A MARKOVIAN POINT PROCESS EXHIBITING MULTIFRACTAL BEHAVIOR AND ITS APPLICATION TO TRAFFIC MODELING

ANDRÁS HORVÁTH
Dipartimento di Informatica, Università di Torino
Corso Svizzera 185, 10149 Torino, Italy, horvath@di.unito.it

MIKLÓS TELEK
Department of Telecommunications, Technical University of Budapest
Sztoczek u. 2, 1521 Budapest, Hungary, telek@hit.bme.hu

This paper introduces a set of Markovian Arrival Processes (MAPs) with a special structure exhibiting multifractal behavior. The considered MAP structure is motivated by the unnormalized Haar wavelet transform representation of finite sequences. A parameter fitting method is also proposed to approximate the multifractal behavior of experimental data sets by MAPs of the given structure. The goodness of the fitting method is evaluated via the log-moment diagrams, the partition function, the Legendre transform, and also by comparing the queue length distribution resulting from the measured data set with the one resulted from the approximating MAP.

1 Introduction

The traffic of high-speed communication networks, carrying the data packets of various applications, shows high variability and burstiness over several time scales (references to many measurement studies are provided by Willinger et al.[1]). The statistical analysis of some experimental traffic traces suggested a self-similar behavior over a range of time scales. Since measured data sets are finite (the large ones contain $10^6 - 10^8$ samples) the statistical properties of these data sets could be studied only over a range of time scales and the asymptotic behavior is determined from the range of known time scales.

The importance of the observed self-similar behavior lies in the fact that the queue length and the waiting time distribution of packet queues with self-similar traffic significantly differs from the ones with "regular traffic".

In the early 90's the research was focused on checking the self-similar behavior and the evaluation of the scaling parameter of self-similarity (referred to as Hurst parameter). It has to be noted that the majority of the practically applied statistical tests (e.g., variance-time plot, R/S plot) checks only the second order properties of the data sets and provides information only on the second order self-similarity of the analyzed data set. (Actually, there are

self-similar processes, like the fractional Gaussian noise [2], that are completely characterized by their first and second order behavior, but measured data sets are usually far more complex).

The first observations of self-similarity in measured traffic processes resulted in a fertile research for applying complex mathematical models in traffic engineering. The two main goals of these research efforts were to find "solvable" mathematical models with identical (or similar) properties and to create random sequence generators with predefined statistical properties. Some of the considered models are: fractional Gaussian noise [3,2], traditional [4] and fractional ARIMA processes [5], fractal [6] and Markovian models (MMPP, MAP) [7,8,9]. A valuable advantage of Markovian models is that effective numerical methods are available to analyze systems with Markovian arrival processes [10,11]. Furthermore, Markovian models also represent a simple and computationally effective way of generating random data series.

Unfortunately, some of the statistical properties of measured data sets differ from the ones of self-similar processes. The fact that the observed scaling behavior differs from the one of self-similar processes suggested the application of multifractal models to better capture the behavior of measured data sets [12]. A common approach to study multifractal models is wavelet analysis. Riedi et al. proposed a wavelet model to approximate the scaling behavior of measured data sets and based on this model they presented an algorithm to generate random sequences with similar scaling behavior [13]. The proposed method shows a good fit according to several statistical tests, but it is computationally rather expensive and does not allow any numerical analysis of queues.

Based on the mentioned advantages of Markovian models there is a need to approximate multifractal behavior with Markovian models. In this paper we propose Markovian models of a special structure to approximate the multifractal scaling behavior of measured data sets.

The flexibility of Markovian models in exhibiting complex stochastic behavior along the practically interesting time scales is known from previous works [8,7,9]. This paper attempts to extend this set of results with approximating multifractal models. The proposed MAP structure is motivated by the unnormalized Haar wavelet transform representation of finite sequences as it was applied in the multifractal wavelet model of Riedi et al. [13].

The rest of the paper is organized as follows. Section 2 summarizes the basic concepts of multiscale analysis. The proposed MAP structure is introduced in Section 3. Section 4 presents various properties of the proposed MAP structure and an example in fitting with measured data sets. The paper is concluded in Section 5.

2 Multiscale analysis

This section gives an overview of three methods which we will use in this paper to carry out the multiscale analysis of a stationary sequence of numbers $\mathcal{X} = \{X_i, i \geq 1\}$. As far as traffic modeling is concerned, this sequence may represent any kind of important characteristics of the traffic load arriving to the network. That is, $\mathcal{X}$ may be the series of interarrival times, the series of the number of arrivals in successive time-slots or the number of bytes per arrival. (See [14] for an exhaustive study of multifractal properties of time series describing different aspects of TCP traffic.)

In the sequel $\mathcal{X}^{(m)}$ denotes the corresponding aggregated process with level of aggregation m

$$\mathcal{X}^{(m)} = \left\{ X_i^{(m)}, i \geq 1 \right\} = \left\{ \frac{X_1 + \ldots + X_m}{m}, \ldots, \frac{X_{(i-1)m+1} + \ldots + X_{mi}}{m}, \ldots \right\}.$$

This section is organized as follows. First we look at statistical scaling in Section 2.1 that aims at determining the Hurst parameter of the process. In Section 2.2 a method to analyze multifractal scaling, that results in the Legendre spectrum, is described with which we will compare our approximating MAP with the real traffic trace. The fitting procedure is based on another way of examining a finite sequence of numbers, the Haar wavelet transform (Section 2.3), because it allows us to compute the desired properties of our MAP structure analytically.

The introduction to statistical and multifractal scaling given hereinafter is partly based on [15] and [14].

2.1 Statistical scaling

Recently, it has been agreed [16,17,2] that when one studies a traffic trace the most significant parameter to be estimated is the degree of self-similarity, usually given by the so-called Hurst-parameter. The aim of the statistical approach, based on the theory of self-similarity, is to find the Hurst-parameter.

The standard definition of self-similarity is stated for continuous-time processes: $\mathcal{Y} = \{Y(t), t > 0\}$ is self-similar if

$$Y(t) \stackrel{d}{=} a^{-H} Y(at), \forall t > 0, \forall a > 0, 0 < H < 1, \tag{1}$$

where $\stackrel{d}{=}$ denotes equality in the sense of finite-dimensional distributions and H is the Hurst-parameter. The most broadly applied signal model satisfying (1) is the fractional Brownian motion [18,17,2] whose power lies in its simple

parameterization. It is fully determined by its mean, variance and the Hurst-parameter.

There are several, different definitions of self-similarity involving stationary sequences $\mathcal{X} = \{X_i, i \geq 1\}$; in the context of traffic modeling these are more appropriate than the one given by (1). A stationary discrete-time stochastic process $\mathcal{X} = \{X_i, i \geq 1\}$ is said to be *exactly self-similar* if

$$\mathcal{X} \stackrel{d}{=} m^{1-H}\mathcal{X}^{(m)} \tag{2}$$

for all aggregation levels m. In other words $\mathcal{X}$ is said to be exactly self-similar if $\mathcal{X}$ and $\mathcal{X}^{(m)}$ are identical within a scale factor in the sense of finite-dimensional distributions. (We note here that if $\mathcal{X}$ is the incremental process of an exactly self-similar continuous-time process it satisfies (2) for all aggregation levels m.) A stationary sequence is said to be *asymptotically self-similar* if (2) holds as $m \to \infty$. A covariance-stationary sequence $\mathcal{X}$ is *exactly second-order self-similar* or *asymptotically second-order self-similar* if $m^{1-H}\mathcal{X}^{(m)}$ has the same variance and auto-correlation as $\mathcal{X}$ for all aggregation level m, or as $m \to \infty$.

As it was proposed in [15] one may perform a test of self-similarity by analyzing the behavior of the absolute moments of the aggregated process. If $\mathcal{X}$ is exactly self-similar

$$log(\mathbf{E}(|X^{(m)}|^q)) = log(\mathbf{E}(|m^{H-1}X|^q)) = q(H-1)log(m) + log(\mathbf{E}(|X|^q)). \tag{3}$$

According to (3), in case of a self-similar process plotting $log(\mathbf{E}(|X^{(m)}|^q))$ against $log(m)$ for a fixed q results in a straight line. The slope of the line is $q(H-1)$. Based on the above observations the test is performed as follows. Having a series of length N, the moments may be estimated as

$$\mathbf{E}(|X^{(m)}|^q) = \frac{1}{\lfloor N/m \rfloor} \sum_{i=1}^{\lfloor N/m \rfloor} |X_i^{(m)}|^q,$$

where $\lfloor x \rfloor$ denotes the largest integer number smaller or equal to x. To test for self-similarity $log(\mathbf{E}(|X^{(m)}|^q))$ is plotted against $log(m)$ and a straight line is fitted to the curve. If the straight line shows good correspondence with the curve, then the process is self-similar and its Hurst-parameter may be calculated by the slope of the straight line.

It is worth pointing out that (2) and stationarity imply that either $\mathbf{E}(X) = 0$, or $\mathbf{E}(X) = \pm\infty$, or $H = 1$. But $H = 1$ implies as well that $X_i = X_j$, $\forall i, j$ almost surely. As a consequence, to test for statistical self-similarity makes sense only having zero-mean data, i.e., the data has to be centered before the

analysis. The variance-time plot, which is used widespread to gain evidence of self-similarity, is the special case with $q = 2$. It depicts the behavior of the 2nd moments for the centered data. On the other hand, as we show later, multifractal analysis may be carried out on data with non-zero mean as well.

2.2 Multifractal scaling

As it is described above, statistical tests of self-similarity try to gain evidence through examining the behavior of the absolute moments $\mathbf{E}(|X^{(m)}|^q)$. Multifractal analysis looks at the behavior of the absolute moments as well, but in a different manner which may result in more detailed information on the sequence. While the above described statistical view looks for a single number, the Hurst parameter, that completely describes the behavior of $\mathbf{E}(|X^{(m)}|^q)$ for any q, multifractal analysis results in a *spectrum* to illustrate the behavior of the absolute moments.

As for self-similarity we start the discussion with a continuous-time process $\mathcal{Y} = \{Y(t), t > 0\}$. The scaling of the absolute moments of the increments is observed through the *partition function*

$$T(q) = \lim_{n\to\infty} \frac{1}{-n} \log_2 \mathbf{E}\left[\sum_{k=0}^{2^n-1} |Y((k+1)2^{-n}) - Y(k2^{-n})|^q\right]. \qquad (4)$$

Then, a multifractal spectrum, the so-called *Legendre spectrum* is given as the *Legendre transform* of (4) by

$$f_L(\alpha) = T^*(\alpha) = \inf_q (q\alpha - T(q))$$

Since $T(q)$ is always concave, the Legendre spectrum $f_L(\alpha)$ may be found by simple calculations using

$$T^*(\alpha) = q\alpha - T(q), \text{ and } (T^*)'(\alpha) = q \text{ at } \alpha = T'(q). \qquad (5)$$

Let us mention here that there are also other kinds of fractal spectrum defined in the fractal world (see for example [19]). The Legendre spectrum is the most attractive one from numerical point of view, and even though in some cases it is less informative than, for example, the large deviation spectrum, it provides enough information in the cases considered herein.

In the case of a discrete-time process $\mathcal{X}$ we assume that we are given the increments of a continuous-time process. This way, assuming that the sequence we examine consists of $N = 2^L$ numbers, the sum in (4) becomes

$$S_n(q) = \sum_{k=0}^{N/2^n-1} |X_k^{(2^n)}|^q, \ 0 \le n \le L, \qquad (6)$$

where the expectation is ignored. Ignoring the expectation is accurate for small n, i.e., for the finer resolution levels. In order to estimate $T(q)$, we plot $\log_2(S_n(q))$ against $(L-n)$, $n = 0, 1, ..., L$, then $T(q)$ is found by the slope of the straight line fitted to the plotted points. If the straight line shows good correspondence with the plotted points, that is, if $\log_2(S_n(q))$ scales linearly with $\log(n)$, then the sequence $\mathcal{X}$ can be considered a multifractal process.

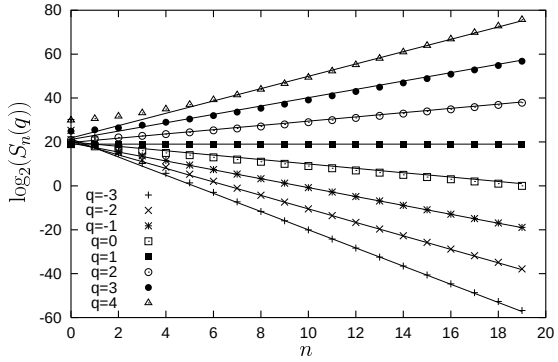

Figure 1. Scaling of log-moments with linear fits for the interarrival times of the Bellcore *pAug* trace

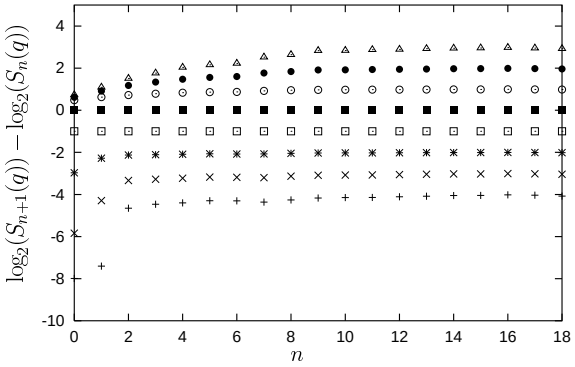

Figure 2. Increments of log-moments for the interarrival times of the Bellcore *pAug* trace

Figure 1, 2 (the same legend applies to this figure as to Figure 1), 3 and 4 illustrate the above described procedure to obtain the Legendre spectrum

of the famous Bellcore *pAug* traffic trace (the trace may be found at [20]). Figure 1 depicts the scaling behavior of the log moments calculated through (6). With q in the range $[-3, 4]$, excluding the finest resolution levels $n = 0, 1$ the moments show good linear scaling. For values of q outside the range $[-3, 4]$ the curves deviate more and more from linearity. As, for example, in [13] one may look at non-integer values of q as well, but, in general, it does not provide notably more information on the process. To better visualize the deviation from linearity Figure 2 depicts the increments of the log-moment curves of Figure 1. Completely horizontal lines would represent linear log-moment curves.

The partition function $T(q)$ is depicted in Figure 3. The three slightly different curves differ only in the considered range of the log-moments curves, since different ranges result in different linear fitting. The lower bound of the linear fitting is set to 3, 5 and 7, while the upper bound is 18 in each case. (In the rest of this paper the fitting range is 5 - 18 and there are 200 moments evaluated in the range $[-5, +5]$.) Since the partition function varies only a little (its derivative is in the range $[0.8, 1.15]$), it is not as informative as its Legendre transform is (Figure 4). According to (5) the Legendre spectrum is as wide as the range of derivatives of the partition function is. That is, the more the partition function deviates from linearity the wider the Legendre spectrum is. The Legendre transform significantly amplifies the scaling information, but it is also sensitive to the considered range of the log-moments curves.

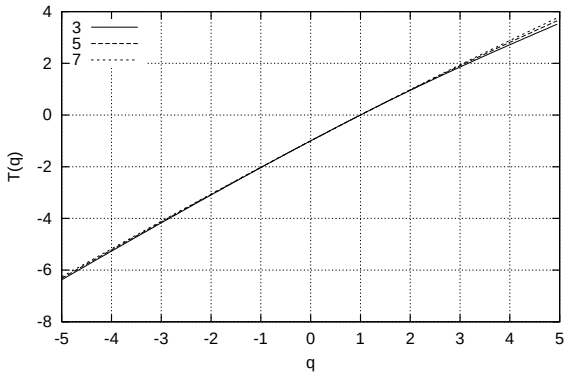

Figure 3. Partition function estimated through the linear fits shown in Figure 1

See [13] for basic principles of interpreting the spectrum. We mention here

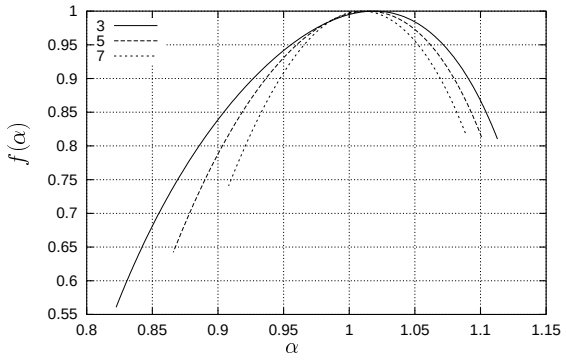

Figure 4. The Legendre transform of the partition function (Figure 3) results in the Legendre spectrum

only that a curve like the one depicted in Figure 4 reveals a *rich multifractal spectrum*. On the contrary, as it was shown in [21], the fractional Brownian motion (fBm) has a trivial spectrum. The partition function of the fBm is a straight line which indicates that its spectrum consists of one point, i.e., the behavior of its log-moments is identical for any q.

2.3 The unnormalized Haar wavelet transform

The third way we mention here to carry out multiscale analysis is the Haar wavelet transform. The choice of using the *unnormalized* version of the Haar wavelet transform is motivated by the fact that it suits more the analysis of the Markovian point process introduced further on.

The multiscale behavior of the finite sequence $X_i, 1 \leq i \leq 2^L$ will be represented by the quantities $c_{j,k}, d_{j,k}, j = 0, \ldots, L$ and $k = 1, \ldots, 2^L/2^j$. The finest resolution is described by $c_{0,k}, 1 \leq k \leq 2^L$ which gives the finite sequence itself, i.e., $c_{0,k} = X_k$. Then the multiscale analysis based on the unnormalized Haar wavelet transform is carried out by iterating

$$c_{j,k} = c_{j-1,2k-1} + c_{j-1,2k}, \tag{7}$$

$$d_{j,k} = c_{j-1,2k-1} - c_{j-1,2k}, \tag{8}$$

for $j = 1, \ldots, L$ and $k = 1, \ldots, 2^L/2^j$. The quantities $c_{j,k}, d_{j,k}$ are the so-called scaling and wavelet coefficients of the sequence, respectively, at scale j and position k. The procedure to obtain the scaling and wavelet coefficients

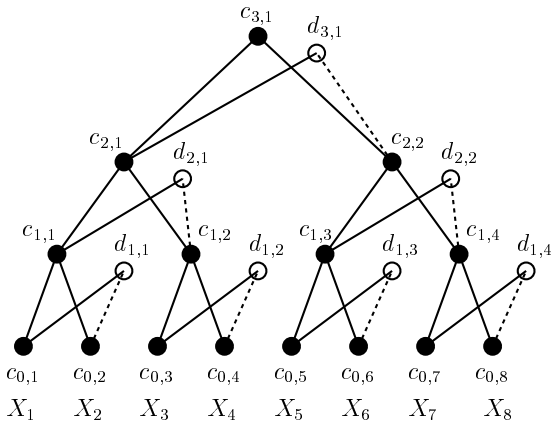

Figure 5. Haar wavelet transform

is depicted in Figure 5 for a series of length 8. In the figure, the scaling (wavelet) coefficients are drawn as filled (empty) circles, and solid (dashed) lines are connected to quantities that are taken into account with positive (negative) sign. At each scale the coefficients are represented by the vectors $\mathbf{c}_j = [c_{j,k}]$ and $\mathbf{d}_j = [d_{j,k}]$ with $k = 1, \ldots, 2^L/2^j$. For what concerns $\mathbf{c}_j$, the higher j the lower the resolution level at which we have information on the sequence. The information that we lose as a result of the step from $\mathbf{c}_{j-1}$ to $\mathbf{c}_j$, is conveyed by the sequence of wavelet coefficients $\mathbf{d}_j$. It is easy to see that $\mathbf{c}_{j-1}$ can be perfectly reconstructed from $\mathbf{c}_j$ and $\mathbf{d}_j$. As a consequence the whole $X_i, 1 \leq i \leq 2^L$ sequence can be constructed (in a top to bottom manner) based on a normalizing constant, $\mathbf{c}_L = c_{L,1} = \sum_{i=1}^{2^L} X_i$, and the $\mathbf{d}_j, j = 1, \ldots, L$ vectors.

By taking the expectation of the square of (7) and (8) we get

$$E[c_{j,k}^2] = E[c_{j-1,2k-1}^2] + 2E[c_{j-1,2k-1}c_{j-1,2k}] + E[c_{j-1,2k}^2], \tag{9}$$

$$E[d_{j,k}^2] = E[c_{j-1,2k-1}^2] - 2E[c_{j-1,2k-1}c_{j-1,2k}] + E[c_{j-1,2k}^2]. \tag{10}$$

Let us assume that the series we analyze is stationary[a]; then, by summing (9)

[a] Clearly, since we are given only a finite trace, we cannot have strong evidence that the series under analysis is obtained from a stationary process. However, we apply this assumption

and (10) and rearranging the equation, we have

$$E[c_{j-1}^2] = \frac{1}{4}\left(E[d_j^2] + E[c_j^2]\right). \tag{11}$$

Similarly, by consecutive application of (11) from one scale to another the $E[d_j^2], j = 1, \ldots, L$ series completely characterize the variance decay of the $X_i, 1 \leq i \leq 2^L$ sequence apart from a normalizing constant ($\mathbf{c}_L = c_{L,1} = \sum_{i=1}^{2^L} X_i$). This fact allows us to realize a series with a given variance decay if it is possible to control the 2nd moment of the scaling coefficient with the chosen synthesis procedure. This is why the $E[d_j^2], j = 1, \ldots, L$ series plays an important role in the subsequent discussion. Basically, we attempt to capture the multifractal scaling behavior via this series.

3 The proposed MAP structure

To exhibit multifractal behavior we propose to apply a special MAP, a Markov modulated Poisson process (MMPP) whose background CTMC has a symmetric[b] n-dimensional cube structure and the arrival intensities are set according to the variation of the arrival process at the different time scales. We believe that other MAP structures can also exhibit multifractal behavior. Our special choice is motivated by the generation of the Haar wavelet transform. Basically the Haar wavelet transform evaluates the variation of the data set at different aggregation levels (time scales), and similarly, the proposed MAP structure provides different variation of the arrival rate at different time scales.

The composition of the proposed MAP structure follows a very similar pattern as the generation of the Haar wavelet transform. Without loss of generality, we assume that the time unit is such that the long term arrival intensity is one. A MAP of one state with arrival rate 1 represents the arrival process at the largest (considered) time scale.

At the next time scale, $1/\lambda$, an MMPP of two states with generator

$$\begin{bmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{bmatrix}$$

and with arrival rates $1 - a_1$ and $1 + a_1$ ($-1 \leq a_1 \leq 1$) represents the variation of the arrival process. This composition leaves the long term average arrival rate unchanged.

---

because it simplifies the discussion significantly.

[b] We also investigated the effect of applying asymmetric n-dimensional cubes. According to our experience, the asymmetric models perform similarly to the symmetric ones but have more parameters.

In the rest of the composition we perform the same step. We introduce a new dimension and generate the n-dimensional cube such that the behavior at the already set time scales remains unchanged. E.g., considering also the $1/\gamma\lambda$ ($\gamma > 1$) time scale an MMPP of four states with generator

$$\begin{array}{|cc|cc|}
\hline
\bullet & \lambda & \gamma\lambda & \\
\lambda & \bullet & & \gamma\lambda \\
\hline
\gamma\lambda & & \bullet & \lambda \\
& \gamma\lambda & \lambda & \bullet \\
\hline
\end{array}$$

and with arrival rates $(1-a_1)(1-a_2)$, $(1+a_1)(1-a_2)$, $(1-a_1)(1+a_2)$ and $(1+a_1)(1+a_2)$ ($-1 \leq a_1, a_2 \leq 1$) represents the variation of the arrival process. With this MMPP, parameter a_1 (a_2) determines the variance of the arrival process at the $1/\lambda$ ($1/\gamma\lambda$) time scale. If γ is large enough ($>\sim 30$) the process behavior at the $1/\lambda$ time scale is independent of a_2. The proposed model is also applicable with a small γ. In this case, the only difference is that the model parameters and the process behavior of different time scales are dependent.

Finally, Figure 6 introduces the MAP structure with 3 levels. Following the same construction rule one can compose any high level MAP structure of this kind.

$$\begin{array}{|cc|cc|cccc|}
\hline
\bullet & \lambda & \gamma\lambda & & \gamma^2\lambda & & & \\
\lambda & \bullet & & \gamma\lambda & & \gamma^2\lambda & & \\
\hline
\gamma\lambda & & \bullet & \lambda & & & \gamma^2\lambda & \\
& \gamma\lambda & \lambda & \bullet & & & & \gamma^2\lambda \\
\hline
\gamma^2\lambda & & & & \bullet & \lambda & \gamma\lambda & \\
& \gamma^2\lambda & & & \lambda & \bullet & & \gamma\lambda \\
\hline
& & \gamma^2\lambda & & \gamma\lambda & & \bullet & \lambda \\
& & & \gamma^2\lambda & & \gamma\lambda & \lambda & \bullet \\
\hline
\end{array}$$

$$\begin{array}{|l|}
\hline
(1-a_1)(1-a_2)(1-a_3) \\
(1+a_1)(1-a_2)(1-a_3) \\
(1-a_1)(1+a_2)(1-a_3) \\
(1+a_1)(1+a_2)(1-a_3) \\
(1-a_1)(1-a_2)(1+a_3) \\
(1+a_1)(1-a_2)(1+a_3) \\
(1-a_1)(1+a_2)(1+a_3) \\
(1+a_1)(1+a_2)(1+a_3) \\
\hline
\end{array}$$

Figure 6. The generator matrix of the proposed MAP structure with 3 levels and the associated arrival rates

A level n MAP of the proposed structure is composed by 2^n states and has $n+2$ parameters. Parameters γ and λ define the considered time scales, and parameters $a_1, a_2, \ldots, a_n$ determine the variance of the arrival process at the n considered time scales. It can be seen that the ratio of the largest and the smallest considered time scales is γ^n. Having a fixed n (i.e., a fixed cardinality of the MAP), any large ratio of the largest and the smallest considered time scales can be captured by using a sufficiently large γ.

3.1 Analysis of the proposed MAP structure

In the case of an m-phase MMPP with descriptors $\mathbf{D_0}$ and $\mathbf{D_1}$, the distribution of the sum of k consecutive interarrival times can be viewed as PH distribution of order mk whose descriptors are the following. In order to describe the initial probability vector of the PH distribution we need the MAPs stationary probability vector embedded at arrival epochs which can be obtained by

$$\boldsymbol{\tau} = (\boldsymbol{\pi}\mathbf{D_1}\mathbf{e})^{-1}\boldsymbol{\pi}\mathbf{D_1} \tag{12}$$

where $\boldsymbol{\pi}$ is the stationary probability vector of the CTMC with infinitesimal generator $\mathbf{D_0}+\mathbf{D_1}$, and $\mathbf{e}$ is a vector of ones. Using (12) the initial probability vector of length $m \cdot k$ is given by

$$\mathbf{t}_{(k)} = [\boldsymbol{\tau}\ 0 \cdots 0],$$

and the generator matrix $\mathbf{T}$ that describes the state-transitions among the transient states is of size $mk \times mk$ and is given by

$$\mathbf{T}_{(k)} = \begin{bmatrix} \mathbf{D_0} & \mathbf{D_1} & \mathbf{0} & \cdots & \\ \mathbf{0} & \mathbf{D_0} & \mathbf{D_1} & \mathbf{0} & \cdots \\ & & \ddots & \ddots & \\ & \cdots & \mathbf{0} & \mathbf{D_0} & \mathbf{D_1} \\ & & \cdots & \mathbf{0} & \mathbf{D_0} \end{bmatrix}.$$

Let us denote by $Z_{(k)}$ a PH distributed random variable with descriptors $\mathbf{t}_{(k)}$ and $\mathbf{T}_{(k)}$. Then, applying (10), the second moment of the wavelet coefficients can be calculated as

$$E[d_j^2] = 2E[Z^2_{(2^{j-1})}] - 2\sum_{i=1}^{km} Pr(z_l = i)E[Z_{(2^{j-1})}|z_l = i]E[Z_{(2^{j-1})}|z_0 = i]$$

where z_l denotes the last transient phase before absorption while z_0 denotes the initial phase. The first term of the right hand side can be obtained by

$$E[Z^2_{(2^{j-1})}] = 2\mathbf{t}_{(2^{j-1})}\mathbf{T}^{-2}_{(2^{j-1})}\mathbf{e},$$

while the second term can be calculated based on

$$Pr(z_l = i)E[Z_{(2^{j-1})}|z_l = i] = -\mathbf{t}_{(2^{j-1})}\mathbf{T}^{-2}_{(2^{j-1})}\mathbf{T}_{(2^{j-1})}\mathbf{e_i},$$

where $\mathbf{e_i}$ denotes the vector whose only nonzero entry is 1 at position i, and

$$E[Z_{(2^{j-1})}|z_0 = i] = -\mathbf{e_i}\mathbf{T}^{-1}_{(2^{j-1})}\mathbf{e}.$$

In order to compute the quantities in the above expressions one has to determine some entries of $\mathbf{T}_{(k)}^{-1}$. Note that this can be done easily by performing computations on matrices of size $m \times m$ since

$$\mathbf{T}_{(k)}^{-1} =$$

$$\begin{bmatrix} \mathbf{D_0}^{-1} & -\mathbf{D_0}^{-1}\mathbf{D_1}\mathbf{D_0} & \mathbf{D_0}^{-1}(\mathbf{D_1}\mathbf{D_0})^2 & \cdots & (-1)^{k-1}\mathbf{D_0}^{-1}(\mathbf{D_1}\mathbf{D_0})^{k-1} \\ \mathbf{0} & \mathbf{D_0}^{-1} & -\mathbf{D_0}^{-1}\mathbf{D_1}\mathbf{D_0} & \cdots & (-1)^{k-2}\mathbf{D_0}^{-1}(\mathbf{D_1}\mathbf{D_0})^{k-2} \\ \mathbf{0} & \mathbf{0} & \mathbf{D_0}^{-1} & \cdots & \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{D_0}^{-1} \end{bmatrix}.$$

3.2 A parameter fitting method

We apply a simple numerical procedure to fit a MAP of the given structure with a measured data set. Actually, our heuristic approach is composed by "engineering considerations" based on the properties of the measured data set and a parameter fitting method.

First, we fix the value of n. According to our experience a "visible" multiscaling behavior can be obtained from $n = 3 \sim 4$. The computational complexity of the fitting procedure grows exponentially with the dimension of the MAP. The response time with $n = 6$ (MAP of 64 states) is still acceptable (in the order of minutes).

Similar to [13], we set the γ and the λ parameters based on the inspection of the data set. Practically, we define the largest, T_M, and the smallest, T_m, considered time scales and calculate γ and λ from

$$T_M = \frac{1}{\lambda}\ ; \qquad T_m = \frac{1}{\gamma^n \lambda},$$

where $\gamma > 1$.

The extreme values of T_M and T_m can be set based on simple practical considerations. For example when the measured data set is composed of N arrival instances, T_M can be chosen to be less than the mean time of $N/4$ arrivals, and T_m can be chosen to be greater than the mean time of 4 arrivals. A similar approach was applied in [13]. These boundary values can be refined based on a detailed statistical test of the data set. For example, if the scaling behavior disappears beyond a given time scale T_M can be set to that value.

Having the γ and the λ parameters we apply a downhill simplex method to find the optimal values of the variability parameters $a_1, a_2, \ldots, a_n$. The goal function that our parameter fitting method minimizes is the sum of the relative errors of the second moment of Haar wavelet coefficients up to a

predefined time scale S:

$$\min_{a_1,\ldots,a_n} \sum_{j=1}^{S} \frac{|E[d_j^2] - E[\hat{d}_j^2]|}{E[d_j^2]} .$$

4 Numerical analysis

This section presents a collection of numerical analysis results using the proposed Markovian model. The first subsection investigates the multifractal scaling properties of the considered MAP structure. The second subsection presents the comparison of the Bellcore *pAug* trace and its approximating MAP.

4.1 Multifractal scaling properties of the proposed MAP structure

The logarithmic moments of the MAP structure

As mentioned in Section 2 the scaling behavior of the data samples is usually checked using log-moment diagrams. The log-moment diagram plots the logarithm of different moments of the m aggregated process against $log(m)$. Linear curves in the log-moments plot suggest scaling behavior. Figures 7 and 9 show the log-moment plots of the MAPs with the following sets of parameters:

- $n = 5, \lambda = 1/2^{18}, \gamma = 8, a_1 = a_2 = a_3 = a_4 = a_5 = 0.3$
- $n = 5, \lambda = 1/2^{18}, \gamma = 8, a_1 = a_2 = a_3 = a_4 = a_5 = 0.5$

In the range of 4 to 18 the curves are very close to linear, as it can be seen from the increment plots in Figures 8 and 10. Based on Figures 8 and 10, we conclude that the considered MAP exhibits scaling behavior over the time scales from 4 to 18.

Mono-fractal scaling is assumed when the slope of the log-moment curve is linear with q and a non-linear relation suggests multifractal scaling. The slope of the log-moment curves of the two MAPs are collected in Table 1. None of the considered MAPs show a linear relation, but they are not too far from that. The slopes of the MAP with larger variability parameters are further from linearity.

The other test of multifractal scaling considered is the analysis of the partition function and its Legendre transform. The visual inspection of a partition function curve (e.g., in Figure 3) is very hard because the slope of the

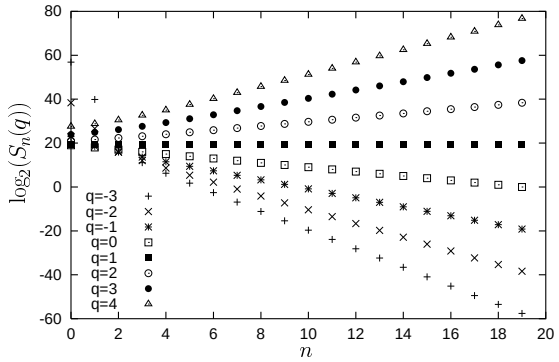

Figure 7. Scaling of log-moments for $a_i = 0.3$

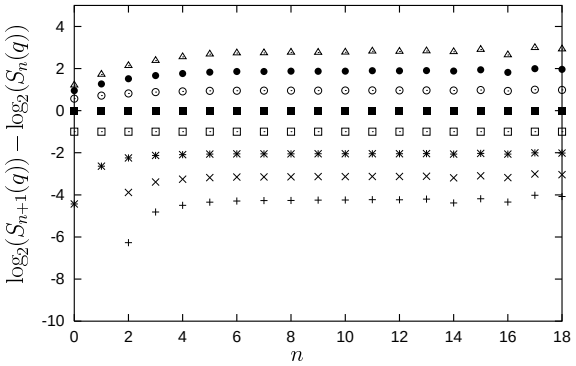

Figure 8. Increments of the log-moment curves for $a_i = 0.3$

curve carries the required information that is very hard to see. The Legendre transform of the partition function somehow amplifies the important information. The differences of the scaling behavior of two multifractal processes are much more perceptible from the Legendre transform. It is analyzed in the following subsection.

The effect of variability parameters on the Legendre spectrum

Legendre transform presents the scaling behavior of a multifractal process in a way which is not closely related to the "physical understanding" of the

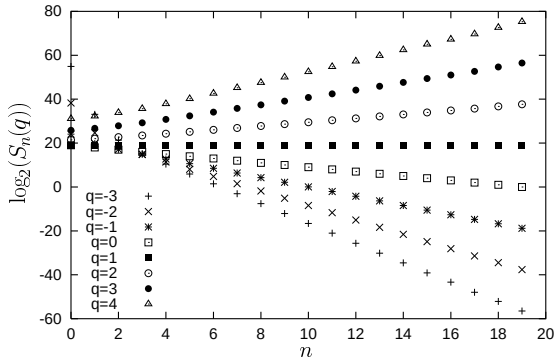

Figure 9. Scaling of log-moments for $a_i = 0.5$

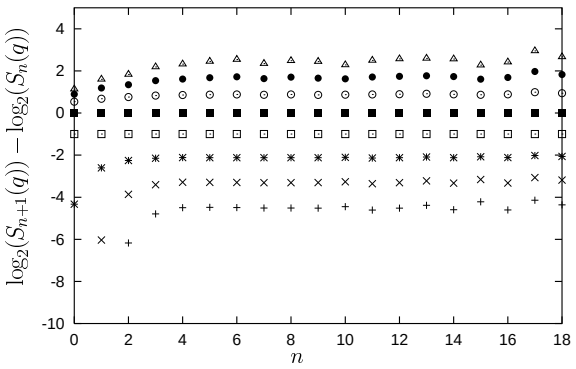

Figure 10. Increments of the log-moment curves for $a_i = 0.5$

process. It is useful in visualizing the scaling behavior, but it is not easy to interpret. We should emphasize again that according to the definition of the partition function (4) and its Legendre transform only infinite data sets can be analyzed. In our experiments we always use finite data sets because the measured traffic samples are finite and the considered MAPs exhibit scaling behavior only through a range of time scales. Using finite data sets we approximate the partition function according to (6). The $n \to \infty$ limiting behavior is approximated via the linear fit of the function over the available time scales. According to our experiment the range of time scales for which the linear line

Table 1. The slope of the log-moments

q	$a_i = 0.3$	$a_i = 0.5$
-5	-6.5319	-6.8912
-4	-5.3821	-5.6784
-3	-4.2457	-4.4719
-2	-3.1298	-3.2766
-1	-2.0446	-2.1076
0	-1.0000	-1.000
1	0	0
2	0.9602	0.8895
3	1.8904	1.7085
4	2.7995	2.4900
5	3.6498	3.2132

is fitted significantly affects the Legendre transform curves (as it is presented in 4). We found that the Legendre transform curves are comparable only when the same range of time scales are considered. Due to this fact in the following set of results the $n = 4, \lambda = 1/2^{15}, \gamma = 8$ parameters, as well as the range of time scales of the linear fit are fixed. We only analyze the effect of the variability parameters, $a_i, i = 1, \ldots, 4$, on the Legendre spectrum.

Figure 11 shows the Legendre transform of the considered MAP structure with uniform variability parameters. It can be seen that low variability, $a_i = 0.1, i = 1, \ldots, 4$, results in a narrow Legendre transform and large variability, $a_i = 0.9, i = 1, \ldots, 4$, results in a wide one. Theoretically, the Legendre transform of a mono-fractal is a single point, since a mono-fractal has the same scaling parameter for each moment. On the other hand, a wide Legendre transform curve represents a "rich multifractal spectrum". Note that our MAP structure results in a Poisson process when $a_i = 0, i = 1, \ldots, 4$, which is not a scaling process (its log-moment curves are not linear).

Figure 12 - 14 display the effect of different variability patterns on the Legendre transform. In Figure 12 the average of the variability parameters is fixed, $\sum_{i=1}^{4} a_i/4 = 0.5$, and the variability parameters form an arithmetic sequence with different increments. In Figure 13 the a_2, a_3, a_4 parameters are fixed to 0.5 and the variability parameter associated with the slowest time scale, a_1, is changed. In Figure 14 $a_1 = a_2 = a_3 = 0.5$ and a_4 varies. From Figures 12 - 14, one can conclude that the variability at the slowest time scale effects the width of the Legendre transform curve. The higher a_1 is the wider the Legendre transform curve is. According to Figure 14 the variability

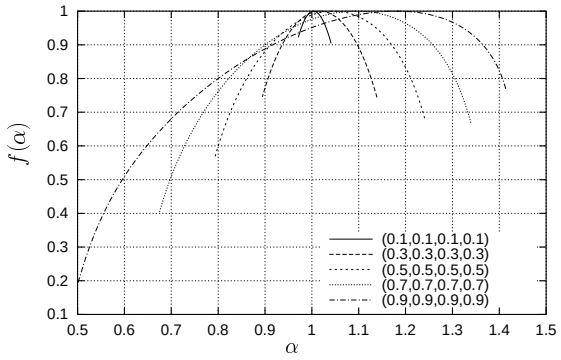

Figure 11. Effect of increasing variability at each time scale

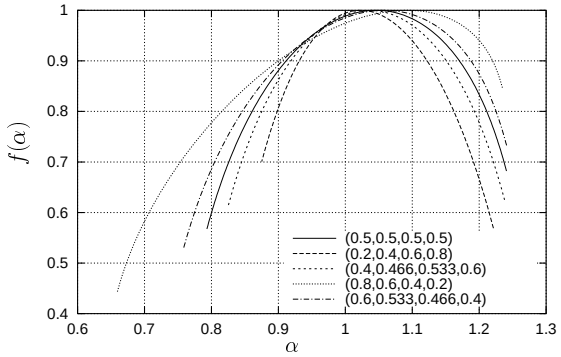

Figure 12. Effect of different variability patterns on the Legendre spectrum

parameter of the fastest time scale does not affect the width of the Legendre transform too much. Instead it turns the curve a bit.

4.2 Approximating the Bellcore pAug *trace*

To test the properties of the proposed MAP structure and the fitting method, we fit a MAP to the famous Bellcore *pAug* trace [20]. This data set is composed of $10^6 \sim 2^{20}$ interarrival times. We applied the fitting method with $n = 5$ and several different predefined setting of γ, λ. We found that the goodness of the fitting is not very sensitive to the predefined parameters around the

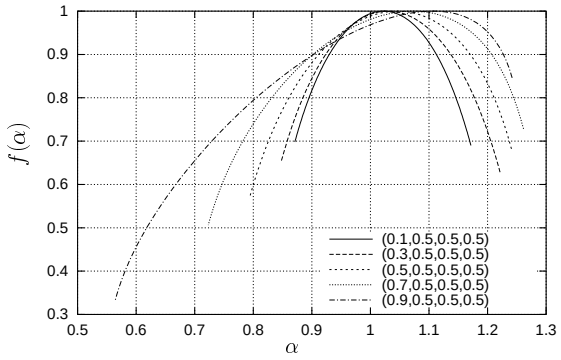

Figure 13. Effect of increasing variability at each time scale

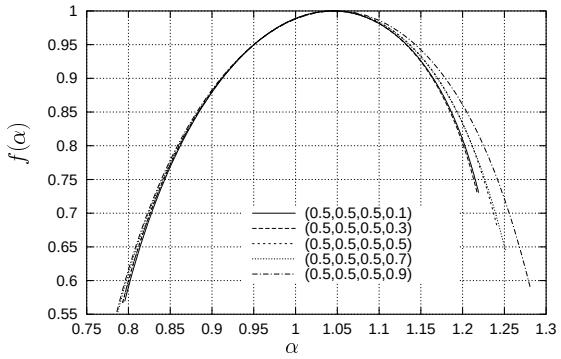

Figure 14. Effect of different variability patterns on the Legendre spectrum

reasonable region. The best "looking" fit is obtained when T_m is the mean time of 16 arrivals and $\gamma = 8$. In this case T_M is the mean time of $16 * 8^5 = 2^{19}$ arrivals which corresponds to the coarsest time scale we can analyze in the case of the Bellcore *pAug* trace. The simplex method minimizing the relative error of the second moment of the Haar wavelet coefficients over $S = 12$ time scales resulted in: $a_1 = 0.144, a_2 = 0.184, a_3 = 0.184, a_4 = 0.306, a_5 = 0.687$. The result of fitting the second moment of the Haar wavelet transform at different aggregation levels is plotted in Figure 15. Since our fitting method intends to minimize a sum of these differences the obtained small differences come from the structural limits of the applied MAP with the given fixed

parameters. At small time scales the fitting seems to be perfect. There is only a small oscillation of the curves. At larger time scales the oscillation seems to enlarge. The slope of the curves are almost equal in the depicted range. (Note that the $E[d_k^2]$ is also increasing with the aggregation level.)

First, we compared the multiscaling behavior of the obtained MAP with the one of the original data set via the log-moment curves. Figure 16 depicts the logarithm of different moments of the aggregated process, $\log_2(S_n(q))$, as a function of the aggregation level, n. In the figure, the symbols represent the log-moment curves of the fitting MAP and the solid lines indicate the corresponding log-moment curves of the Bellcore *pAug* trace. In the range of $n \in (3, 19)$ the log-moment curves of the fitting MAP are very close to the ones of the original trace. The log-moment curves of the approximate MAP are also very close to linear in the considered range.

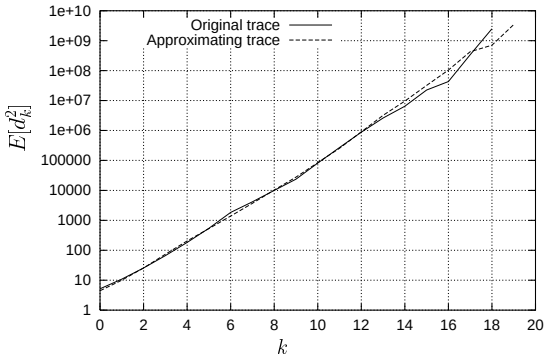

Figure 15. The second moment of the Haar wavelet transform at different aggregation levels

The partition functions of the fitting MAP and the original trace are depicted in Figure 17. As it is mentioned in the previous section, the visual appearance of the partition function is not very informative about the multifractal scaling behavior. Figure 18 depicts the Legendre transform of the partition functions of the original data set and the approximating MAP. The visual appearance of the Legendre transform significantly amplifies the differences of the partition functions. In Figure 18, it can be seen that both processes exhibit multifractal behavior but the original data set has a bit richer multifractal spectrum. The difference of the Legendre transforms comes from the differences of the high negative and high positive moments (< -3 and > 4), which are not provided in Figure 16. The reason why the Legendre

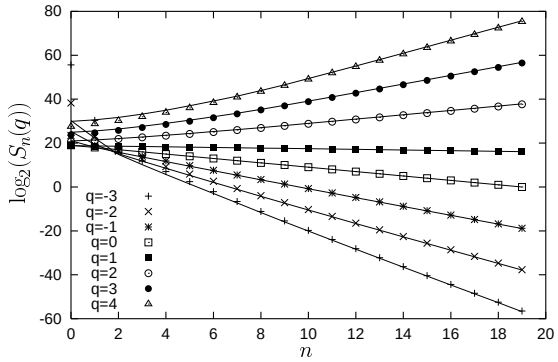

Figure 16. Scaling of log-moments of the original trace and the fitting MAP

transform of the approximating trace seems to be the rotation of the Legendre spectrum of the original trace can be found examining the partition function. One can observe in Figure 17 that the trace generated by the MAP gives higher (lower) values for $T(q)$ for high (low) values of q than the original trace. This difference appears as a rotation in Figure 18.

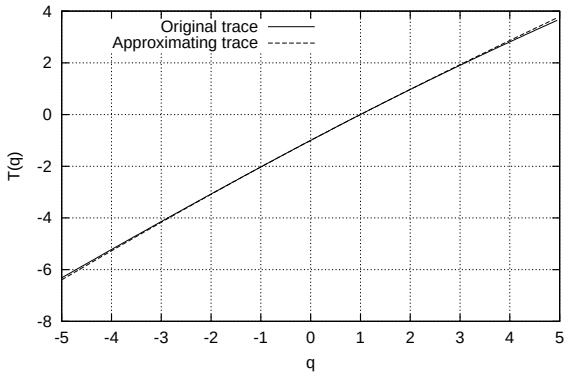

Figure 17. Partition function estimated through the linear fits shown in Figure 16

Actually, Figure 18 shows a rather poor fitting of the multifractal spectrum. Based only on this figure one cannot accept the proposed fitting method. We believe that the differences in the Legendre transforms have to be handled with care. The Legendre transform might overemphasize the

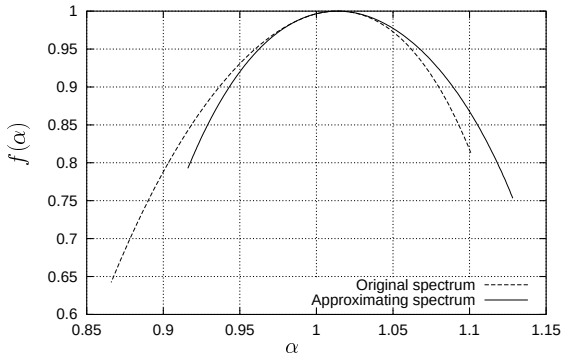

Figure 18. The Legendre transform of the original data set and the one of the approximate MAP

differences of the multifractal spectrum and it is very sensitive to the applied numerical procedure as it was shown in Figure 4. The Legendre spectrum of the approximate multifractal wavelet models proposed in [13] also show significant differences from the one of the original trace (Figure 9 in [13]).

The above tests of the proposed MAP fitting method considered the statistical properties of the original and the approximate processes. From an applications point of view, especially from telecommunication related applications point of view, one of the most important criteria of the goodness of fitting is the queuing behavior resulted by the arrival processes.

We also compared the queuing behavior of the original data set with the one of the approximate MAP assuming deterministic service time and different queue utilization, ρ. The utilization was set by properly choosing the value of the deterministic service time. As it is mentioned above, we apply an artificial time unit such that the overall average arrival rate is 1 (one arrival per time unit). Using this time unit, the service time (< 1) is equal to the utilization. Figures 19 - 22 depict the queue length distribution resulting from the original and the approximate arrival processes. The queue length distribution curves show a quite close fit. The probability of an empty queue, which is not displayed in the figures, is the same for the MAP as for the original trace since the MAP has the same average arrival intensity as the original trace. The fit is better with a higher queue utilization, which might mean that different scaling behaviors play a dominant rule at different utilizations, and the ones that are dominant at high utilization are better approximated by the proposed MAP.

5 Conclusion

The paper presents a MAP structure that is able to exhibit multifractal scaling behavior according to the commonly applied statistical tests. The proposed MAP structure is constructed similarly as the unnormalized Haar wavelet transform of finite sequences.

A heuristic fitting method is also proposed to approximate data sets with multifractal scaling behavior by a MAP with the considered structure. Our numerical experiences show a good fitting according to the majority of the

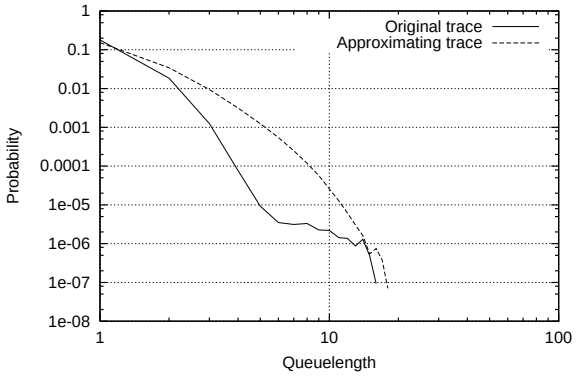

Figure 19. Queue-length distribution at $\rho = 0.2$

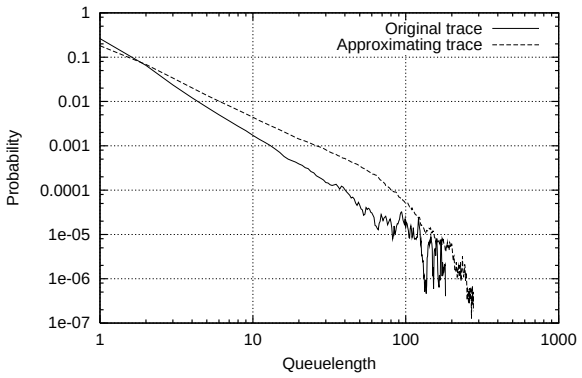

Figure 20. Queue-length distribution at $\rho = 0.4$

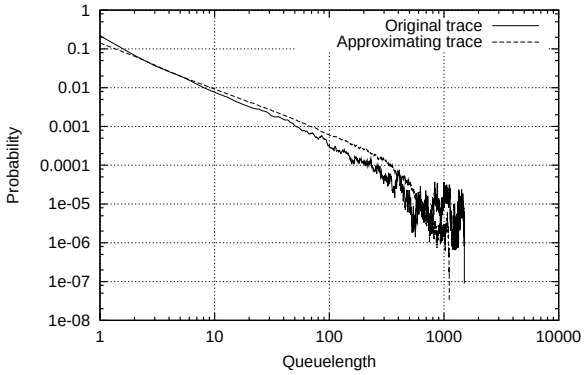

Figure 21. Queue-length distribution at $\rho = 0.6$

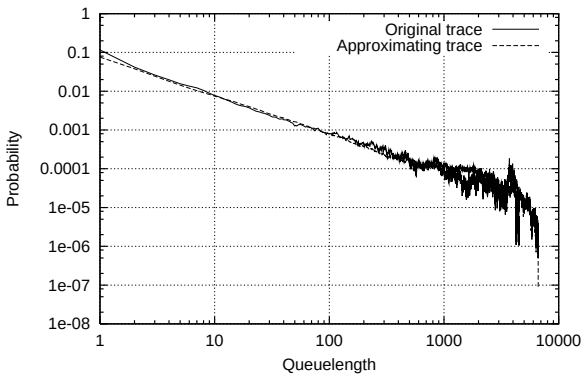

Figure 22. Queue-length distribution at $\rho = 0.8$

performed tests except the comparison of Legendre spectra of the original and the approximate arrival processes.

From telecommunication applications point of view it is promising that the queue length distribution of the original arrival process fits in with the one of the approximate arrival process.

Acknowledgments

This work was partially supported by OTKA grant n. T-30685. András Horváth acknowledges the support of the Italian Ministry for University and Scientific Research, through the Planet-IP project

References

1. W. Willinger, M. S. Taqqu, and A. Erramilli. A bibliographical guide to self-similar traffic and performance modeling for high speed networks. In *Stochastic Networks: Theory and Applications*, pages 339–366. Oxford University Press, 1996.
2. I. Norros. On the use of fractional brownian motion in the theorem of connectionless networks. *IEEE Journal on Selected Areas in Communications*, 13:953–962, 1995.
3. B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1969.
4. G. E. P. Box, G. M Jenkins, and C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliff, N.J., third edition, 1994.
5. C. W. J. Granger and R. Joyeux. An introduction to long-memory time series and fractional differencing. *Journal of Time Series Analysis*, 1:15–30, 1980.
6. B. Ryu and S. B. Lowen. Point process models for self-similar network traffic, with applications. *Stochastic models*, 14, 1998.
7. A. T. Andersen and B. F. Nielsen. A markovian approach for modeling packet traffic with long-range dependence. *IEEE Journal on Selected Areas in Communications*, 16(5):719–732, 1998.
8. S. Robert and J.-Y. Le Boudec. New models for pseudo self-similar traffic. *Performance Evaluation*, 30:1997, 57-68.
9. A. Horváth, G. I. Rózsa, and M. Telek. A map fitting method to approximate real traffic behaviour. In *8th IFIP Workshop on Performance Modelling and Evaluation of ATM & IP Networks*, pages 32/1–12, Ilkley, England, July 2000.
10. M.F. Neuts. *Structured stochastic matrices of M/G/1 type and their applications*. Marcel Dekker, 1989.
11. G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. SIAM, 1999.
12. A. Feldman, A. C. Gilbert, and W. Willinger. Data networks as cascades: Investigating the multifractal nature of internet wan traffic. *Computer*

communication review, 28(4):42–55, 1998.
13. R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk. A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory*, 45:992–1018, April 1999.
14. R. H. Riedi and J. Lévy Véhel. Multifractal properties of TCP traffic: a numerical study. Technical Report 3129, INRIA, February 1997.
15. M. Taqqu, Vadim Teverovsky, and Walter Willinger. Is network traffic self-similar or multifractal? *Fractals*, 5:63–73, 1997.
16. W. E. Leland, M. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions in Networking*, 2:1–15, 1994.
17. I. Norros. A storage model with self-similar imput. *Queueing Systems*, 16:387–396, 1994.
18. G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.
19. R. H. Riedi. An introduction to multifractals. Technical report, Rice University, 1997. Available at `http://www.ece.rice.edu/~riedi`.
20. The internet traffic archive. http://ita.ee.lbl.gov/index.html.
21. J. Lévy Véhel and R. H. Riedi. Fractional brownian motion and data traffic modeling: The other end of the spectrum. In C. Tricot J. Lévy Véhel, E. Lutton, editor, *Fractals in Engineering*, pages 185–202. Springer, 1997.