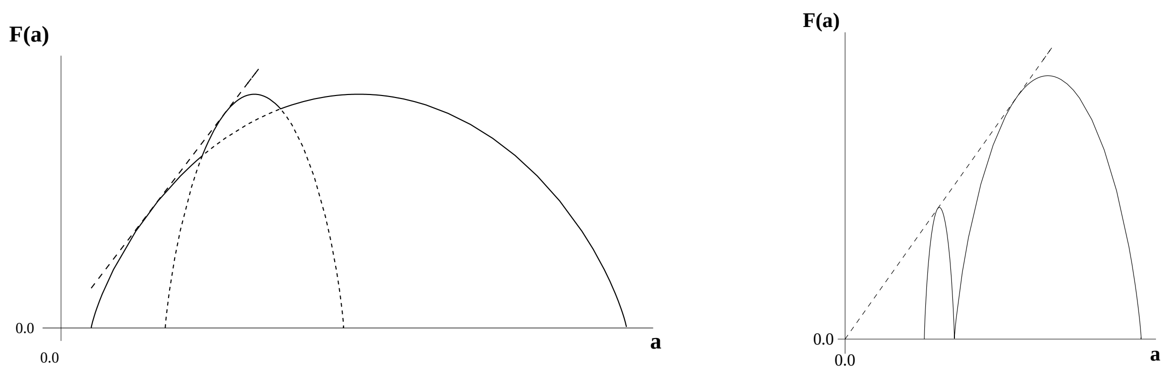


Figure 4: The spectrum of the lumping $\mu = \mu_1 + \mu_2$ of two measures (meaning that their supports are disjoint) is simply the maximum of the individual spectra. This will in general result in a non-concave spectrum as shown in two cases here. The dashed parts show the internal bisector of the axes and the spectra of the binomial measures μ_1 and μ_2 where they do not coincide with the spectrum of μ.

The failure of being concave is here a direct consequence of a sort of **phase transition**: The major contributor to the singularities of strength α changes from one of the measures μ_1 and μ_2 to the other.

Similar phenomena have been observed with the somewhat richer class of self-affine measures μ in the plane [R1]. Here, $\tau(q) = \max(\Gamma_1(q), \Gamma_2(q))$ where the $\Gamma_k(q)$ represent the contributions of the two eigenspaces to the singularities of μ. The phase transition is here between the two eigenspaces as the major contributor.

In conclusion, we take non-concave spectra as evidence for the absence of a 'universal' multiplicative law, and not much can be said in general. If the spectrum consists of clearly identified bumps, though, we may assume the presence of several such laws.

3 Numerical methods

We are given a sequence of positive numbers $(Z_i)_{i=1}^N$ which may represent any interesting information on the traffic load passing through a gateway. (In Section 4 we will reserve this letter for one particular data). Among many other ways of analyzing we may interpret the data as a sampling a) of a random measure where $Z_i/\sum_i Z_i$ is the probability for a byte to arrive at time t_i, or b) of a stochastic process.

In the former case, a multifractal analysis is in order. In the latter case, it has been widely agreed [LTWW, N1, N2] that the most important fractal statistical parameter to be estimated is the degree of *long range dependence* (LRD), usually measured through the *Hurst exponent* H. Among the various methods of estimating H [TTW1, TTW2] the method of moments comes closest to the multifractal approach.

In the first part of this section we make precise the notion of multifractal and statistical scaling of moments. We will argue that the multifractal method is more natural when positive data such as traffic data is given since there is no need to center the data before analyzing. A more sophisticated comparison of the two approaches is postponed to Subsection 4.3 where the necessary data analysis is available.

In the second part we will elaborate on the more sophisticated estimation of the large deviation spectrum f_{G}. We will show, in particular, how to reduce the double limit to a single one. Recall that f_{G} provides more detailed information than the method of moments, i.e. the Legendre spectrum f_{L}.

3.1 Multifractal scaling

As was explained in more detail in Section 2, the most straightforward approach to measuring local singularity exponents of a measure is through the partition function $\tau(q)$. One considers then the data $(Z_i)_{i=1}^N$ as a sampling of a measure μ on $[0,1]$ at scale $\delta = 1/N$ and defines the *partition sum* through

$$S_m^Z(q) := \sum_{k=1}^{N/m} (\overline{Z}_k^{(m)})^q,$$

where

$$\overline{Z}_k^{(m)} := \sum_{l=1}^{m} Z_{(k-1)m+l}$$

provides a sampling of μ at scale $\delta_m = m/N$.

If $\log S_m^Z(q)$ is in good approximation linearly depending on $\log m$, we say that the data exhibits *multifractal scaling*, in short: Z_i is a multifractal. The slope of the linear law, usually obtained by least square fitting, is denoted by $\tau^Z(q)$, or shortly $\tau(q)$:

$$\log S_m^Z(q) \simeq \tau^Z(q) \cdot \log m + \text{const.} \tag{11}$$

In order to visualize the quality of a linear approximation of the graph of $\log_2 S_m^Z(q)$ versus $\log_2 m$ it is useful to look at the piecewise increments of $\log_2 S_m^Z(q)$, i.e.

$$\tau_m^Z(q) := \tau(Z,m,q) := \log_2 S_{2m}^Z(q) - \log_2 S_m^Z(q) \tag{12}$$

as a function of $\log_2 m$. If (11) holds in good approximation then $\tau_m(q) \simeq \tau(q)$ independently of m.

In praxis, one computes $\tau(q)$ through a least square fitting rather than through averaging $\tau_m(q)$. The behavior of the latter, though, can be used to determine the *scaling region*, i.e. the range of m in which the fitting is performed.

Furthermore, as the slope $\tau(q)$ varies often only little, typically in the range $[1/2, 2]$, its plot may appear to be almost linear to the naked eye. Therefore, displaying its Legendre transform f_{L} is generally more informative. Note, that f_{L} may show **negative** values. These correspond to very rare Hölder exponents as explained in Subsection 2.3.

3.2 Statistical scaling

The statistical approach we refer to relies on the notion of *self-similar processes*. Consider a process Y_i with stationary increments $X_i = Y_i - Y_{i-1}$. Assume that Y_i is H-self-similar, i.e.

$$Y_{mi} \stackrel{d}{=} m^H Y_i.$$

Let

$$X_k^{(m)} = 1/m \sum_{l=1}^{m} X_{(k-1)m+l} = 1/m \overline{X}_k^{(m)}.$$

Then,

$$X_k \stackrel{d}{=} m^{1-H} X_k^{(m)}. \tag{13}$$

As was proposed by Taqqu, Teverovsky & Willinger recently [TTW2], a test of self-similarity could be performed through the behavior of the absolute moments. An estimator of $\mathbb{E}|X^{(m)}|^q$ would be

$$\hat{\mathbb{E}}|X^{(m)}|^q := \frac{1}{N/m} \sum_{k=1}^{N/m} |X_k^{(m)}|^q.$$

In analogy to multifractal analysis we consider rather

$$S_m^X(q) := \sum_{k=1}^{N/m} |\overline{X}_k^{(m)}|^q = m^{q-1} N \cdot \hat{\mathbb{E}}|X^{(m)}|^q.$$

If X_i is self-similar in the sense of (13) we have

$$\log \mathbb{E}|X^{(m)}|^q = \log \mathbb{E}|m^{H-1}X|^q = b(q)\log m + \text{const.}$$

Moreover, $b(q)$ is then linear in q

$$b(q) = q(H-1).$$

Thus, Taqqu et al. [TTW2] propose a test of self-similarity which we translate into our setting: Determine whether there is γ such that

$$\log S_m^X(q) \simeq \gamma(q) \cdot \log m + \text{const.} \tag{14}$$

If γ depends linearly in q then the data may be called self-similar in the sense of (13), and

$$\gamma(q) = q - 1 + b(q) = qH - 1.$$

It is worth noting, first of all, that (13) implies together with stationarity that

$$\text{either} \qquad \mathbb{E}X = 0, \qquad \text{or} \qquad \mathbb{E}X = \pm\infty, \qquad \text{or} \qquad H = 1.$$

But $H = 1$ implies that $Y_t = t \cdot Y_1$ almost surely. In other words, the concept of statistical self-similarity (14) makes sense only after centering the data, i.e. in our context when setting

$$X_i := Z_i - \hat{\mathbb{E}}Z = Z_i - 1/N \sum_{i=1}^{N} Z_i.$$

Multifractal analysis as presented here, on the other hand, is perfectly fitted to positive data such as data traffic measurements.

Further comparison of the multifractal and the statistical approach is postponed to Subsection 4.3. Here, we would like to add that a multifractal analysis of processes with arbitrary increments is being developed using wavelets [J1, J2]. The application of this approach, though, is beyond the scope of this paper and has to be presented in forthcoming publications [RL].

3.3 Estimation of the large deviation spectrum f_{G}.

As mentioned before, the most attractive multifractal spectrum from numerical point of view is the one obtained through the Legendre transform

$$f_{\mathrm{L}}(\alpha) := \tau^*(\alpha).$$

In cases where the multifractal formalism holds, f_{L} provides as much information as f_{G}. In general, though, f_{G} provides more detailed information (see 4.1.1). The most prominent situation in which the formalism has been established are the so-called self-similar measures which generalize the construction of the Binomial measure. As has been pointed out before, we have

$$f_{\mathrm{G}}(\alpha) = f_{\mathrm{L}}(\alpha) = \tau^*(\alpha) \qquad \text{at } \alpha = \tau'(q),$$

provided this derivate exists. Such situations are indeed encountered with many multiplicative measures such as the self-similar and the self-affine ones.

In a concrete application, one will compute $\tau(q)$ as the slope of a least square fitting of a log-log plot of $S_\delta^Z(q)$ against δ. It is essential to check the quality of this fitting. If one does find a nearly linear behavior we say that we have **multifractal scaling**. It is then legitimate to assume that one properly estimated $\tau(q)$.

When there are reasons to believe that f_{G} is not concave, it becomes necessary to design a specific estimation procedure [L1]. Mathematical examples with such a behavior include the lumping and superposition of binomial measures (compare Subsection 2.3) as well as the discontinuous self-similar measures [MR, RM2]. As is shown later, certain aspects of data traffic provide wonderful real world examples (see 4.1.1).

The procedure is based on a classical tool in density estimation: the kernel method, or more precisely the double kernel method ([D]). Such a method has been used extensively with good success in density estimations, and rather precise theorems are known that assess the quality of the results. In particular, the use of smooth kernels allows to obtain more regular estimates. The difficulty we meet here is that f_{G} is not the density corresponding to the α's, but rather a double logarithmic normalization of this density.

At the starting point of the double kernel method stands the observation that $N_n^\varepsilon(\alpha)$ can be re-written as the following convolution: $N_n^\varepsilon(\alpha) = 2^{n+1}\varepsilon K_\varepsilon * \rho_n(\alpha)$, where ρ_n is the density of the coarse Hölder exponents $\alpha(C)$ ($C \in G_{1/2^n}$) as defined in (7), and K_ε is the rectangular kernel: $K(x) = 1$ for $x \in [-1/2, 1/2]$, $K(x) = 0$ else, and $K_\varepsilon(x) = 1/\varepsilon K(x/\varepsilon)$. It is easily shown that in fact any choice of a compactly supported kernel K leads to the same result.

A classical problem in density estimation is the choice of an ε such that $K_\varepsilon * \rho_n$ is as close as possible to the limiting density ρ. In this setting, ε becomes a function of n. Adapting classical techniques allows to design an algorithm which provides an optimal sequence ε_n. In particular, we have ${f_{\mathrm{G}}}^{\epsilon_n}(\alpha) = (1/n)\log N_n^{\varepsilon_n}(\alpha) \to f_{\mathrm{G}}(\alpha)$ assuming that such a sequence exists.

Such is the case in the following context. Assume that the distribution μ is a finite superposition or lumping of binomial measures. Let $f_n^{\varepsilon_n}$ correspond to the rectangular kernel and $g_n^{\varepsilon_n}$ correspond to the triangular kernel ($T(x) = 1 - x$ for $x \in [0, 1], T(x) = 0$ for $x > 1, T(-x) = T(x)$). [L1] shows that if $(\varepsilon_n, \varepsilon_n')$ minimizes $\sup_\alpha(f_n^{\varepsilon_n}(\alpha) - g_n^{\varepsilon_n'}(\alpha))$, then $f_n^{\varepsilon_n}(\alpha) \to f(\alpha)$ for all α.

Numerical experiments using this technique are displayed in Fig. 5 and 6 where the theoretical f_G along with its kernel estimate are shown for a superposition and a lumping of two binomial measures.

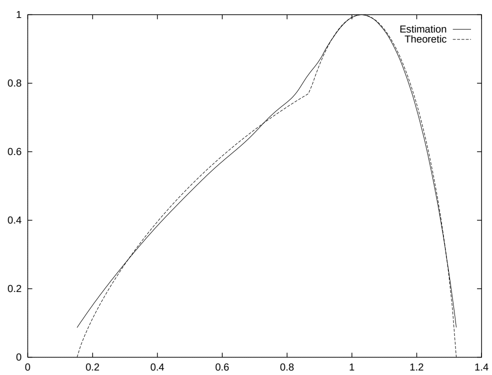


Figure 5: Theoretical (dashed) and estimated (solid) f_G spectrum for the superposition of two binomial measures.

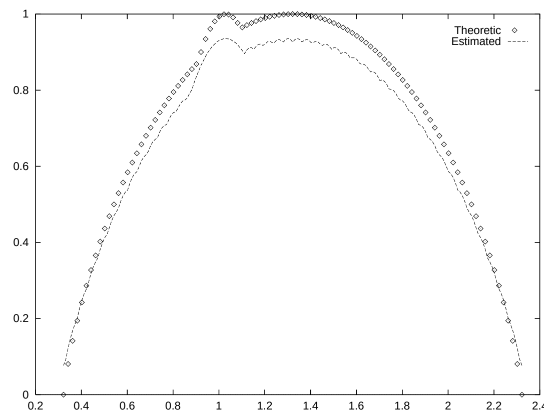


Figure 6: Theoretical (dashed) and estimated (solid) f_G spectrum for the lumping of two binomial measures.

4 Numerical analysis of the TCP traffic

The two traces analyzed in this study are TCP–data traffic collected at the gateway of a LAN, corresponding to two hours at Berkeley and to eight hours at CNET labs, respectively. Both traces contain about 2^{21} observations, consisting of the arrival time T_i of the packet number i, the size Z_i of the packet in number of bytes, as well as information regarding sender and recipient.

This allowed to distinguish between the traffic entering the LAN, called *incoming*, and the traffic leaving the LAN, called *outgoing*. We would like to stress that the incoming traffic is, therefore, generated by a WAN traffic, thus (WAN-LAN), while the outgoing one is produced by a LAN traffic, thus (LAN-WAN). Both traffic streams, however, have to pass through the same gateway, producing the measured traffic load Z_i. This is what we mean by 'traffic'. When confusion is possible we will call it *combined traffic* in the sequel.

4.1 Trace recorded at Berkeley

In the sequel we will address the various parts of the recorded data, i.e. the 'number of bytes of packet' Z_i, the 'inter-arrival time between packets' T_i, the 'number of bytes arriving per time' B_k, and 'number of packets arriving per time' P_k, as *aspects* of the traffic.

4.1.1 Bytes per packet

In a first attempt we investigated Z_i,i.e. the aspect 'bytes per packet'. Not only is this aspect the most easily accessible, it poses also no problems when separating incoming and outgoing traffic.

At first sight, one would not expect an interesting behavior, since TCP packets are mostly either very small (ACK, NACK) or very large. This reasoning, however, considers only the histogram at the finest time scale. As we are about to show, the aspect Z_i reveals essential characteristics of the traffic. In particular, though there is only little variability in the data Z_i, there is excellent multifractal scaling, meaning that there are clusters of very large (small) packets, clusters of clusters, and so on. In other words, there is no 'averaging'. To the contrary the multifractal properties of Z_i allow to clearly distinguish between outgoing and incoming traffic and demonstrate the striking difference between traffic generated at the University of Berkeley versus the one produces at the research laboratories of CNET.

Let us report on the multifractal scaling first. We found excellent multifractal scaling of $S_m^Z(q)$ for all q when choosing $m = 25, 50, 100, \ldots, 1600$ as well as when taking $m = 1, 2, 4, \ldots, 2^{19}$ (see Fig. 7). The linear behavior is of exceptional quality especially in the scaling region $m = 2^7, \ldots, 2^{19}$ (see Fig. 8). Also the collapse of the estimators ${f_{\mathrm{G}}}^m$ obtained at various levels of aggregation m is very satisfactory (see Fig. 9).

The $\tau(q)$ obtained from a least square fitting of $\log S_m^Z(q)$ against $\log m$ yields a narrow multifractal spectrum f_{L}. Nevertheless, we find $\tau'(q)$ to fill the range [.9, 1.2] and a Legendre spectrum $f_{\mathrm{L}} = \tau^*$ in excellent agreement with the f_{G} obtained by coarse graining (see Fig. 9).

The most striking feature of the multifractal spectrum f_G is its departure from a pure bell-shape: it shows a bump or a 'non-concavity' (see Fig. 9) and is a perfect example where coarse graining (f_{G}) provides more information than the method of moments (f_{L}). This phenomenon found a most satisfying explanation when incoming and outgoing traffic were separated. While the former shows a clear bell-shaped spectrum, we find two 'bells' for the latter (see Fig. 10 and also Fig. 16). A most simple interpretation would be that there are two multiplicative laws underlying the outgoing traffic. In other words, at Berkeley, heavy outgoing traffic seems to be widely independent of the moderate outgoing traffic.

With the statistical approach the scaling behavior is less favorable. For the centered data $X_i := Z_i - \hat{\mathbb{E}}Z$ the scaling behavior of the partition sum $S_m^X(q)$ was acceptable only for *positive* q (see Fig. 11). For negative q the log–log plots were far from being linear, obviously due to very small values $|X_k^{(m)}| << 1$ produced when centering the data. From the point of view of scaling, thus, it can be regarded as a disadvantage having to center the data in the 'statistical scaling approach'.

But the estimates $\gamma_m^X(q)$ suffer from high variability even in the positive q range. The estimation of H in particular depends highly on the choice of the scaling region and cannot be called robust (Subsection 4.3). Nevertheless, the scaling function $\gamma(q)$ we obtained is fairly linear for $q \geq 1$. This linearity corresponds to a Legendre transform γ^* which 'stops' at the point with slope $q = 1$ (see Fig. 12). The decreasing part, though displayed, is not reliable since there was poor scaling for negative q (see Fig. 11). Most importantly, with the centered data X_i no significant difference could be observed between incoming and outgoing traffic, i.e. the reliable increasing part of the γ^*-spectra are practically identical (see Fig. 12). We note on the side that the f_{G} spectra of X do not show such a clear difference either.

4.1.2 Time per packet

The next simple aspect from the point of view of an analysis is 'time per packet'. We have chosen to keep the absolute arriving times rather than the interarrival times when separating incoming and outgoing traffic. This takes into account the mutual influence of the two data streams as it is observed at the gateway.

The multifractal scaling was excellent in the scaling region $2^8, \ldots, 2^{18}$ (see Fig. 13). The partition function $\tau^T(q)$ obtained by least square fitting is found to be almost linear. So, one could suspect that $\tau^T(q)$ does not contain more information than one single parameter such as the slope. But a closer look, in particular a comparison with the *inverse aspect* P (packets arriving per time) reveals the accuracy and the significance of the spectrum $f_{\mathrm{L}} = \tau^*$. We elaborate on this issue in 4.1.3.

4.1.3 Aspects with respect to time

From point of view of an application in traffic modeling the most interesting aspect are certainly the ones with respect to 'time'. When cumulating data into time-intervals of equal length one has to be aware, though, that additional information may be created. We are about to explain in detail several methods we compared. As they all show the same τ and f_{G} (see Fig. 14) we consider the multifractal behavior as being established.

The three methods of cumulating are the following, all based on new time instances $t_k = k \cdot \delta t$ with a fixed time interval δt.

- Compute the 'number' of packets P_k arriving in the artificial time interval $I_k := [t_k, t_{k+1}]$, accounting for each effective arrival interval $[T_i, T_{i+1}]$ according to its 'time spent in I_k':
$$P_k = P_k(\delta t) = \sum_i \frac{|[T_i, T_{i+1}] \cap I_k|}{|[T_i, T_{i+1}]|}$$
where $|\cdot|$ denotes the length of an interval. Similarly, the 'number' of bytes B_k 'arriving' in I_k is
$$B_k = B_k(\delta t) = \sum_i Z_i \cdot \frac{|[T_i, T_{i+1}] \cap I_k|}{|[T_i, T_{i+1}]|}.$$
With this method, artificial information has been created.

- Approximate uniform inter-arrival times δt as closely as possible by choosing actual time instances $T_{l(k)}$ such that
$$T_{l(k)} \leq T_{l(k-1)} + \delta t < T_{l(k)+1}.$$
For this procedure to make sense, we need $\delta t \geq T_{i+1} - T_i$ for all i. Now, take $P^*_k(\delta t) := l(k+1) - l(k)$ and, consequently,
$$B^*_k = B^*_k(\delta t) = \sum_{i=l(k)}^{l(k+1)-1} Z_i.$$
No additional information has been created for the price of having non-uniform time intervals. In order to switch from time scale δt to $2 \cdot \delta t$, $4 \cdot \delta t$, $8 \cdot \delta t$ etc. one may
  - repeat the procedure with time step $2\delta t$, i.e. consider $P^*_k(2 \cdot \delta t)$, $P^*_k(4 \cdot \delta t)$, etc. and similarly for B^*,
  - or simply cumulate recursively, i.e. consider $P^{**}_k(\delta t) := P^*_k(\delta t)$, $P^{**}_k(2^{l+1} \cdot \delta t) := P^{**}_{2k-1}(2^l \delta t) + P^{**}_{2k}(2^l \delta t)$, and similarly for B^{**}.

  It is clear that the computation is faster in the second case but that the approximation is better in the first case. The piecewise increments τ_m showed no significant difference for the three methods, leading to virtually the same estimates of $\tau(q)$ (see Fig. 14).

Of particular interest is the aspect 'packets per time' P_k since it is 'inverse' to the aspect 'time per packet' T_i. Theory [MR, RM1, RM2] says that the spectra should be related by the formula
$$f^{(T)}(\alpha) = \alpha f^{(P)}(1/\alpha),$$
resp.
$$\tau^{(T)} = -q^{(P)} \qquad \text{and} \qquad \tau^{(P)} = -q^{(T)}.$$
Note, that the transformation $(\alpha, f) \mapsto (1/\alpha, 1/\alpha \cdot f)$ exchanges the bisector of the axes $f = \alpha$ with the horizontal line $f \equiv 1$, the two typical touching lines of a spectrum. Whence, the 'symmetry' of the spectra $f^{(T)}$ and $f^{(P)}$ which is clearly visible in Fig. 15.

Several remarks are in order.

- First, the fact that the predicted symmetry between $f^{(T)}$ and $f^{(P)}$ holds with high accuracy proves that these spectra contain significant and correct information.

- Second and most importantly, we would like to stress that the main difference between the observed outgoing and incoming traffic remains hidden to an analysis based on statistical scaling analysis of the interarrival times T_i, since this difference manifests itself in the part of the spectrum corresponding to *negative* q. There, the scaling of centered data is poor (see Fig. 11, 19, 20, 21 and 22).

- Finally, all aspects reflecting the 'traffic load' (Z_i, P_k and B_k) show that the outgoing traffic observed at Berkeley is clearly more regular and less bursty than the incoming traffic (see Fig. 16).

4.2 Traffic at CNET Labs

An analysis of traffic at the LAN at CNET Laboratories showed again excellent multifractal scaling, certainly superior to the statistical scaling (see Subsection 4.3, as well as Fig. 19 and 22).

It was again striking to find such a clear difference between incoming and outgoing traffic. This time, the outgoing traffic showed very small α, in fact only $\alpha \leq 1$, while the incoming and the combined traffic showed almost no tendency towards burstiness (see Fig. 17 and 18). Traffic with 'left sided' spectra are very irregular, in other words 'bursty'. Comparing with the findings for the traffic at Berkeley it has to be concluded that it is not necessarily the incoming traffic which contributes the bulk of the bursty traffic at the gateway of a LAN.

4.3 Statistical versus multifractal scaling

Recall that we denoted the given positive data by Z_i and that we set

$$X_i := Z_i - \hat{\mathbb{E}}Z = Z_i - 1/N \sum_{i=1}^{N} Z_i.$$

in order to obtain centered data. For Z_i a multifractal analysis is in order while X_i is suitable for a statistical analysis estimating the parameter of self-similarity H.

In order to compare the two scaling behaviors note that $\sum_{k=1}^{N/m} X_k^{(m)} = 0$, whence

$$S_m^Z(2) \;=\; \sum_{k=1}^{N/m} (\overline{Z}_k^{(m)})^2 = m^2 \sum_{k=1}^{N/m} (Z_k^{(m)})^2$$

$$= \quad m^2 \sum_{k=1}^{N/m} (X_k^{(m)} + \hat{\mathbb{E}} Z)^2 = m^2 \sum_{k=1}^{N/m} (X_k^{(m)})^2 + mN(\hat{\mathbb{E}} Z)^2$$
$$= \quad S_m^X(2) + mN(\hat{\mathbb{E}} Z)^2. \tag{15}$$

At this point a conceptual difficulty arises. The statistical test looks for asymptotic behavior as $m \to \infty$, while multifractal analysis is formulated in terms of the limit $\delta_m = m/N \to 0$. The equation above shows now that scaling can't be perfect for both simultaneously except in the trivial case $\gamma(2) = \tau(2) = 1$.

Nevertheless, the scaling behavior of both, S_m^Z and S_m^X may be acceptable in a *scaling region*, i.e. a range of values of m with nearly linear behavior in log-log coordinates. Some remarks are in order.

First, $S_N^Z(q) = (S_N^Z(1))^q$ implies $\tau_N^Z(q) \simeq q$, and $S_N^X(q) = 0 =$ const yields $\gamma_N^X(q) \simeq 0$. This forces the scaling to break down as $m \to N$. It is clear that the effect on γ_m^X is stronger since both, τ and γ increase with q (see Fig. 19, 20, 21 and 22). In the limit $m \to 1$, on the other hand, imprecision of the measurement will affect the scaling behavior.

Second, $\tau(q)$ is a convex function as has been shown before (see also [B]). Since $\tau(0) = -1$ and $\tau(1) = 0$, we must have $\tau(2) \leq 1$. On the other hand, $\gamma(2) \leq 1$ follows from $H \leq 1$. Thus, the term of order m^1 will not be essential for $m \to 1$.

All this makes clear that scaling should be expected rather in the multifractal limit $m/N \to 0$. If scaling is present there, (15) suggests

$$\tau(2) = \min(1, \gamma(2)) = \gamma(2) = 2H - 1. \tag{16}$$

Equation (16) illustrates, first of all, the way in which multifractal analysis goes beyond the 'mono-fractal' statistical tests based on an estimation of H. As an estimator of H, however, $1/2(\tau(2) + 1)$ is certainly of limited use.

To illustrate this point in some detail, and in order to show what different conclusions a statistical approach would provide we put forward some estimations of H through $\hat{H}_s := 1/2(\gamma(2) + 1)$ and $\hat{H}_m := 1/2(\tau(2) + 1)$ which are obtained by least square fitting of the log-log plot of the corresponding partition sum in a suitable scaling region.

In conclusion, $\hat{H}_m$ is considerably larger than $\hat{H}_s$ throughout the data and is, thus, of limited use. Table 1 illustrates, though, that one has to take the whole spectrum $\tau(q)$ into consideration, not only one value such as $\tau(2)$.

With the statistical method of moments, on the other hand, we encounter as an essential problem a general scaling behavior of a lesser quality, and more importantly, a significant dependence of the estimator on the choice of the scaling region. In this context it is essential to note that despite of poor scaling, the partition functions $\gamma^B(q)$ and $\gamma^T(q)$ of

$\hat{H}_m$	**outgoing**	**incoming**	**combined**
Berkeley, Z	.981	.968	.983
Berkeley, B	.957	.916	.952
CNET, Z	.987	.998	.998
CNET, B	.968	.937	.964

Table 1: The multifractal estimator $\hat{H}_m = (\tau(2)+1)/2$ of the Hurst exponent for the aspects Z (bytes per packet) and B (bytes per time). The other two aspects showed no conclusive difference between outgoing and incoming traffic. As this table illustrates, conclusions on the irregularity of data cannot rely on the value $\tau(2)$ only. With the aspect B at CNET, e.g. $\hat{H}_m$ suggests that the incoming traffic is more irregular than the outgoing one, a conlcusion which is certainly denied when taking the whole spectrum into account (Fig. 18).

the centered aspects B (bytes per packet) and T (time per packet) is indeed fairly linear for $q \geq -1$ up to large values, supporting the asumption of self-similarity which forms the bases of the statistical approach. The corresponding Legendre transform $\gamma^*(\alpha)$ forms, thus, an 'edge' at its maximum, i.e. one finds straight lines of these slopes q which touch $\gamma^*(q)$ in only one point (Fig. 12).

In summary, in our context the multifractal analysis is superior to the statistical method of moments for the following reasons:

- Throughout all data considered in this work, the multifractal scaling was more convincing.
- Scaling was typically observed in the multifractal limit $m \to 0$ and not as $m \to N$.
- The necessary procedure of centering the data introduces inaccuracy in the statistical method of moments, whence the scaling is unacceptable for negative q.
- Essential properties concerning regularity and burstiness were found when looking at the whole spectrum and not only at one parameter.
- With the inter-arrival times the properties just mentioned manifest themselves in the range corresponding to negative q and were, thus, even less accessible through the statistical method of moments.

4.4 Discussion

We found that the data traffic at Berkeley and at CNET labs is clearly multifractal. Moreover, we would like to stress that the scaling is extraordinary in both, quality and

$\hat{H}_s$	outgoing	incoming	combined
Berkeley, Z	.828	.795	.832
Berkeley, T	.835	.824	.843
Berkeley, B	.800	.767	.775
Berkeley, P	.790	.815	.790
CNET, Z	.896	.876	.928
CNET, T	.874	.771	.900
CNET, Z	.915	.917	.924
CNET, T	.926	.904	.926

Table 2: The statistical estimator $\hat{H}_s$ for various aspects. The lines at the end are obtained from the same data (as indicated), however with different scaling regions $2^5, \ldots, 2^{10}$ as compared to $2^1, \ldots, 2^{18}$ which was applied at the top 6 lines (compare Fig. 20).

size of the scaling region. The latter is especially astonishing when compared to findings in other fields: the scaling region of TCP traffic spans four to five orders of magnitude as compared to the one to two orders which are typical in other fields [Ha]. We also found statistical self-similarity, but notably only for $q \geq 1$.

We see possible reasons for the high quality of the multifractal scaling in the hierarchy inherent to telecommunication protocols, as well as in the fact that data transfers involve splitting wholes into pieces.

The analysis we provide, in particular the various spectra $f_{\rm G}$ and $f_{\rm L}$, are best interpreted from a 'distant point of view'. Rather than looking at the exact values $f(\alpha)$ one looks at the shapes of the spectra. These spectra reveal telling information about the irregularities as well as regularities of the data traffic. Among such we were in particular able to distinguishing between incoming and outgoing traffic and suggest that heavy and moderate incoming traffic at Berkeley are independent from each other. This could not be achieved with an estimation of the statistical parameter H (see Table 2).

One could wonder which traffic, outgoing or incoming, should be more regular, both being subject to some smoothing effects: the incoming traffic comes from a WAN where a greater number of sources could lead to some averaging, while the outgoing traffic is produced by a LAN where response times are shorter and flow control should be more efficient. Our analysis, especially the comparison of the traffic at Berkeley and at CNET, shows that the very type of LAN has a much greater influence than the mentioned effects. A tentative explanation of the very sporadic outgoing traffic at CNET could be found in the fact that the essential part of this traffic consists in consultations of the World Wide Web (where the server is located outside the LAN). One may suspect that the traffic leaving the LAN is more sporadic since it consists mainly of a control traffic of the web clients (ACKs and NACKS) with a few 'true' data transfers in between. The traffic arriving at the LAN stems mainly from the web server and is –as is most apparent in its aspect B– more regular since it runs constantly on 'heavy mode'. Investigations on Web traffic seem to support our finding that it differs from 'usual' traffic [C].

Proposing as models of arriving traffic loads simple random binomial measures (see Fig. 3) we do not suggest to fit a whole spectrum of scaling exponents $\tau(q)$ to a given trace but to adjust rather few parameters which determine the distribution of the random weights M_0, M_1 as explained in Subsection 2.3. The model in Fig. 3, e.g. is obtained by taking $R_k \equiv 1/2$ and $M_0 \simeq U[0,1]$, $M_1 := 1 - M_0 \simeq U[0,1]$. It is clear that this is too simple a model, especially since its spectrum is perfectly concave. But modeling was not what we aimed for but merely a visual comparison of a random binomial measure with an actual trace. Indeed, having approximately log-normal marginal distributions these models seems to well characterize real data traffic [P].

Now that the multifractal structure of TCP traffic has been demonstrated, several tasks lay ahead. First, it would be useful to calculate essential statistical information such as forecasting for such random measures, or multifractal processes. The knowledge of the whole multifractal spectrum will then allow more precise statistics than the knowledge of one scaling or 'burst' exponent H only. Secondly, 'physically' relevant models with multifractal properties are needed for a better understanding of TCP traffic. These issues will be addressed in forthcoming papers [RL, LR].

Conclusions.

TCP traffic is clearly multifractal with excellent scaling of the comparably large scaling region of four to five orders of magnitude. There are good numerical estimators for the multifractal spectrum of a measure. From this spectrum important information about the scaling structure and burstiness can be obtained which remains hidden to an analysis relying on the Hurst exponent only H. It has been shown that the monofractal assumption, which is implicitly used when modeling with FBM, does not apply.

Acknowledgement.

We gratefully acknowledge the financial support of CNET, France, under contract # 95 8B 069. We would like to express our thanks to Fabrice Clerot and Ilkka Norros for useful comments on an earlier version of this paper.

References

[AP] M. ARBEITER AND N. PATZSCHKE, Self-Similar Random Multifractals, *Math. Nachr.* **181** (1996) pp 5–42.

[B] C. BECK, Upper and lower bounds on the Renyi dimensions and the uniformity of multifractals, *Physica D* **41** (1990) pp 67–78.

[BMP] G. BROWN, G. MICHON AND J. PEYRIERE, On the Multifractal Analysis of Measures, *J. Stat. Phys.*: **66** (1992) pp 775–790.

[CM] R. CAWLEY AND R. D. MAULDIN, Multifractal Decompositions of Moran Fractals, *Adv. in Math.* **92** (1992) pp 196–236.

[CLP] P. COLLET, J. LEBOVITC AND A. PORCIO, The Dimension Spectrum of Some Dynamical Systems, *J. Stat. Phys.* **47** (1987) pp 609–644.

[C] M. CROVELLA AND A. BESTAVROS Self-similarity in World Wide Web traffic. Evidence and possible causes. In *Proceedings of SIGMETRICS '96*, May 1996.

[D] L. DEVROYE, The double kernel method in density estimation, *Ann. Inst. Henri Poincaré* **25** (1989) 533–580.

[Ell] R. ELLIS, Large Deviations for a general Class of Random Vectors, *Ann. Prob.* **12** (1984) pp 1–12.

[EM1] C. J. G. EVERTSZ AND B. B. MANDELBROT, Multifractal Measures, *Appendix B in: 'Chaos and Fractals' by H.-O. Peitgen, H. Jürgens and D. Saupe,* Springer New York (1992) pp 849–881.

[EM2] C. J. G. EVERTSZ AND B. B. MANDELBROT, Self-similarity of the harmonic measure on DLA, *Physica* **A 185** (1992) pp 77–86.

[F1] K. J. FALCONER, Fractal Geometry: Mathematical Foundations and Applications, John Wiley and Sons, New York (1990).

[F2] K. J. FALCONER, The Multifractal Spectrum of Statistically Self-Similar Measures, *J. Theor. Prob.* **7** (1994) pp 681–702.

[FP] U. FRISCH AND G. PARISI, Fully developed turbulence and intermittency, *Proc. Int. Summer school on 'Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics'* (M. Ghil, R. Benzi and G. Parisi, Amsterdam, North-Holland 1985) pp 84–88.

[Gr1] P. GRASSBERGER, Generalized Dimensions of strange attractors, *Phys. Lett. A* **97** (1983) pp 227–230.

[Gr2] P. GRASSBERGER, Generalizations of the Hausdorff Dimension of Fractal Measures, *Phys. Lett. A* **107** (1985) pp 101–105.

[GBP] P. Grassberger, R. Badii, and A. Politi, Scaling Laws for Invariant Measures on Hyperbolic and Nonhyperbolic Atractors, *J. Stat. Phys.* **51** (1988) 135–178.

[GP1] P. Grassberger and I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.* **50** (1983) pp 346–349.

[GP2] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* **9** (1983) pp 189–208.

[HJKPS] T. Halsey, M. Jensen, L. Kadanoff, I. Procaccia and B. Shraiman, Fractal measures and their singularities: The characterization of strange sets, *Phys. Rev. A* **33** (1986) pp 1141–1151.

[Ha] D. Hamburger et al., *Phys. Rev. E* **53** (1996) pp 3342.

[HP] H. Hentschel and I. Procaccia, The Infinite number of Generalized Dimensions of Fractals and Strange Attractors, *Physica D* **8** (1983) pp 435–444.

[HI] T. Hirata and M. Imoto, Multifractal analysis of spatial distribution of microearthquakes in the Kanto region, *Geophys. J. Int.* **107** pp 155-162.

[HW] R. Holley and E. Waymire, Multifractal dimensions and scaling exponents for strongly bounded random cascades, *Ann of App. Prob.* **2** (1992) pp 819–845.

[J1] S. Jaffard, Multifractal Formalism for Functions, Ph. D. thesis 1993, see also *CRAS:* **317** (1993) pp 745–750.

[J2] S. Jaffard, Pointwise smoothness, two-microlocalization and wavelet coefficients, *Publicacions Matematiques* **35** (1991) pp 155–168.

[JKP] M. Jensen, L. Kadanoff and I. Procaccia, Scaling structure and thermodynamics of strange sets, *Phys. Rev. A* **36** (1987) pp 1409–1420. *Adv. Appl. Math.* accepted for publication (1995).

[KP] J.-P. Kahane and J. Peyrière, Sur Certaines Martingales de Benoit Mandelbrot, *Adv. Math.* **22** (1976) pp 131–145.

[LTWW] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, On the self-similar nature of Ethernet Traffic (extended version), *IEEE/ACM Transactions on Networking* (Feb. 1994) pp 1–15.

[L1] J. Lévy Véhel, Numerical Computation of the Large Deviation Multifractal Spectrum, In *CFIC'96*, Rome, 1996.

[L2] J. Lévy Véhel, Fractal Approaches in Signal Processing, *Fractal Geometry and Analysis, The Mandelbrot Festschrift, Curacao 1995*, C.J.G. Evertsz, H.-O. Peitgen, R.F. Voss (eds.), World Scientific, 1996.

[L3] J. Lévy Véhel, Introduction to the Multifractal Analysis of Images, In *Fractal Image Encoding and Analysis*, Yuval Fisher (ed.), Springer Verlag, 1996.

[LC] J. Lévy Véhel and C. Canus, Hausdorff dimension estimation and application to multifractal spectrum computation, Technical report, INRIA, 1996.

[LR] J. Lévy Véhel and R. Riedi, Simulating TCP Traffic using Weakly Self-Affine Functions, submitted to *Fractals in Engineering 97*.

[LV] J. Lévy Véhel and R. Vojak, Multifractal analysis of Choquet capacities: Preliminary Results, *Adv. Appl. Math.* accepted for publication (1995).

[LT] J. Lévy Véhel and C. Tricot, Continuous Multifractal Spectra, *preprint*.

[M1] B. B. Mandelbrot, "The Fractal Geometry of Nature", New York Freeman (1982).

[M2] B.B. Mandelbrot, Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence, In *Statistical Models and Turbulence*, pages 331–351, La Jolla, California, 1972. Edited by Murray Rosenblatt and Charles Van Atta, New York, Springer. (Lecture Notes in Physics, 12).

[M3] B. B. Mandelbrot, Intermittent turbulence in self similar cascades : divergence of high moments and dimension of the carrier, *J. Fluid. Mech.* **62** (1974) p 331.

[M4] B. B. Mandelbrot, New 'Anomalous' Multiplicative Multifractals: Left sided $f(\alpha)$ and the modeling of DLA, *Physica A* **168** (1990) pp 95–111.

[M5] B. B. Mandelbrot, Negative Fractal Dimensions and Multifractals, *Physica A* **163** (1990) pp 306–315.

[MEH] B. B. Mandelbrot, C. J. G. Evertsz and Y. Hayakawa, Exactly self-similar left-sided multifractal measures, *Physical Review A* **42** 8 (1990) pp 4528–4536.

[ME] B. B. Mandelbrot and C. J. G. Evertsz, Multifractality of the harmonic measure on fractal aggregates and extended self-similarity, *Physica* **A** **177** (1991) pp 386–393.

[MR] B. B Mandelbrot and R. H. Riedi, Inverse Measures, the Inversion formula and Discontinuous Multifractals, *Adv. Appl. Math.* **18** (1997) pp 50–58.

[N1] I. Norros, A storage model with self-similar input, *Queueing Systems* **16** (1994) pp 387–396.

[N2] I. Norros, On the use of fractional Brownian motion in the theory of connectionless networks, *COST/* **242** (1994).

[O] L. Olsen, Random geometrically graph directed self-similar multifractals, *Pitman Research Notes Math. Ser.* **vol 307** (1994).

[OWY] E. Ott, W. Withers and J. Yorke, Is the Dimension of Chaotic Attractors invariant under Coordinate Changes? *J. Stat. Phys.* **36** (1984) pp 687–697.

[PR] R. Pastor-Satorras and R. Riedi, Numerical Estimates of Generalized Dimensions D_q for Negative q, *J. Phys. A: Math. Gen.* **29** (1996) L391–L398.

[P] V. Paxson. Emperically-derived analytic models of wide-area TCP connections. *IEEE/ACM Transactions on Networking*, **2** (1994) 316–326.

[R1] R. Riedi *An Improved Multifractal Formalism and Self-Affine Measures*, Ph. D. thesis, ETH Zürich (1993).

[R2] R. H. Riedi, An Improved Multifractal Formalism and Self-Similar Measures, *J. Math. Anal. Appl.* **189** (1995) pp 462–490.

[R3] R. H. Riedi, Seven approaches to Multifractal Analysis: an introduction, *forthcoming.*

[RL] R. Riedi and J. Lévy Véhel, Multifractal Models for TCP Traffic, *forthcoming.*

[RM1] R. H. Riedi and B. B Mandelbrot, Multifractal Formalism for Infinite Multinomial Measures, *Adv. Appl. Math.* **16** (1995) pp 132–150 .

[RM2] R. H. Riedi and B. B Mandelbrot, Inversion formula for Continuous Multifractals, *Adv. Appl. Math.* **19** (1997) pp 332–354.

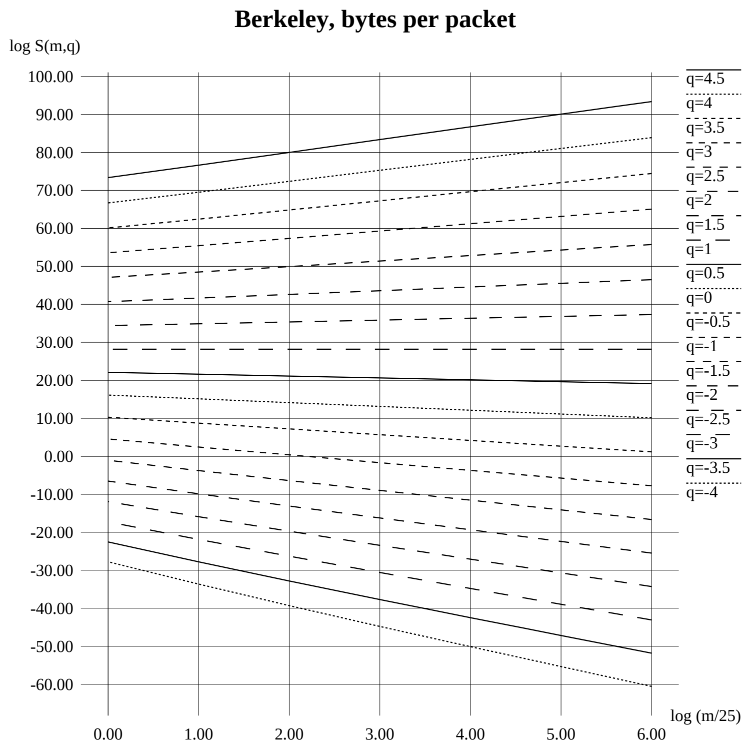
[RM3] R. H. Riedi and B. B Mandelbrot, Exceptions to the Multifractal Formalism for Discontinuous Measures, to appear in *Math. Proc. Cambr. Phil. Soc.* (May 1997).

[TTW1] M. Taqqu, V. Teverovsky and W. Willinger, Estimators for Long-Range Dependence: An empirical study, *Fractals.* **3** (1995) pp 785–798.

[TTW2] M. Taqqu, V. Teverovsky and W. Willinger, Is the Ethernet data self-similar or multiaffine?, *Draft, Boston Univ.* (March 1996).

[V] T. Vicsek, Fractal Growth Phenomena, *World Scientific*, Singapore, 1989.

[VL] R. Vojak and J. Lévy Véhel, Multifractal Description of Road Traffic Structure, *IFAC'94, Tanjin (China), August 24-26* (1994).



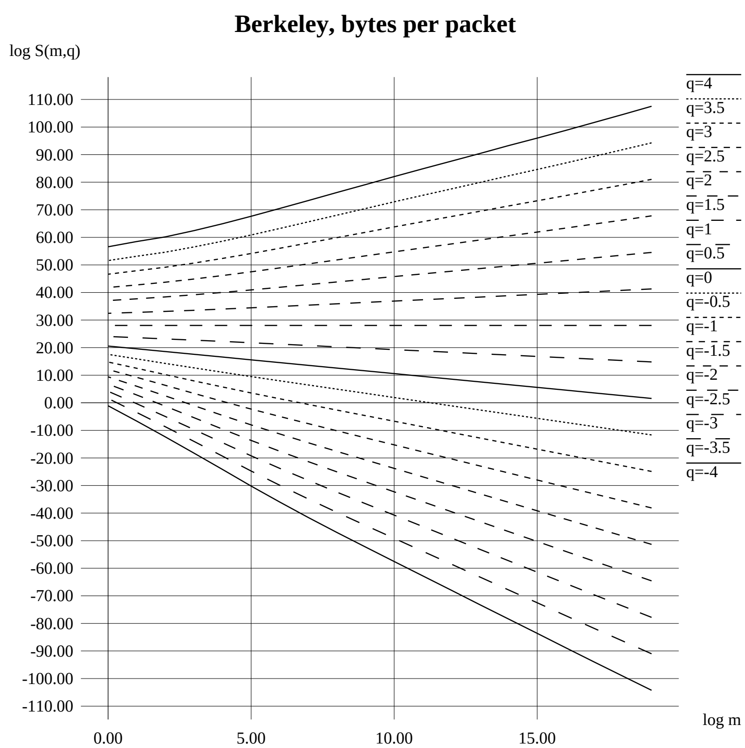


Figure 7: The scaling behavior of Z_i (bytes per packet) for the (combined) traffic observed at Berkeley demonstrated in log-log plots of $S_m^Z(q)$ against m, where from top to bottom q runs through (4.5), 4, 3.5, $3, \ldots, -3.5, -4$. The ordinate is $\log_2 S_m^Z(q)$. In the figure on the top $m = 25, 50, 100, \ldots, 1600$ and the abscissa is $\log_2(m/25)$. In the figure on the bottom $m = 1, 2, 4, \ldots, 2^{19}$ and the abscissa is $\log_2(m)$.

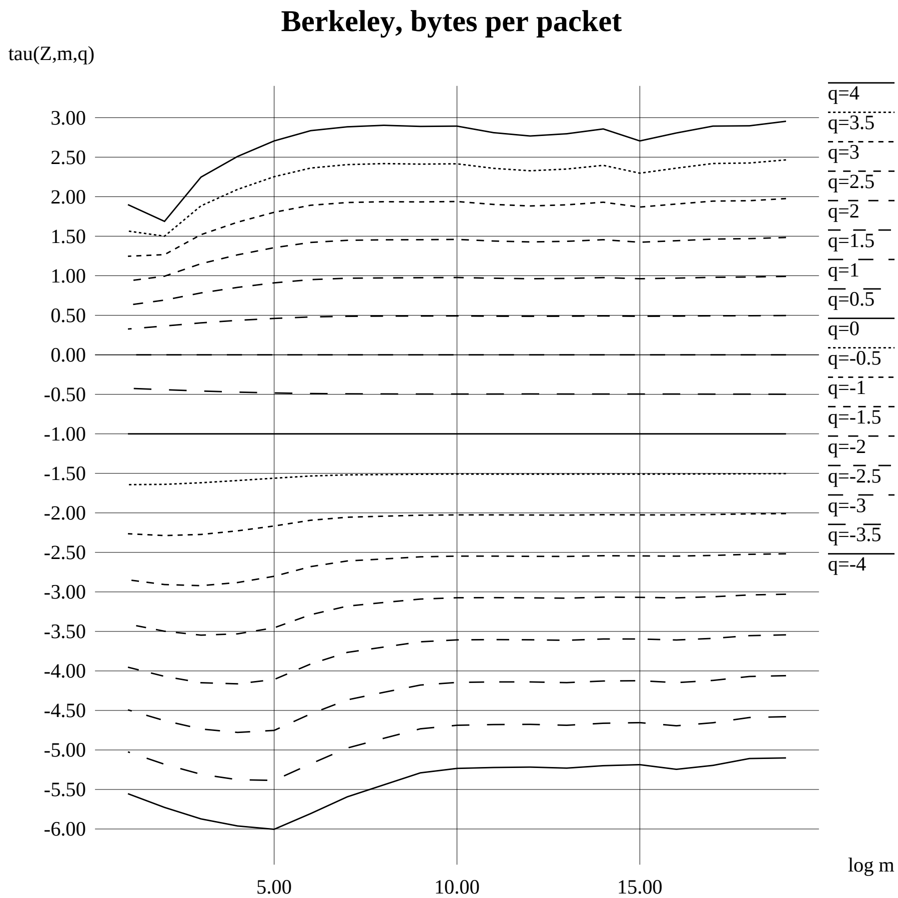


Figure 8: The piecewise increments of the log-log plot in Figure 7 (top), i.e. $\tau_m(q) = \log_2 S^Z_{2m}(q) - \log_2 S^Z_m(q)$ as defined in (12) against $\log_2 m$. Note that we display the whole available range $m = 1, 2, 4, \ldots, 2^{19}$. This demonstrates impressively that the graphs are indeed very close to being linear. Any average of the $\tau_m(q)$ provides a valuable estimate of $\tau(q)$. Nevertheless, $\tau(q)$ has been calculated via a least square fitting in the 'scaling region' $m = 2^7, \ldots, 2^{19}$.

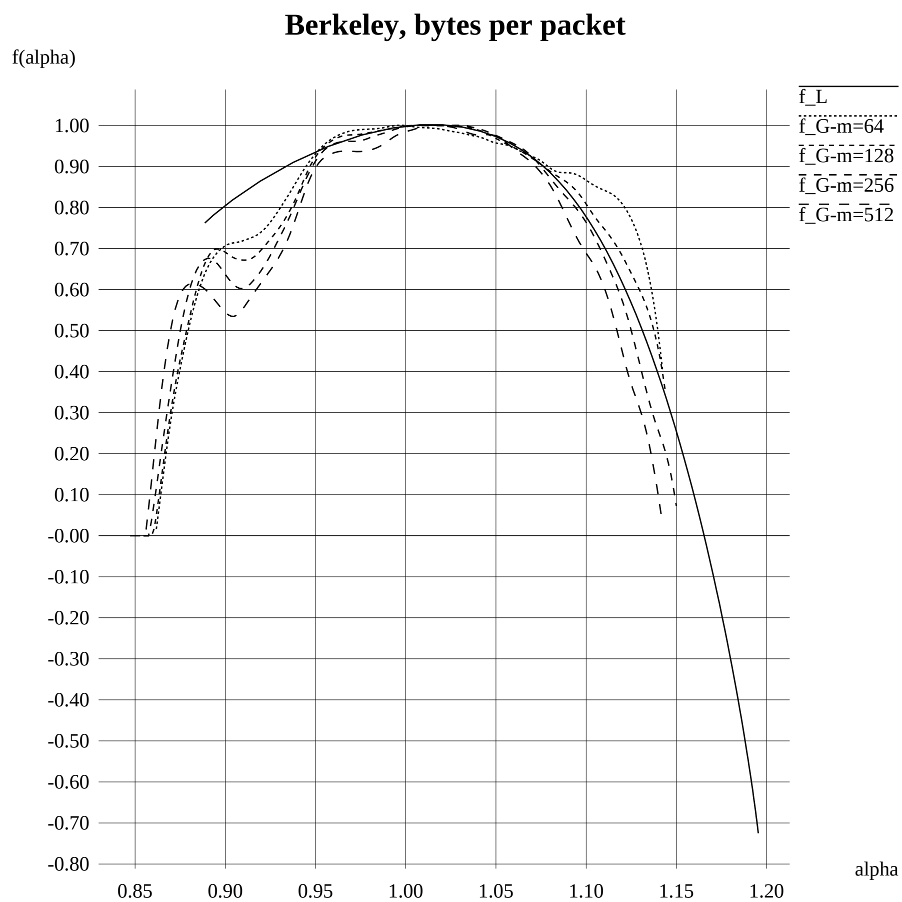
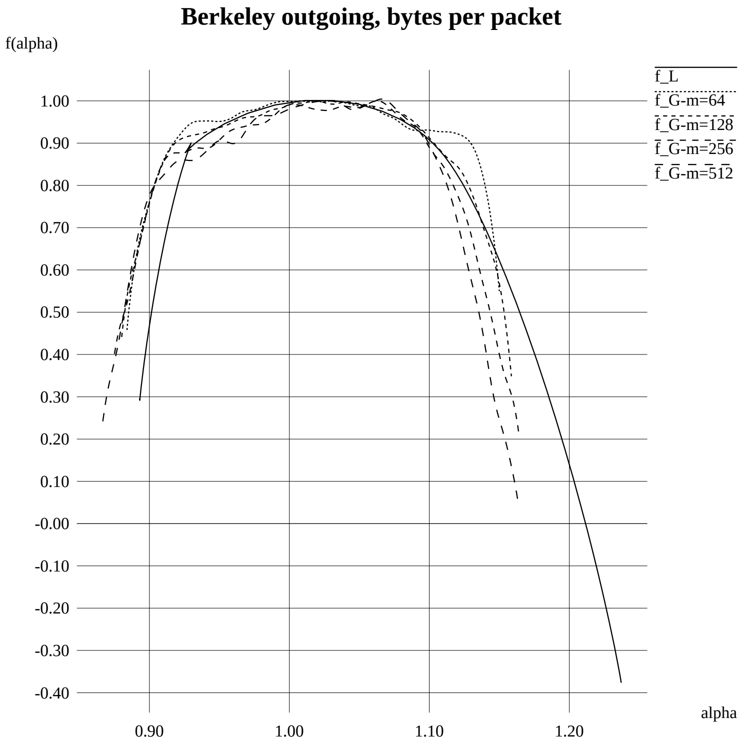


Figure 9: Approximations $f_{\mathrm{G}}{}^{\epsilon_m}(\alpha)$ to f_{G} on the scales $\delta_m = m/N$, $m = 2^6, \ldots, 2^9$ (compare Subsection 3.3) for the same data as in Fig. 7 and Fig. 8. For comparison, the Legendre transform $f_{\mathrm{L}}(\alpha)$ (bold) of $\tau(q)$ is shown as well. Note that $f_{\mathrm{L}}(\alpha)$ is concave by definition. The negative values of $f_{\mathrm{L}}(\alpha)$ correspond to Hölder exponents α which are so rare that they will not be observable when looking at one realization of the process only (compare (3) as well as the remarks in Subsection 2.3).



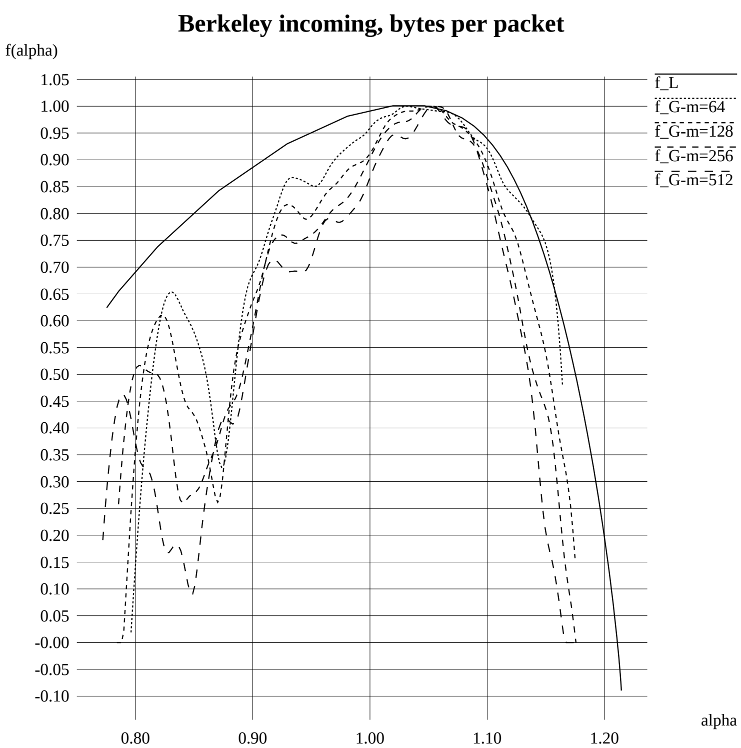


Figure 10: The spectra ${f_{\mathrm{G}}}^{\epsilon_m}(\alpha)$ and $f_{\mathrm{L}}(\alpha)$ as in Fig. 9, however, with a separate analysis for the outgoing traffic, displayed on the top, and for the incoming traffic, shown on the bottom.

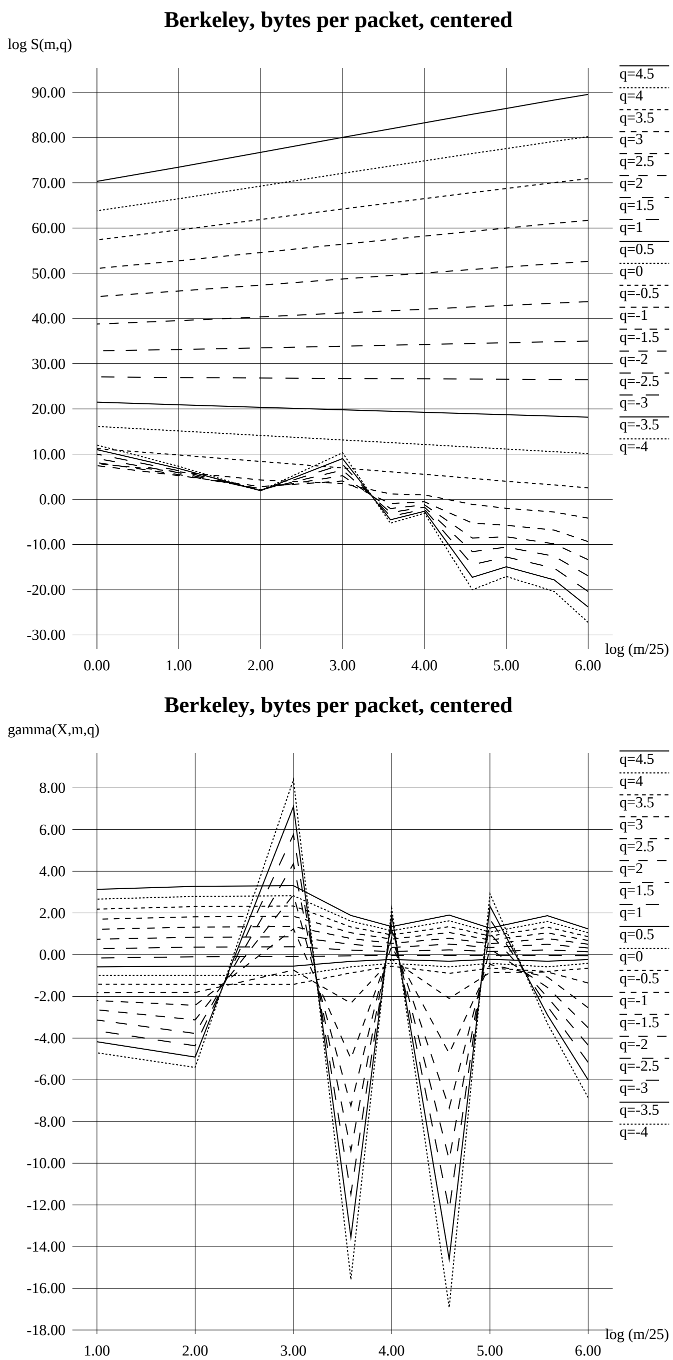


Figure 11: Scaling behavior of the centered data $X_i = Z_i - \hat{\mathbb{E}}Z$ for the same source as in Fig. 7 and Fig. 8, demonstrated in $\log_2 - \log_2$ plots of the partition sum $S_m^X(q)$ (top). On the bottom, the stepwise increments $\gamma_m^X(q) = \gamma(X, m, q) = \log_2 S_{2m}^X(q) - \log_2 S_m^X(q)$ (compare (12)) are. It is clear that one cannot talk of scaling for negative q. But also for positive q we find high variations of up to 100%. Here, the cumulation parameter m varies from $25, 50, \ldots, 1600$. In Fig. 20 a close-up on the range $m = 1, 2, \ldots, 2^{19}$ is shown. This is discussed further in Subsection 4.3.

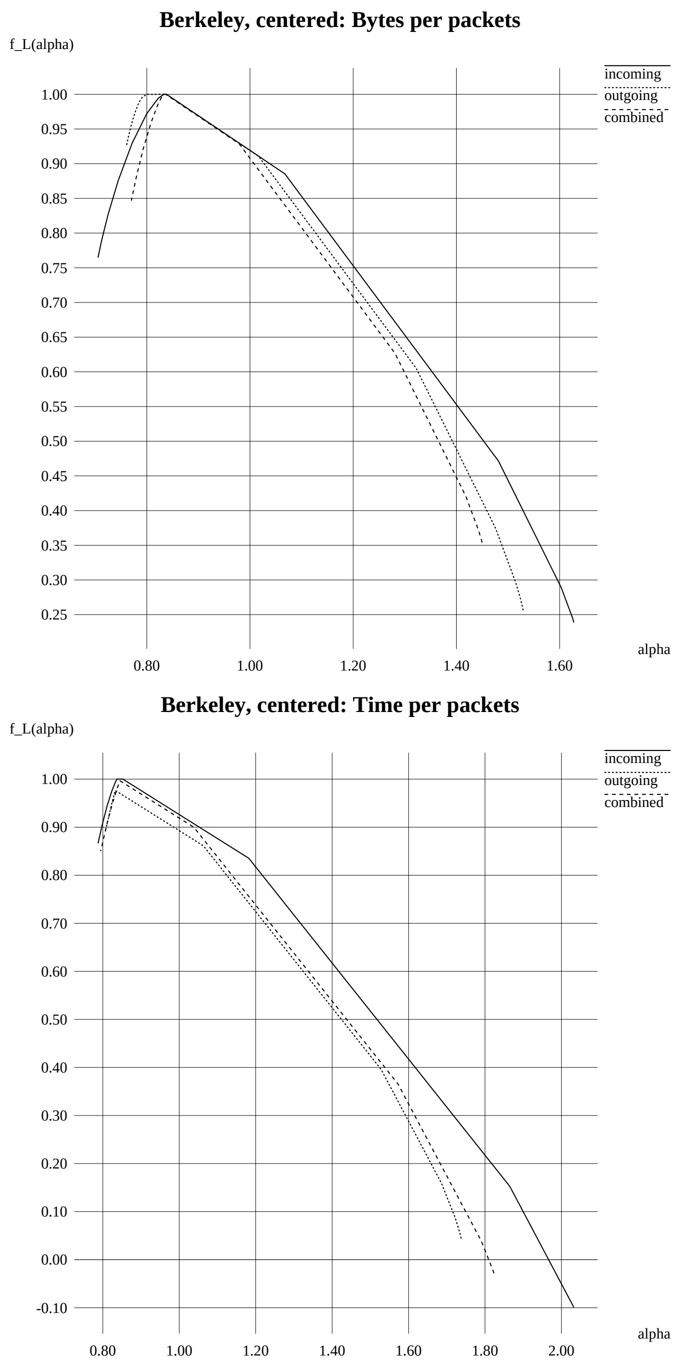


Figure 12: The Legendre transform of $\gamma(q)$ and $\gamma^T(q)$ which were obtained by least square fitting of Fig. 20 and 21 in the scaling region $m = 2^4, \ldots, 2^9$. The complete curves have been calculated despite of poor scaling in the range $q < 0$ which corresponds to the decreasing part of the spectra. The difference between incoming and outgoing traffic shows only with the aspect 'bytes per packets'. From an analysis of the interarrival times, the two traffic look alike. In particular, the LRD parameter H (peak of the spectra) lie in the range $[.832, .845]$. In this context it is important to mention that the spectra show a clear edge at the maximum which is related to the linearity of $\gamma(q)$ for q starting around -1 up to high values.

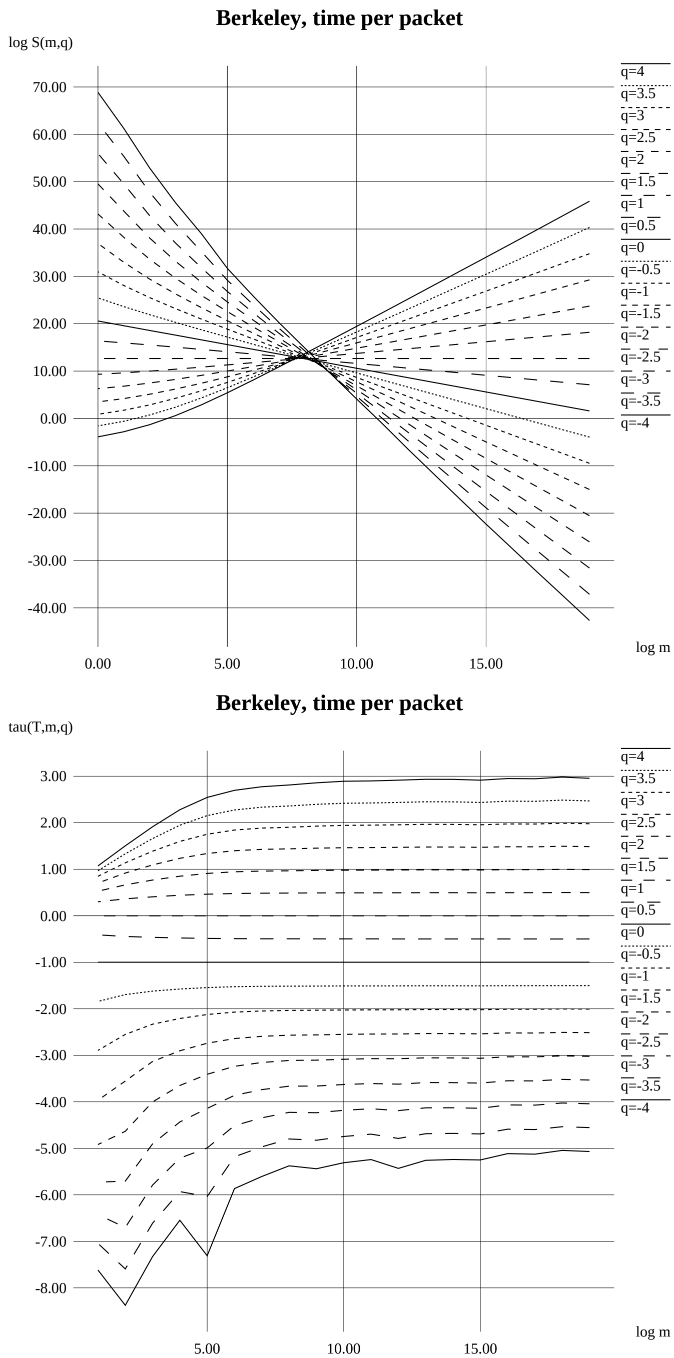


Figure 13: Scaling of the aspect T_i (time per packet) for the traffic observed at Berkeley. On top the usual log-log plot of the partition sum $S_m^T(q) = \sum (\overline{T}_k^{(m)})^q$ of the cumulated data $\overline{T}_k^{(m)}$. On the bottom, the step wise increments $\tau_m^T(q) = \tau(T, m, q) = \log S_{2m}^T(q) - \log S_m^T(q)$ are shown. Scaling is excellent in the region $2^8, \ldots, 2^{18}$. Though $\tau^T(q)$ is nearly linear, especially for positive q, it provides a correct and informative spectrum (see 4.1.2).

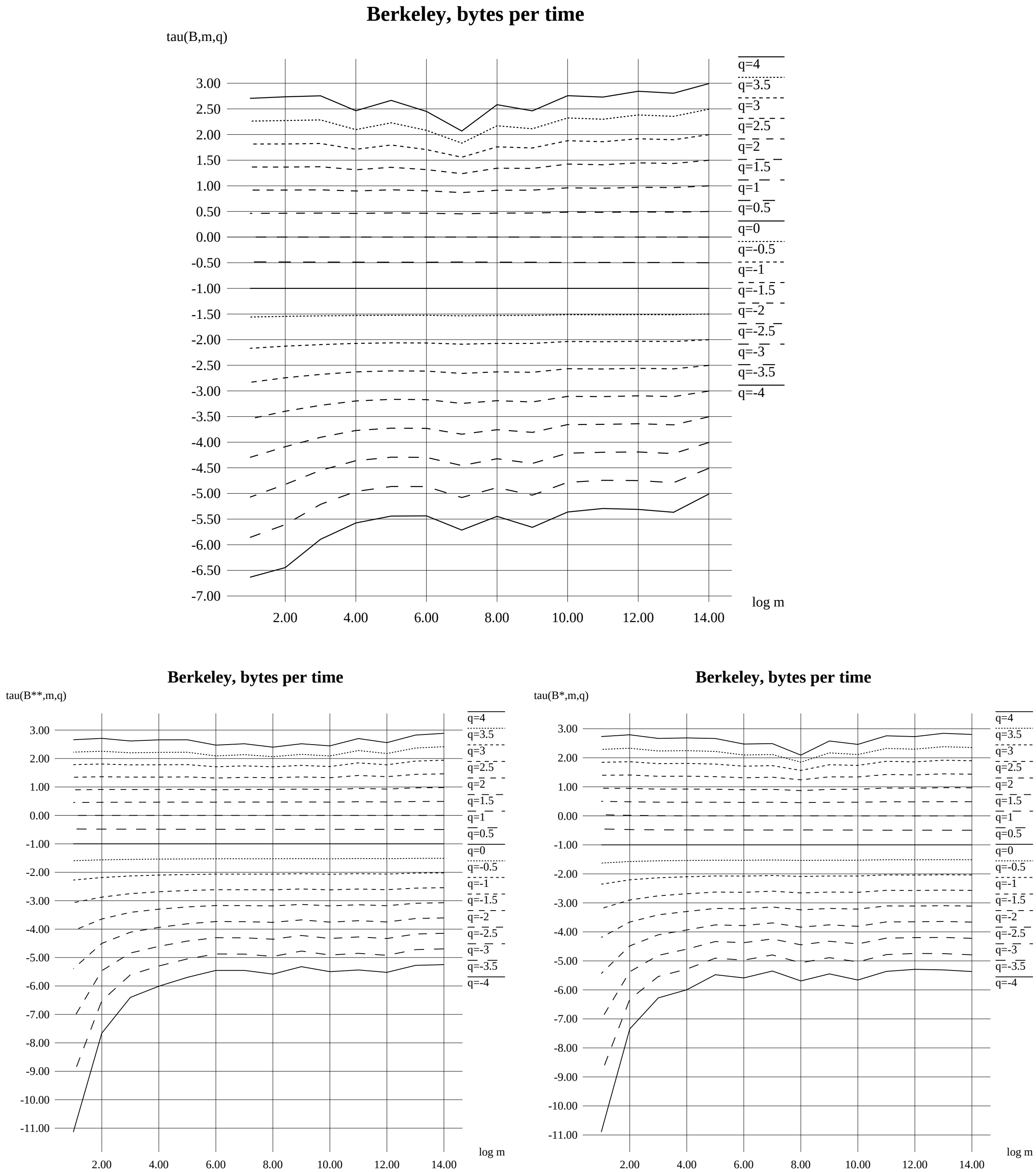


Figure 14: Comparing different ways of cumulating data into time intervals of approximately equal size. Again, we show the stepwise increments $\tau_m^{(\cdot)}(q)$ of the partition sum $\log_2 S_m^{(\cdot)}(q)$ as a function of $\log_2(m)$. It is clear that the first method of cumulation (yielding the data set B) shows the best scaling behavior. All three methods give nearly indistinguishable least square estimates of the partition function $\tau(q)$.

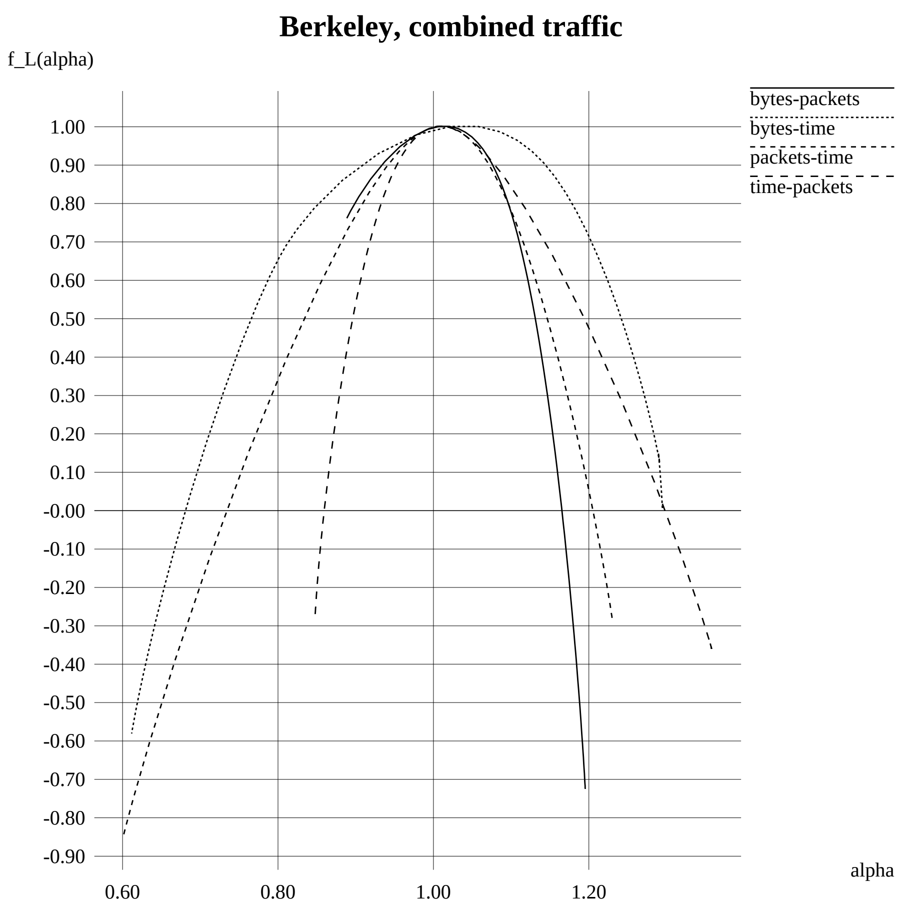


Figure 15: The Legendre spectra $f_{\mathrm{L}}(\alpha)$ for all 'aspects' of the data traffic at Berkeley. The most variation in Hölder exponents is found with B_k (bytes per time). Note again the symmetry between the spectra of P_k and T_i (Fig. 16).

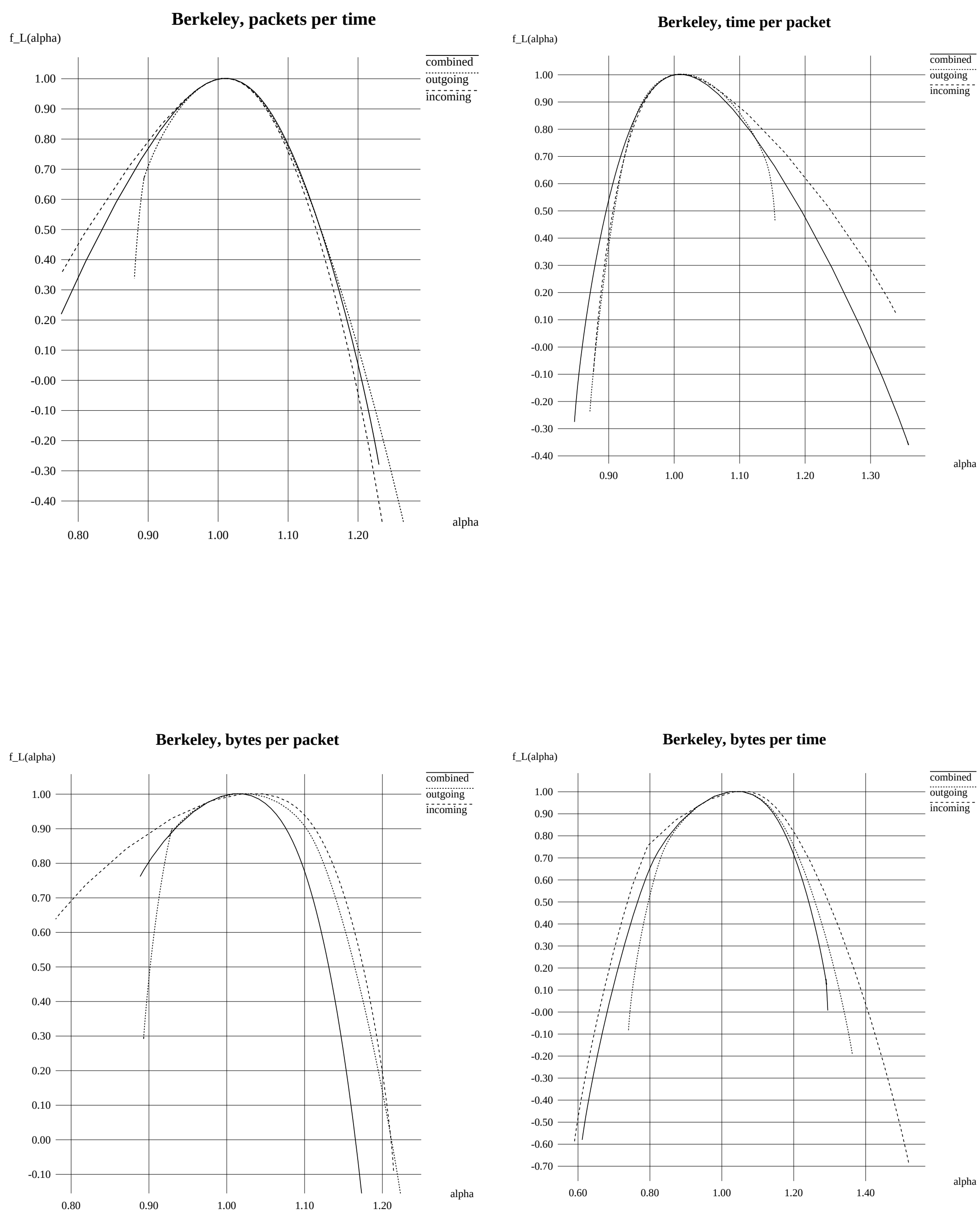
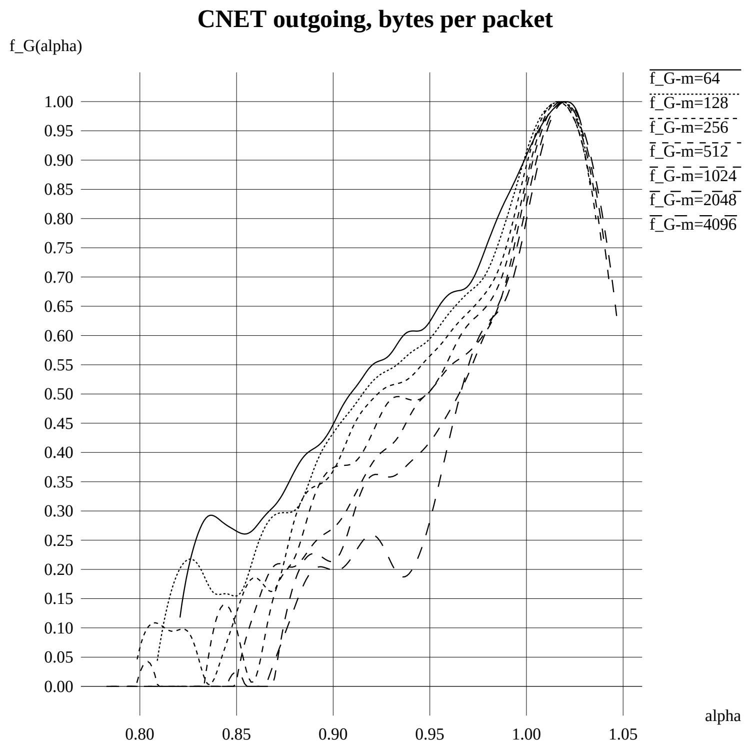


Figure 16: At Berkeley, the incoming traffic shows characteristically different spectra f_{L} than the outgoing traffic and the combined traffic. This becomes apparent when studying any aspect. Note the symmetry of the spectra of the aspects P_k and T_i ('packets per time' and 'time per packet'). These aspects provide samplings of measures which are *inverse* to each other [MR, RM2, RM3]. Theory says that their spectra must be related through the formula $f^P(\alpha) = \alpha f^T(1/\alpha)$, which is perfectly met here.



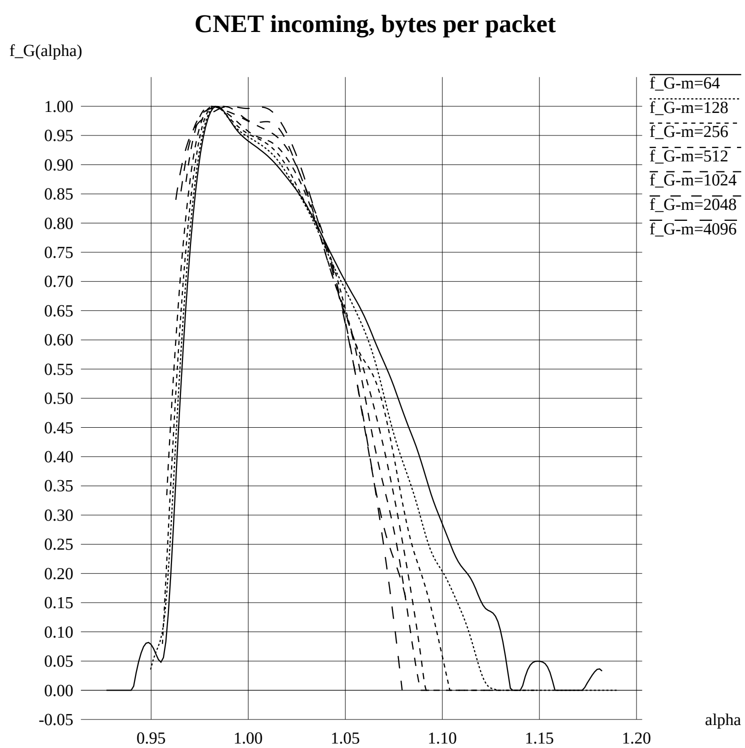


Figure 17: The difference between incoming and outgoing traffic at CNET, hence between incoming and outgoing traffic, is again very striking. Here, we display approximations $f_{\mathrm{G}}{}^{\epsilon_m}(\alpha)$ to the f_{G} spectrum obtained on various levels of aggregation m for the aspect Z 'bytes per packets': on top the strongly left-sided spectra for the outgoing traffic, on the bottom the right-sided spectra for incoming traffic. The spectra of the combined traffic are very similar to the ones of the incoming traffic, which is why they are not shown.

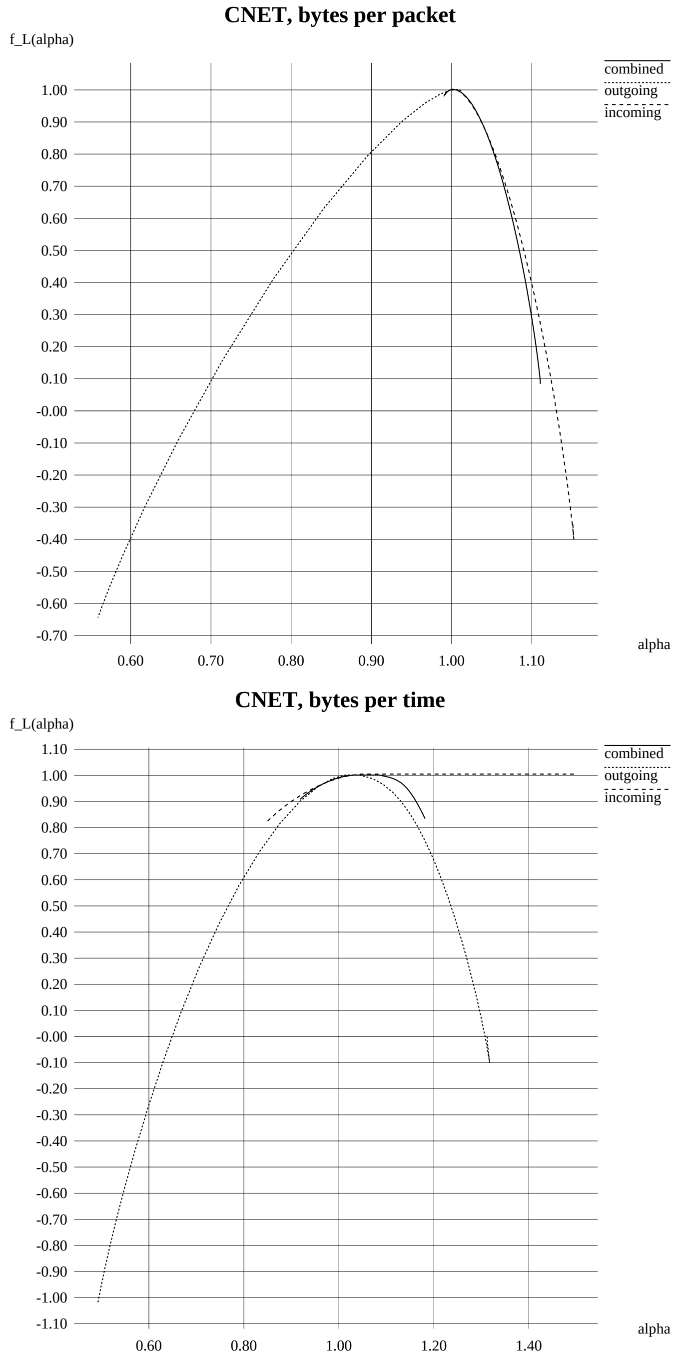


Figure 18: The characteristic difference between incoming, outgoing and combined traffic at CNET is best visible in an analysis of Z, i.e. 'bytes per packets' (on top), demonstrated here with the spectra f_{L}. For the outgoing traffic f_{L} exists of a considerably broad increasing part, having no decreasing part (dashed). Such spectra have been termed 'left-sided' [M4, MEH, ME, RM1]. For the incoming as well as the combined traffic the f_{L} spectrum found is 'right sided' and comparably narrow. This is in excellent agreement with Fig. 17. For a comparison we also display the f_{L} spectrum of B (bytes per time) on the bottom. Also here we find a left sided spectrum for the outgoing traffic and right sided spectra for the others.

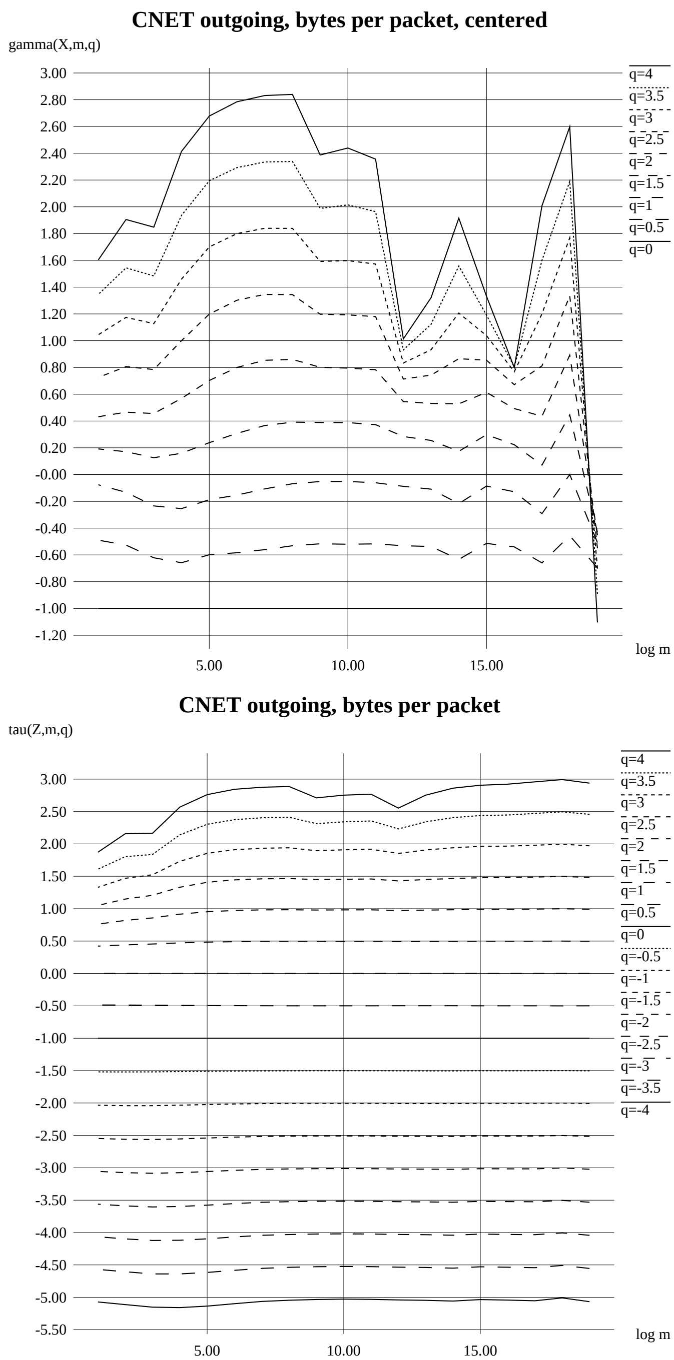


Figure 19: The superior scaling of non-centered data is most obvious when looking at the aspect B (bytes per packet) for the outgoing traffic at CNET. Here, as in the remaining figures, we show for various aspects A the statistical scaling of moments, i.e. $\gamma_m^{A'}(q) = \gamma(A', m, q) = \log_2 S_{2m}^{A'}(q) - \log_2 S_m^{A'}(q)$ for the centered aspect A' (here on top) as well as the multifractal scaling of moments, i.e. $\tau_m^A(q) = \log_2 S_{2m}^A(q) - \log_2 S_m^A(q)$ for the non-centered data A (here on the bottom). (For the ease of notation we write Z' instead of X here.)

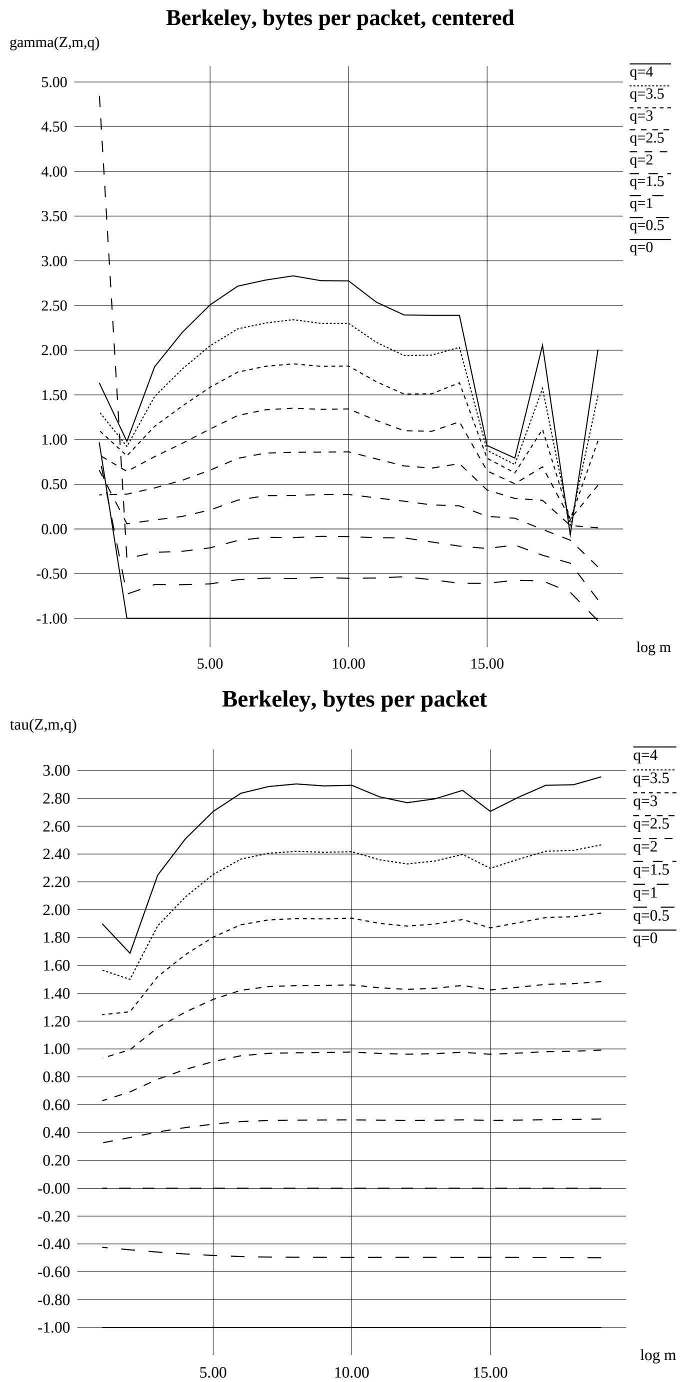
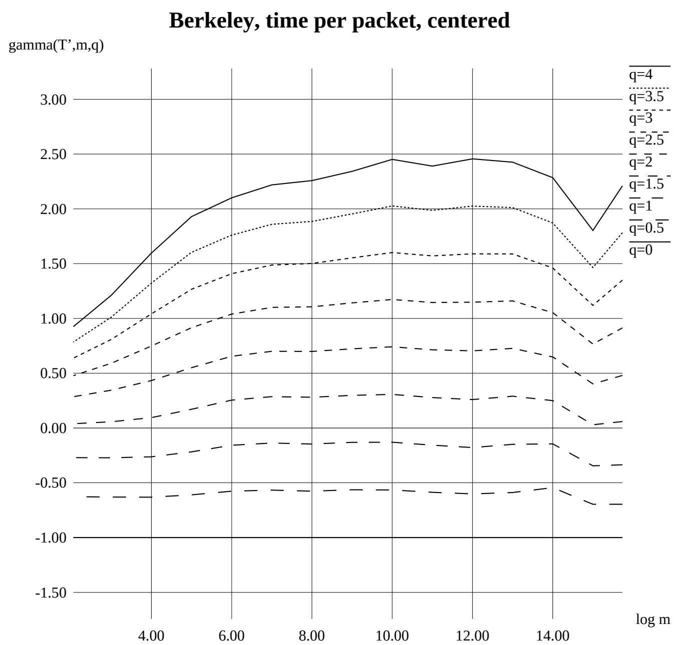


Figure 20: Statistical versus multifractal scaling of moments for the data considered in Fig. 8. Choosing the scaling region to be $2^4, \ldots, 2^{10}$ or $2^{10}, \ldots, 2^{15}$ yields the same estimate for the partition function $\tau(q)$ of the non-centered data, but introduces a drastic change in the estimation of $\gamma(q)$, the partition function of the centered data. (See Subsection 4.3.)



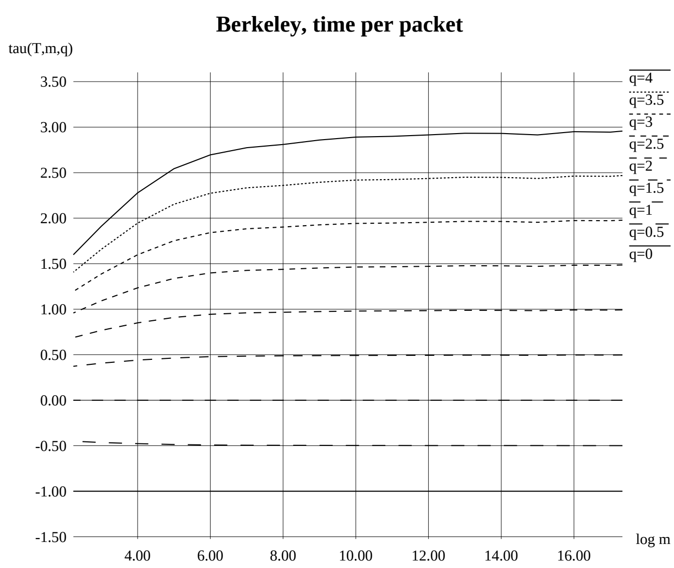


Figure 21: As with the aspect Z (bytes per packet) the statistical approach is found to be less robust than the multifractal analysis when studying the interarrival times of packets. (See Fig. 19 and 20.)

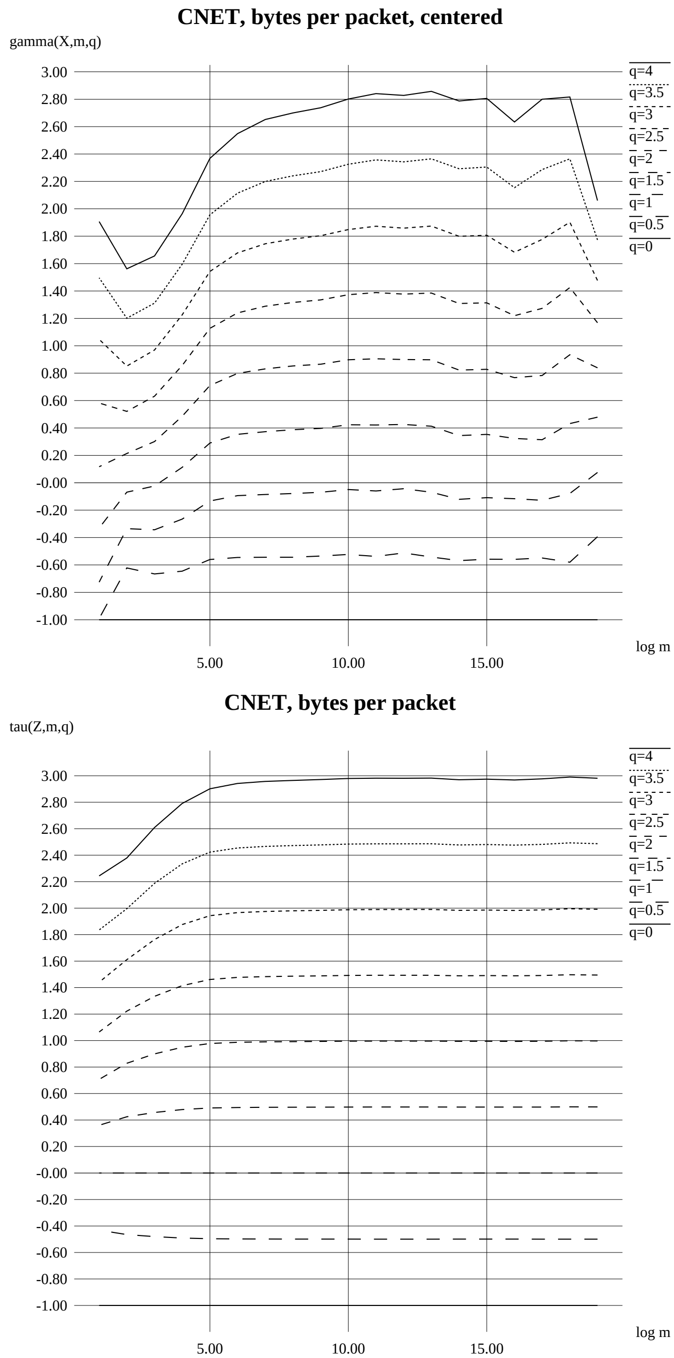


Figure 22: For the traffic at CNET labs, the same conclusions as above hold when comparing the statistical and the multifractal scaling of moments. Note that the partition functoin $\tau(q)$ found here is linear for $q > 0$ in agreement with the right sided spectrum shown in Fig. 17.