Simulation Modelling Practice and Theory 26 (2012) 32–48

Contents lists available at SciVerse ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

A multiscaling traffic model for UDP streams

Larissa O. Ostrowsky [a], Nelson L.S. da Fonseca [a,*], Cesar A.V. Melo [b]

[a] State University of Campinas, Institute of Computing, Campinas, Brazil
[b] Federal University of Amazonas, Institute of Computing, Mananus, Brazil

ARTICLE INFO

Article history:
Received 22 May 2011
Received in revised form 3 December 2011
Accepted 4 April 2012
Available online 11 May 2012

Keywords:
Traffic model
Multifractal traffic models
UDP non-responsive traffic flows

ABSTRACT

Although most of the traffic carried over the Internet uses the Transmission Control Protocol (TCP) as the transport layer protocol, it is of paramount importance to develop models for streams that use the User Datagram Protocol (UDP), since these streams are inelastic and, consequently, they can jeopardize the acquisition of bandwidth by TCP streams. This paper introduces a traffic model for UDP streams and its performance is compared to those of other traffic models. The proposed model can be used to generate streams of aggregated UDP sources in simulation experiments.

1. Introduction

The Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP), the two transport protocols most widely used in the Internet, provide, respectively, connection and connectionless services at the transport layer. They differ in various aspects. TCP furnishes reliable delivery services, with a transmission window governed by both flow control and congestion control mechanisms which introduce processing overhead as well as limit the potential bandwidth a connection can use. UDP, on the other hand, neither furnishes reliable data delivery nor has the mentioned overhead and limitation, so it is more appropriate for real-time applications such as voice over IP and video on demand.

Although the deployment of real-time applications on the Internet has increased, roughly 80% of the traffic still uses TCP as the transport protocol which justifies the large number of traffic models for TCP streams that has been proposed [1–3]. The proportion of UDP traffic in the Internet does not diminish the need for accurate models for such type of traffic given that UDP does not reduce its transmission rate under congestion situations and consequently it can jeopardize the bandwidth acquired by TCP streams under congestion. Understanding and being able to reproduce the behavior of UDP streams at the packet level is a key for the assessment of the efficacy of congestion control mechanisms. Nonetheless, not too much attention has been paid to such type of model, with the most popular models focusing on the behavior at the flow level [4,5].

Moreover, the scaling nature of Internet traffic has been subject to debate in recent years. Some authors advocate the use of monoscaling models (monofractal, self-similar) while others prefer the use of multiscaling models (multifractal). The authors of this paper analyzed a large number of IP, TCP and UDP traces to shed light on scaling nature of these types of traffic [6]. Results suggest that TCP streams determine the scaling nature of IP flows as well as that UDP flows are multifractal [6]. Such multifractal nature can be understood by the diversity of packet size transported by the UDP protocol ranging from small packet such as those of VoIP to large ones generated by media streaming applications.

This paper proposes a traffic model for UDP streams, firstly introduced in [7], which reproduces the traffic at the packet level as well as its scaling characteristics. The model is not oriented to the traffic generated by specific applications but rather to the aggregation of the packets generated by applications which use UDP protocols at the transportation layer. This model

* Corresponding author.
E-mail addresses: lari_ost@yahoo.com.br (L.O. Ostrowsky), nfonseca@ic.unicamp.br (N.L.S. da Fonseca), cavmelo@ICOMP.ufam.edu.br (C.A.V. Melo).

1569-190X/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.simpat.2012.04.003

was based on the identification of marginal distributions existing in UDP flows, as well as the identification of those distributions needing to be considered in greater detail. The model is a 4-state one and its accuracy is compared to that of other multiscaling models proposed in the literature. It can be used to generate synthetic data for simulation experiments at low computational cost. Besides providing an extensive evaluation of the proposed model and expanded comparison with related models, this paper differs from [7] by the precise characterization of the multifractal nature of UDP traffic as well as by the multifractal analysis of the traffic generated by the proposed model.

The following section reviews the concept of scaling in network traffic. Section 3 describes various distributions of UDP traffic streams. Section 4 presents the construction of the model. Section 5 overviews two multifractal models commonly used in the literature. Section 6 provides an analysis of its performance. Section 8 concludes the paper.

2. Scaling nature of network traffic

Since the seminal work of Leland et al. [8], several studies have shown that network traffic presents scale invariance, or scaling, which is the absence of any specific time scale for which the "burstiness" of a traffic stream can be characterized. Hence, an accurate description of traffic must account for a variety of time scales. Such traffic presents long range dependency which implies that the auto-correlation of the traffic decays very slowly, or hyperbolically; moreover this auto-correlation is non-summable across all time scales. LRD is characteristic of most network traffic and it impacts critically on the dimensioning and performance of queues since the loss of packets does not decrease substantially with an increasing buffer. Various types of network traffic can be modeled by self-similar or (mono) fractal processes such as local area network traffic and some wide area network traffic. Scaling of such monofractal traffic is measured by a single constant value: the Hurst parameter. Any self-similar (or monofractal) q order process $\overline{X}(t)$ has statistical moments defined by [9]:

$$E|X(t)|^q = E|X(1)|^q \cdot |t|^{qH} \tag{1}$$

where H is the Hurst parameter. This structure restricts the burstiness of a process to a uniform pattern across different time scales.

Multifractal processes have also been used for modeling network traffic. These processes have richer form of scaling behavior which is associated with non-uniform local variability, i.e. these processes have non-linear behavior at different moments. In addition to long range dependencies, multifractal traffic has a high level of activity on small time scales which differs significantly from that of monofractal traffic on the same scales. Actually, the burstiness on small scales diverges from those on larger scales. In these processes, the local regularity of sample paths are described by a collection of scaling exponent, a generalization of the Hurst parameter. Moreover, the incremental process cannot be described by a gaussian distribution as in monofractal traffic.

The statistical moments of a multifractal process are defined by [9]:

$$E|X(t)|^q = E|X(1)|^q \cdot |t|^{\zeta(q)} \tag{2}$$

where $\zeta(q)$ is the scaling function.

Some wide area network traffic can be modeled as multifractal traffic as will be shown in the next section. The use of a monofractal process to model such multiscale traffic leads to misleading results since this can considerably overestimate the bandwidth needed by the traffic [10].

In the wavelet domain, the relationship established in Eq. (2) is defined by:

$$E|d_X(j,k)|^q \approx 2^{j\zeta(q)} \quad \mathrm{j} \to -\infty \tag{3}$$

where $d_X(j, k)$ is the series of increments (details) obtained by the decomposition of the process $X(t)$ using the discrete wavelet transform. The scaling function $\zeta(q)$ is defined by:

$$\zeta(q) = \alpha_q - \frac{q}{2} \tag{4}$$

where α_q is the scaling exponent. This exponent has its value bound by the burstiness of the traffic. For multifractal processes, it varies at different statistical moments (q).

Abry et al. [9] describe a method called Multiscale Diagram (MD) to determine the occurrence of multifractality in a process. This method consists of verifying the behavior of the function $\zeta(q)$ at different moments. A non-linear function characterizes a multifractal process whereas a linear behavior indicates that the process is monofractal.

Estimating the values of the function $\zeta(q)$ requires the determination of the scaling exponent α_q, as defined in Eq. (4). To determine α_q, the Logscale Diagram (LD) method is used. In this method, α_q is defined by the slope of the curve that is close to the curve generated by the relation between $S_q(j)$ and 2^j on a logarithmic scale. $S_q(j)$ is given by:

$$S_q(j) = \frac{1}{n_j}\sum_{k=1}^{n_j}|d_X(j,k)|^q \approx E|d_X(j,k)|^q$$

where n_j is the number of details $d_X(j, .)$, on time scale j, generated by the decomposition of $X(t)$ using the discrete wavelet transform.

Multifractality can also be detected by the use of a Linear Multiscale Diagram (LMD) which plots $h_q = \zeta_q/q$ against q. In such diagram, monofractality is revealed by horizontal alignment whereas a non-horizontal alignment reveals multifractality.

A multifractal process has non-uniform burstiness and its variations are not captured by a single constant value, such as the Hurst parameter. Whenever the process has non-uniform burstiness the relation between the statistical moments and time scales (Δ) on a logarithmic scale (Eq. (2)) is non-linear.

The aforementioned theory has been used to characterize the scaling of traffic streams in the literature. Before using it in this paper, we address the issues raised in [11]. Several papers in the literature overstate the multiscaling nature of traffic resulting in misuse of tools for identifying scaling as well as the misidentification of the proper scale for analysis. To avoid such type of mistake, we followed the procedure described next [11]:

1. Define the right process from the analyzed traffic traces.
2. Tune the confidence interval estimation procedure implemented by the signal analysis tool.
3. Find the meaningful time scales.

Four processes were suggested by the Veitch et al. [11] for the analysis. The arrival time point process, $X(t)$, is the most used in the literature given the adequacy of the tools based on wavelet to analyze this type of process. Alternatively, the byte arrival process, $W(t)$ can be used since it leads to similar results than those when the arrival point process is used.

Veitch et al. [11] stressed the need for fine tuning the parameters of the tool for the derivation of confidence intervals in a non-Gaussian context. They suggested to set the non-Gaussian option in the tool, since traffic on small time scale does not show Gaussian behavior, which will empirically estimate the confidence intervals from the data. Such changes have been implemented by the authors.

Veitch et al. [11] suggested the use of three time scales: the "Inter-Arrival Time" scale, j^{IAT}, identifies the scale of isolation of individual packets; the "breakup" scale, j^{**}, defines the start of scaling region and the third one, the "biscaling knee" scale, j^{*}, defines the possible regimes changes: multifractal on smaller scales and monofractal on larger scales. All the three scales have been considered in the present analysis.

We investigated the existence of multiscaling pattern in all traces used in this work by looking carefully their logscale diagrams. Table 1 gives information on the traces used, including the values of time scales j^{IAT}, j^{**}, and j^{*}. Moreover, results for the trace BWY-1069762448 are presented in Fig. 1.

Multiscaling is detected, in general, by the existence of a collection of scaling exponents $\{\alpha_q\}$ along the slope of straight lines in a q-LD (Logscale Diagram) in the same range of scales. However, multiscaling shows only evidences of multifractality [11]. Fig. 1 shows that in the range $q \in [1, 6]$, twin scales can be observed: at fine scales, $[j^{**}, j^{*}] = [4, 10]$ and at coarse scales, $[j^{**}, j^{max}] = [10, 15]$. Therefore, the trace exhibits multiscaling in each scale range.

Since the biscaling identified in the Logscale diagram analysis is due to the existence of twin scaling regimes, we analyzed the scaling properties in the range defined by the *breakup* and by the *knee* scales, see Table 2. Moreover, we analized the scaling pattern by plotting the Linear Multiscale Diagram (LMD), c.f Fig. 2, of traces in Table 1. On coarse scales, the h_q of functions all traces show a clearly horizontal shape, suggesting scaling is described by a single value, i.e. monofractal models should represent the traffic dynamics over these scales. On the other hand, on fine (small) scales, all the h_q's functions show a non-horizontal shape suggesting multifractality. By doing such analysis, we conclude that the traces used in this paper are indeed multiscaling.

3. Characterization of marginal distributions

The approach adopted for derivation of UDP traffic model presented here involves an initial identification of the distributions which best characterize the relevant marginal distributions of the UDP traffic streams. Some of these distributions are bimodal such as the distribution of packet size. Rather than defining a special distribution to describe the bimodal shape of some of the UDP stream distributions, we modeled the two distinct regions of the UDP stream distribution using two different well known distribution. The marginal distributions of UDP streams worth modeling using two distinct regions were identified as well as the cut off point that separated the two regions. After identifying the well known distribution for each region of the UDP distribution, the model was derived and its effectiveness evaluated. Models with two, four, and eight states were derived, but that with four states furnished the most precise results.

To characterize the distributions of UDP traffic streams, real network traces available publicly at http://www.nlanr.net were employed. Table 1 shows the traces used. UDP packets were extracted from the IP traces to form the UDP stream associated to the trace. The software Coralreef [12] was employed to process the traces. UDP constitute roughly 20% of the total number of packets in the traces.

Table 1
Traces used in the investigation.

Trace	Date	Start time	time (s)	Rate (MBps)	$j^{IAT}(ms)$	j^{**}	j^{*}
BWY-1069762448	11/25/03	04:49	90	0.082994	4 (0.441773)	4	10
BWY-1069224416	11/18/03	23:43	90	0.224573	4 (0.260058)	6	10
UFL-1095079334	09/13/04	06:11	90	0.181745	4 (0.280426)	7	11

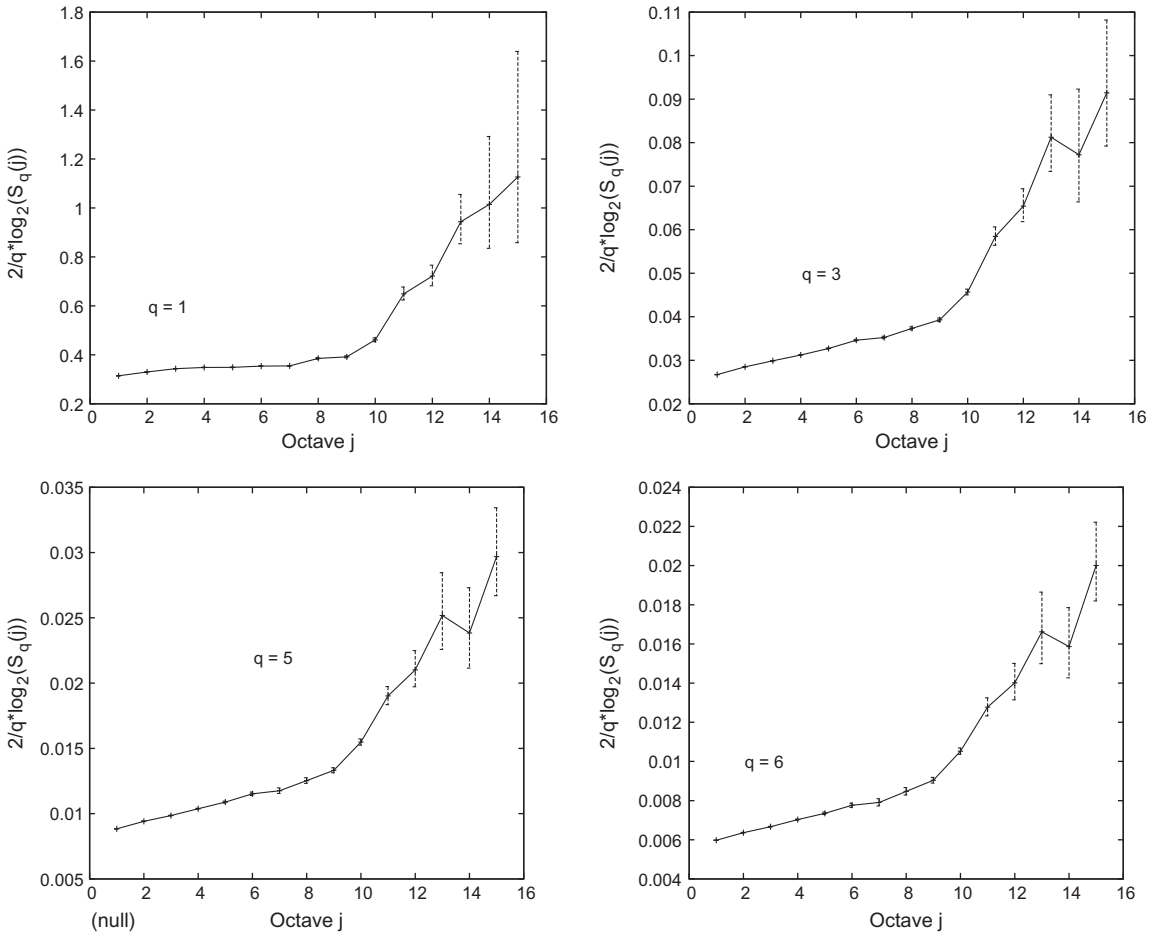


Fig. 1. qth order logscale diagrams for UDP flow of trace BWY-1069762448.

Table 2
Scales defining biscaling regimes as measured from the traces. j^{max} is the coarser scale applied in the analysis.

Trace	Biscaling regimes	
	Fine scale	Coarse scale
BWY-1069762448	$[j^{**}, j^{*}] = [4, 10]$	$[j^{**}, j^{max}] = [10, 15]$
BWY-1069224416	$[j^{**}, j^{*}] = [6, 10]$	$[j^{**}, j^{max}] = [10, 16]$
UFL-1095079334	$[j^{**}, j^{*}] = [7, 11]$	$[j^{**}, j^{max}] = [11, 14]$

The distribution that most influenced in the derivation of the model was that of packet size and that of the duration of the bursts of packets. Packet interarrival time was also investigated, but produced no significant impact on the modeling of UDP streams. The model for a UDP stream distribution was developed employing the SAS and MATLAB softwares. Hypothesis were tested by applying the Kolmogorov–Smirnov test with an alpha level of 0.01. Various statistical distributions used in the tests: Exponential, Beta, Gamma, Normal, Lognormal, Pareto and Weibull. Next, the characterization of UDP stream distributions is described.

The next sub-sections analyzes the relevant distribution for the characterization of UDP streams. The trace BWY-1069224416 will be used to illustrate the specific distributions adopted in the model.

3.1. Packet size distribution

Fig. 3 plots the distribution of packet size, which is clearly bimodal. Such a shape is due to the fact that different application level protocols produce data protocol units of different sizes. Such a bimodal distribution cannot be described by any of the distributions considered. Although a single bimodal distribution could have been derived for packet size, the decision

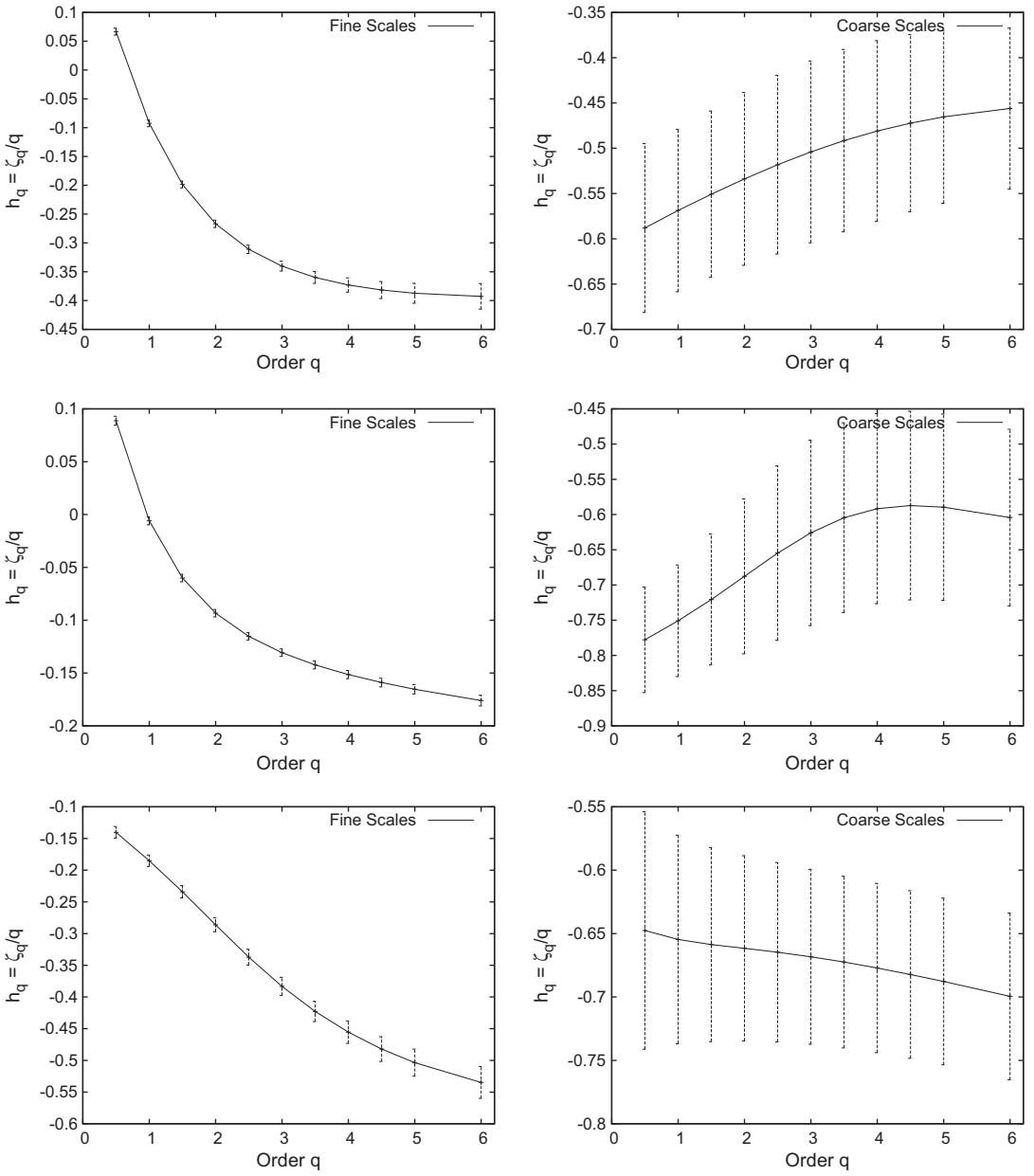


Fig. 2. LMDs of UDP flow of traces BWY-1069762448(top), BWY-1069224416(middle), UFL-1095079334(bottom) over fine scales and coarse scales.

was made to use different distributions for different ranges of packet size. The incorporation of two distinct distributions rather than a single bimodal one involves defining different states in the traffic model. The packet size considered to divide the two regions is 750 bytes, with modes of 120 and 1350 bytes.

Hypothesis tests indicate that the distribution which best characterizes the two regions is the Beta distribution (Fig. 4). In the first region an inverted-J Beta distribution, i.e. $a < 1$ and $b \geqslant 1$, was used to characterize the packet size distribution. In the second region, the packet distribution is characterized by a Unimodal Beta distribution, i.e. $a, b > 1$, with its peak at 1396.9.

3.2. Interarrival time distribution

The choice of using two unimodal distributions to characterize the bimodal distribution of the packet size led to the definition of a model which involves states associated with the two regions of the distribution. The distributions of the packet interarrival time are the ones of packet arrivals belonging to the burst in the same state. For these interarrival time

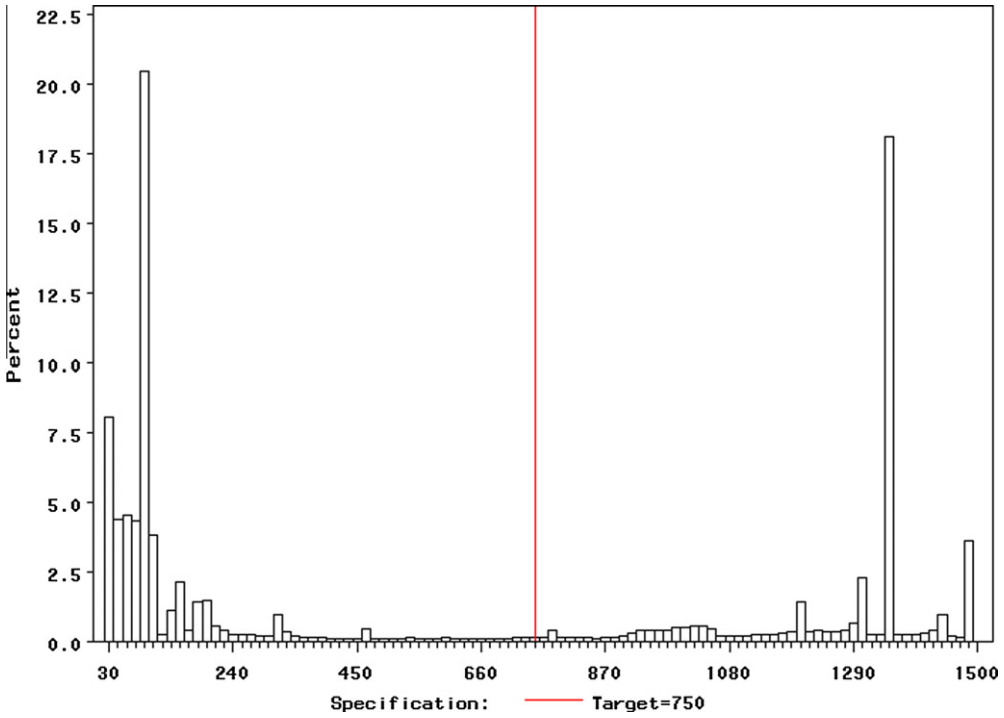


Fig. 3. The whole distribution for trace BWY-1069224416.

distributions, burst of packets with size in the same region of the two packet size distribution were identified in the trace. Arrivals of single packets were also accounted as bursts of size one.

The exponential and the Weibull distributions were the most significant ones for modeling the interarrival time of packets. Moreover, the Weibull distribution fits more precisely the sample distribution (Fig. 5). The Weibull distribution reproduces the behavior of the interarrival time of packets, including the peak values in both regions, the Exponential distribution gives poor results, specially when modeling the interarrival time of packets in the second region (Fig. 5b). In the first region, the Exponential distribution shows a better fitting. Actually, this distribution is very similar to that when the Weibull distribution is used. Such similarity is due to the fact that the Exponential distribution is a special case of the Weibull distribution with the value of its *shape* parameter set to 1(one) [13]. In this example, the value of the shape parameter is equal to 0.88, very close to one, which explains that close fitting (Fig. 5a).

3.3. Duration of bursts in packet distribution

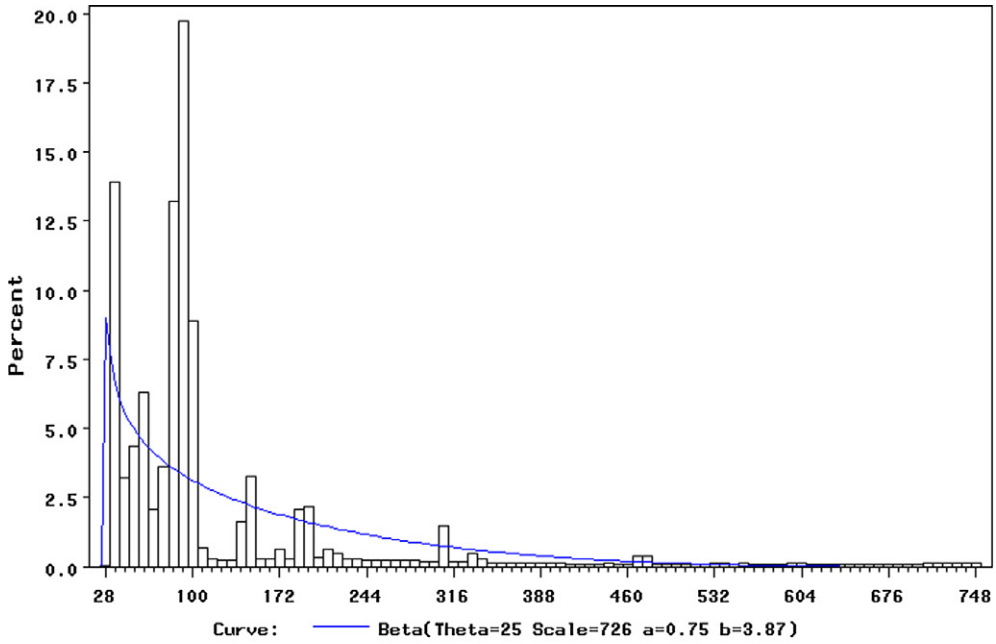
The distributions of the duration of bursts of packets were broken down into two regions since this leads to more accurate traffic models than would a single region used. This can be explained by the fact that the different application layer protocols that generate different packet sizes tend to generate bursts of packets of distinct durations. The cut-off value for burst duration for all traces studied can be seen in Table 3.

Hypothesis tests have indicated that the Uniform distribution is the one which best characterizes the duration of bursts shorter than those values presented in Table 3 for both regions, although the Weibull distribution is the most accurate distribution for bursts longer than the cut-off durations (Table 3).

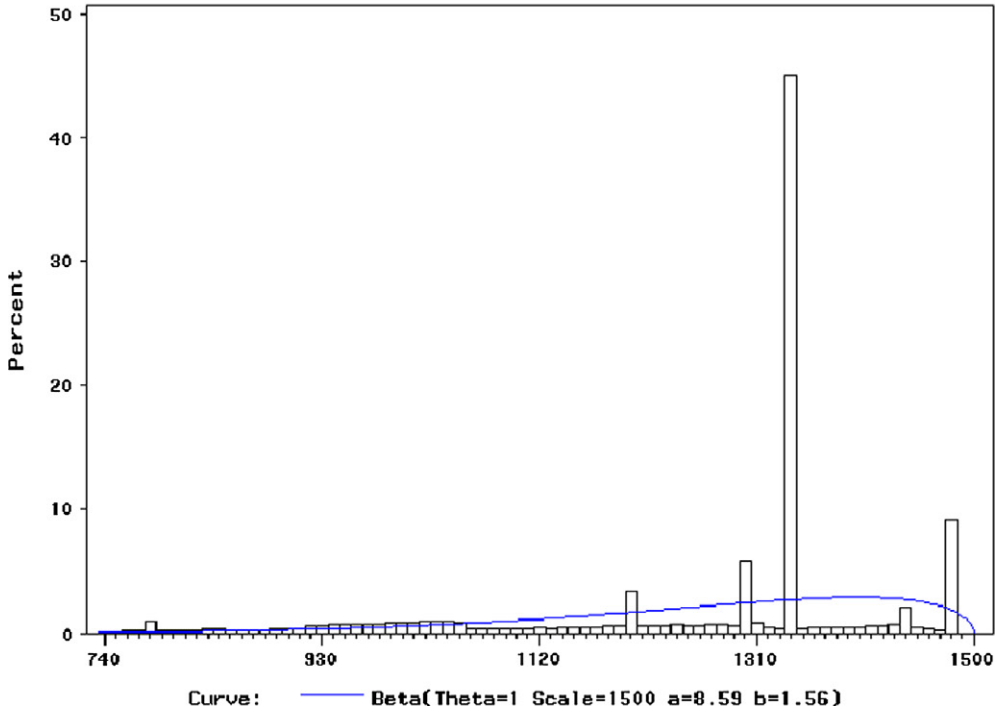
Fig. 6 shows the fitting of the distributions when the burst duration is greater than the cut-off values. It can be seen that the Weibull distribution models more precisely the sample distribution of UDP burst size than do the Exponential distribution for both regions. In the first region, the Exponential distribution and the Weibull distribution show a close fitting but again setting the value of the shape parameter of the Weibull distribution is equal to 1.

4. Proposed traffic model

It was clear from the data obtained that both modes of packet size distribution need to be considered in the modeling process, with the precise modeling of packet size constituted the initial step in the derivation of the model. States were defined for the two regions of the packet size distribution (packets smaller and greater than 750 bytes). To refine the model, various combinations of potential distributions for packet interarrival time and burst duration were then investigated, and the resulting models tested. Both, single distributions and distributions by region were considered.



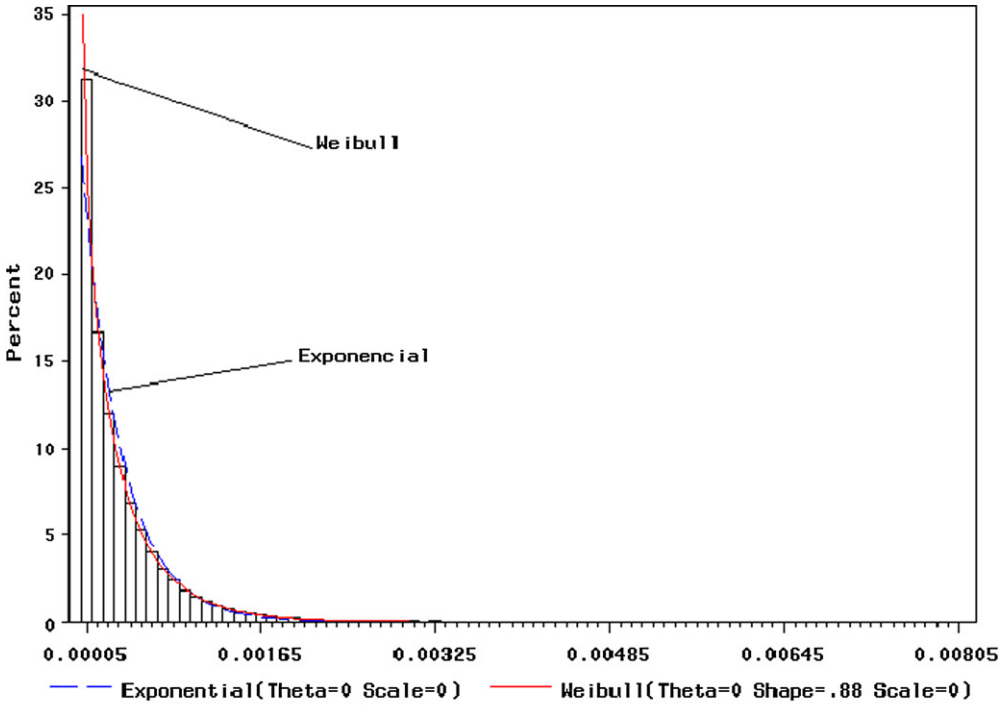
(a) Fitting for Packet Size in the First Region: Beta(theta = 25, scale = 726, $a = 0.75$, $b = 3.85$)



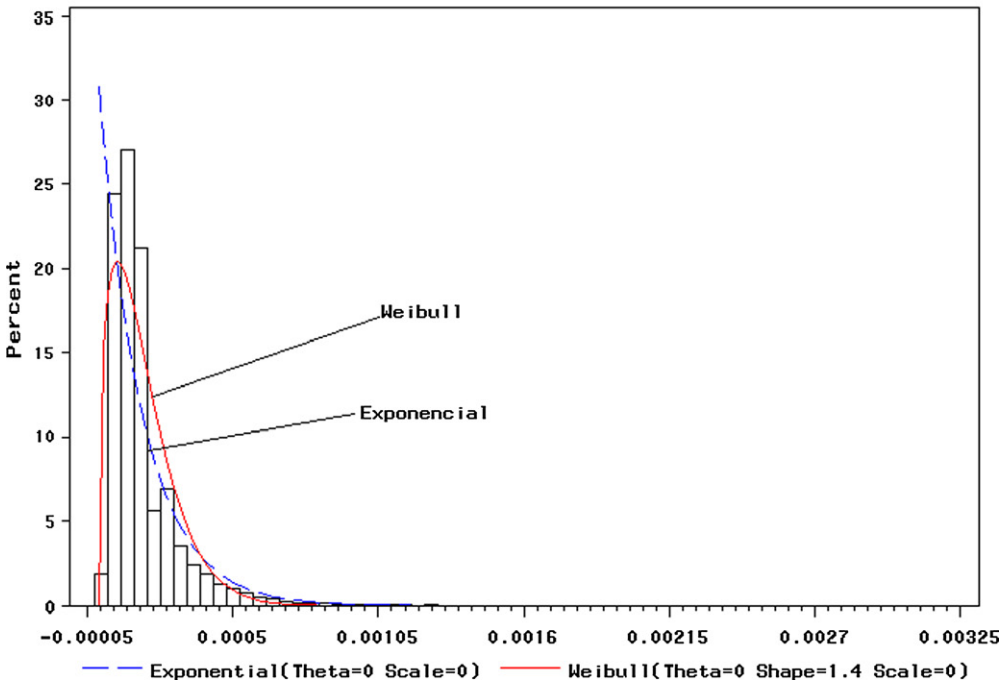
(b) Fitting for Packet Size in the Second Region: Beta(theta = 1, scale = 1500, $a = 8.59$, $b = 1.56$)

Fig. 4. Packet size distribution for trace BWY-1069224416.

Since the size of UDP datagram depends on the application protocol using UDP, for instance datagram carrying DNS messages are small while those carrying RTP messages are typically large. Thus, diverse packet size are expected. The relevance of the packet size to the derivation of the model was somehow predictable. Moreover, the duration of packet burst is tightly



(a) Fitting for Interarrival Time in the First Region: (dashed line) Exponential(theta = 0, scale = 0); (continuous line) Weibull(theta = 0, shape = 0.88, scale = 0)



(b) Fitting for Interarrival Time in the Second Region:(dashed line) Exponential (theta = 0, scale = 0);(continuous line) Weibull(theta = 0, shape = 1.4, scale = 0)

Fig. 5. Interarrival time distribution for trace BWY-1069224416.

Table 3
Cut-off values for burst durations of the traces studied.

Traces	Cut-off values (in s)	
	First region	Second region
BWY-1069762448	0.00128	0.00075
BWY-1069224416	0.00050	0.00018
UFL-1095079334	0.00075	0.00030

coupled with the scaling properties of the traffic stream. It was also expected that the duration of packet bursts would be quite relevant to the model.

The most precise model was one with four states involving two states for packet size mode and two associated to the packet burst duration. The model is presented in Fig. 7; the definition of the states are:

- State 1 – burst of packets smaller than 750 bytes, with duration shorter than x seconds.
- State 2 – burst of packets smaller than 750 bytes, with duration longer than x seconds.
- State 3 – burst of packets larger than 750 bytes, with duration shorter than x seconds.
- State 4 – burst of packets larger than 750 bytes, with duration longer than x seconds.

For transitions (1;3), (3;1), (1;4), (4;1), (2;3), (3;2), (2;4) and (4;2), the transition rate between states is computed as:

$$\lambda_{i,\,j} = \frac{P_{i,\,j}}{\text{residence time in state } i} \tag{5}$$

where

$$P_{i,\,1} = \frac{\text{number of bursts with duration } < x \text{ seconds of packets with less than 750 bytes}}{\text{number of bursts of packets with less than 750 bytes}}$$

$$P_{i,\,2} = \frac{\text{number of bursts with duration } > x \text{ seconds of packets with less than 750 bytes}}{\text{number of burst of packets with less than 750 bytes}}$$

$$P_{i,\,3} = \frac{\text{number of bursts with duration } < y \text{ seconds of packets with more than 750}}{\text{number of bursts of packets with more than 750 bytes}}$$

$$P_{i,\,4} = \frac{\text{number of bursts with duration } > y \text{ seconds of packets with more than 750}}{\text{number of bursts of packets with more than 750 bytes}}$$

For other transitions, i.e. (1;2), (2;1), (3;4) and (4;3), the transition rate between states is equal to zero.

In this model, bursts containing packets smaller than 750 bytes alternate with bursts containing packets larger than 750 bytes.

Table 4 shows the value of the parameters for the beta distribution used to model the packet size of the original traces used, and Table 5 shows the parameters of the distributions that capture the burst size durations. For the uniform distribution parameters a and b indicate the range of the distribution whereas for the Weibull distribution a provides the scale while b gives the shape of the distribution.

The two models were derived and their performance compared to that of the proposed 4-state model, one with two state and the other with eight states.

The states of the 2-state model correspond to the duration of a single burst; one state corresponds to the duration of those bursts composed of packets smaller than 750 bytes, while the other is associated with bursts composed of packets larger than this value. The difference between the 2-state model and 4-state one proposed here is that there is no specific division of burst durations. In other words, the states of the 2-state model are derived by aggregating states of the 4-state model on the basis of packet size.

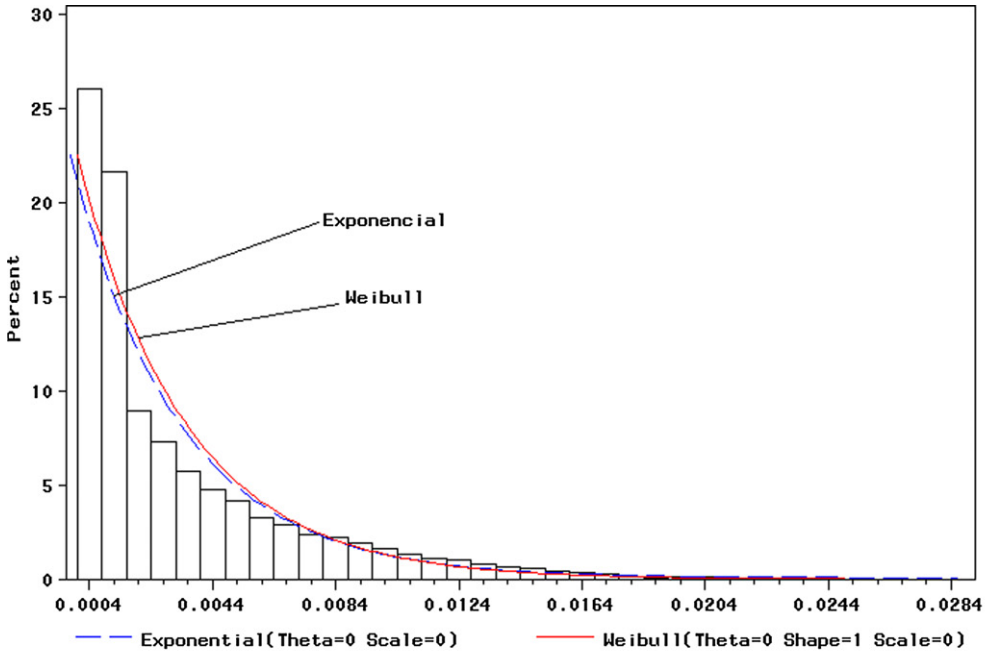
For the derivation of the 8-state model, the packet interarrival time are classified in relation to a given threshold (Table 6 shows the threshold value for traces used). Each state of the 4-state model is thus divided into two according to packet interarrival times, one longer than the threshold value and the other shorter.

5. Multifractal traffic models

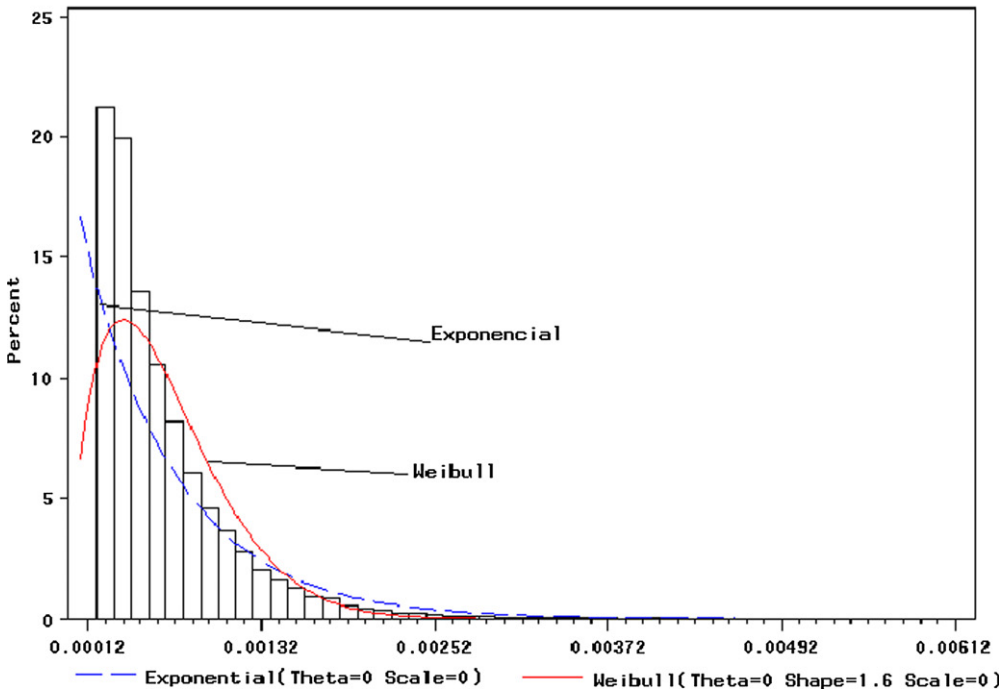
This section describes two multiractal traffic models: the Markovian arrival process and the Multifractal wavelet model which are used in this paper for comparison with the proposed model.

5.1. The Markovian arrival process model

Horvth and Telek [14] introduced a Markovian Arrival Process (MAP) model which is capable of generating multifractal traffic based on multifractal formalism. This formalism makes it possible to derive a time series with target variance decay



(a) Fitting for Burst Size in the First Region: (dashed line) Exponential(theta = 0, scale = 0); (continuous line) Weibull(theta = 0, Shape = 1, scale = 0)



(b) Fitting for Burst Size in the Second Region: (dashed line) Exponential(theta = 0, scale = 0); (continuous line) Weibull(theta = 0, Shape = 1.6, scale = 0)

Fig. 6. Burst size distribution for trace BWY-1069224416.

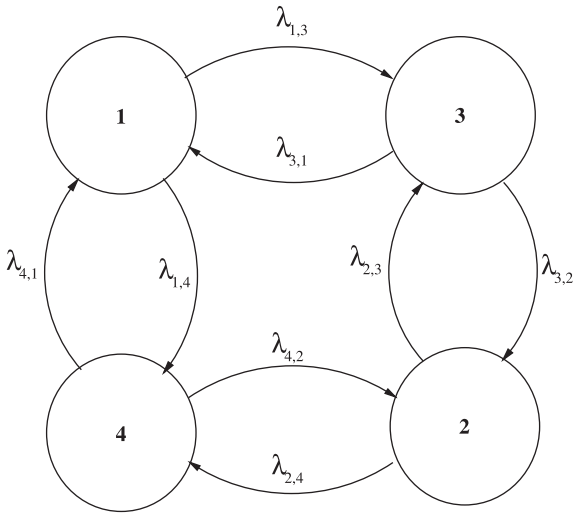


Fig. 7. 4-State model.

Table 4
Beta distribution parameters used to characterize packet size.

Trace	Pkt size < 750 bytes		Pkt size > 750 bytes	
BWY-1069762448	$\alpha = 1.97$	$\beta = 172.94$	$\alpha = 63.39$	$\beta = 447.58$
BWY-1069224416	$\alpha = 1.81$	$\beta = 143.49$	$\alpha = 43.93$	$\beta = 301.72$
UFL-1095079334	$\alpha = 1.44$	$\beta = 90.98$	$\alpha = 25.96$	$\beta = 168.96$

Table 5
Distribution parameters used to characterize burst duration.

Trace	Distribution	Pkt size < 750 bytes		Pkt size > 750 bytes	
BWY-1069762448	Uniform	$a = 0$	$b = 0.00128$	$a = 0$	$b = 0.00075$
	Exponential	$\mu = 0.016$		$\mu = 0.0010$	
	Weibull	$a = 0.0117$	$b = 1.6389$	$a = 0.0016$	$b = 2.0478$
BWY-1069224416	Uniform	$a = 0$	$b = 0.00050$	$a = 0$	$b = 0.00018$
	Exponential	$\mu = 0.0050$		$\mu = 0.0020$	
	Weibull	$a = 0.0035$	$b = 0.9875$	$a = 0.0007$	$b = 1.4995$
UFL-1095079334	Uniform	$a = 0$	$b = 0.00075$	$a = 0$	$b = 0.00030$
	Exponential	$\mu = 0.0048$		$\mu = 0.0014$	
	Weibull	$a = 0.0022$	$b = 1.1830$	$a = 0.0012$	$b = 1.6009$

Table 6
Threshold values for packet interarrival time of the traces studied.

Traces	Threshold values (in s)	
	First region	Second region
BWY-1069762448	0.00027	0.00015
BWY-1069224416	0.00010	0.00022
UFL-1095079334	0.00018	0.00018

rates by controlling the second moment of the scaling coefficients. In other words, by keeping the traffic increments constant across different time-scales, it is possible to generate time series with a multifractal scaling. A special MAP structure is defined on the basis of a Markov Modulate Poisson Process (MMPP) with an associated Continuous Time Markov Chain (CTMC) and a symmetric n-dimensional cube structure. The procedure of synthesis of this MAP model involves three steps:

1 Choose a time-unit such that the long term arrival intensity is one, and one-state MAP with an arrival rate equal to one to represent the arrival rate on the largest time scale;
2 Increase by one the dimensions of the MMPP cube structure (generator), i.e. double the number of states at level n;

3 For each level $n > 0$:
   3.1 Re-define the time-scale as a multiple of $\gamma^n\lambda$, $\gamma > 0$, so that the time-scale set $(1,1/\lambda,\ldots,1/\gamma^{n-1}\lambda)$ remains unchanged, and;
   3.2 Define the arrival rates that capture these changes in the arrival process. At time-scale $1/\gamma^n\lambda$, a new a_n variance parameter should be linked to the arrival rates.

The MAP structure is defined in such a way that the arrival rates represent the variation in the arrival process on different time scales. At level n, the MAP structure defined has $n+2$ parameters and 2^n states; n out of $n+2$ parameters are related to the variance of the arrival process (a_n' parameters), and the other two are the time scale related parameters γ and λ.

A procedure used to derive the MAP parameters from data measurements is described which will allow the derivation of two sequences (S and C), one for each of the traffic traces. These sequences record the interarrival time of packets, and the number of packets that arrive during a time slots of 1s, respectively. Based on sequences S and C, the model is constructed as follows:

1. Calculate the mean of S, i.e. $E[iat]$.
2. Set T_0 equal to the time instant t where the autocorrelation function of C is equal to zero.
3. Calculate the coarsest time scale T_M based on $E[iat]$ and;T_0, i.e. $T_M = E[iat]*T_0$.
4. Calculate the finest time scale T_m based on T_M and γ, i.e. γ equal to eight, n equal to four or five as the authors suggested.
5. Use the length of S to calculate the number of Haar wavelet coefficients that will be derived to T_M. For $T_M = 2^{15}$ the Haar wavelet coefficients for 15 time scales are computed.
6. Use the downhill simplex method [15] calculate the variance parameters $a_1, a_2, \ldots, a_n$. The variance parameters are those obtained when the relative error of the second moment of the Haar wavelet coefficients is minimal.

5.2. The multifractal wavelet model

Riedi et al. [1] presented the Multifractal Wavelet Model (MWM), a model defined on the wavelet domain which aims at reproducing the non-Gaussian behavior (the positiveness and the spikiness) and to capturing the Long Range Dependence (LRD) and the correlation of events logged on data measurements. The first of the aims is achieved by defining the signal content around the time instant $2^{-j}k$ in a multiplicative structure, and decomposing and scaling the initial energy (signal). More specifically, the wavelet coefficients $W_{j,k}$ are defined using the multipliers $A_{j,k}$ and Wavelet scaling coefficients $U_{j,k}$, i.e. $W_{j,k} = A_{j,k} * U_{j,k}$. The multipliers $A_{j,k}$ are symmetric random variables with values chosen in the interval $[-1, 1]$; the wavelet scaling coefficient is a local mean energy value derived from the signal transformation using an Haar wavelet transform algorithm. This transform algorithm computes the wavelet scaling coefficients as following:

$$U_{j+1,2k} = 2^{-1/2}(U_{j,k} + W_{j,k}) \text{ and}$$
$$U_{j+1,2k+1} = 2^{-1/2}(U_{j,k} - W_{j,k})$$

from which we have the necessary and sufficient condition to the positiveness property, i.e. $|W_{j,k}| \leqslant U_{j,k}$. The model spikiness is a result of the random multiplicative structure, since random products can be occasionally extremely large [1].

The LRD and the correlation pattern are captured by controlling the wavelet energy decay rate over the scales:

$$n_j = \frac{var(W_{j-1,k})}{var(W_{j,k})} \text{ for } 1 \leqslant j < n$$

where n is the higher resolution in the analysis.

Actually, n_j can then be expressed as a function of the multipliers $A_{(j)}$, i.e.

$$n_j = \frac{E\left[A^2_{(j-1)}\right]}{E\left[A^2_{(j)}\right]\left(1 + E\left[A^2_{(j-1)}\right]\right)} \tag{6}$$

where $E(A^2_{(0)}) = \frac{E(W^2_{0,0})}{E(U^2_{0,0})}$. Therefore, it is possible to control the LRD and the correlation among events by defining proper pdfs for the multipliers among the scales. A synthesis procedure to generate an MWM is involves three steps:

(1) Compute the coarsest scaling coefficient at time scale j = 0, i.e. $U_{0,0}$;
(2) At scale j and for $k = 0, \ldots, 2^j - 1$, calculate the children scaling coefficients

$$U_{j+1,2k} = \left(\frac{1 + A_{j,k}}{\sqrt{2}}\right)U_{j,k}$$
$$U_{j+1,2k+1} = \left(\frac{1 - A_{j,k}}{\sqrt{2}}\right)U_{j,k}$$

(3) Repeat step 2 until the finest scale $j = n$ is reached.

In Step 1, the coarsest scaling coefficient is computed under the Gaussianity assumption and a small value for the variance, thus reducing the chances of producing non-positive values.

A Haar wavelet transform algorithm with n levels can be used to derive the wavelet and scaling coefficients from N data samples to fit the MWM to data measurements. $\lceil N2^{-n} \rceil$ coarsest-scale wavelet scaling coefficients are sufficient to estimate the moments at scale zero ($j = 0$). These moments are used to calculate the level of energy in each scale j which is the variance of the signal at scale j (see Eq. (6)).

Two marginal distributions, the symmetric Beta and point mass distributions can be applied to derive the multipliers $A_{j,k}$ since both distributions can be easily shaped by adequately fitting their parameters. The procedure supplied in [1] was used to estimate the Beta distribution parameters and to generate the multipliers $A_{j,k}$ in this paper.

6. Numerical results

To evaluate the effectiveness of the proposed model, simulation experiments were conducted using both synthetic traces generated by the analytical models and real network traces (Table 1). The aim was to compare results produced using a specific model with those resulting by the use of real network trace.

For the models proposed in this paper, packets are generated according to the interarrival time and packet size distribution associated with each state, with the Weibull distribution used for modeling interarrival time. The traffic generator stays in each state of the model (residence time) a random duration given by the distribution of the duration of the burst. The parameters of the distribution were the ones obtained by the characterization shown in Tables 4 and 5.

Synthetic traces for the Multifractal Wavelet Model (MWM) were generated by using the tool available at [16] which produces a positive multifractal process with variance on different time scales that match the variance of the real network trace (see Section 5 for details). The parameters of the MWM were computed using the functions train-beta-mwm [16] and the values extracted from the real network trace. The tool was set to generate bytes in intervals of one millisecond to compose the traces used.

To produce results for the MAP model with 16 and 32 states, the TANGRAM II software [17] was employed. Exact solutions were obtained using the method block GTH. Parameters in this model were defined using the values obtained from the real network traces, according to the procedure described in Section 5. The mean queue length obtained analytically was multiplied by the mean packet size revealed in the traces. The utilization level varied with the experiment as well as service rates.

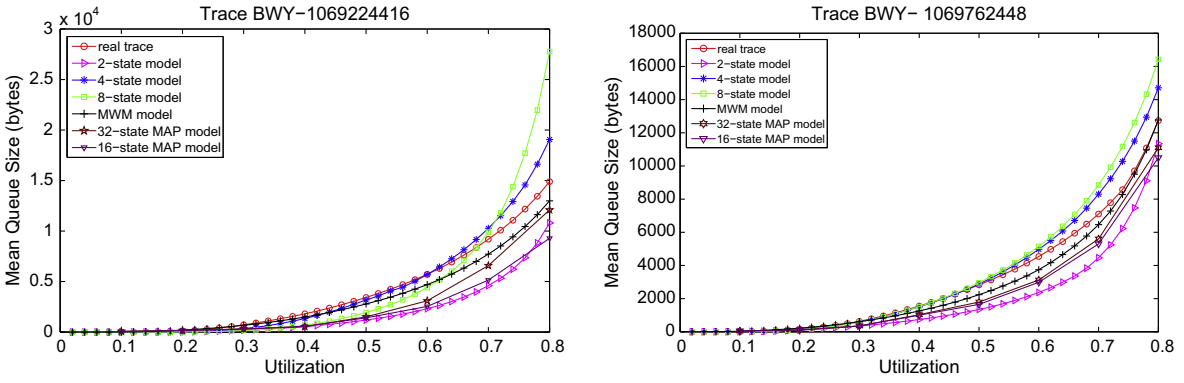
Two set of experiments were conducted. For the first, a queue with an infinite buffer was simulated, and the mean queue length collected as a function of the level of utilization. For the second group of experiments, finite queues were simulated for different levels of utilization and the loss rate values were collected.

Fig. 8 shows the queue length as a function of the level of utilization obtained by varying the service rate. This shows that both the MWM and MAP models underestimate the queue length, which can lead to misleading results when these models are employed for dimensioning purposes. The 2-state model also underestimates the queue size, however the other models, the 4-state and the 8-state models, overestimate the queue length. This overestimation is due to the fact that these two models generate bursts at higher frequency than that observed in real network traces. Moreover, these two models reproduce the high order moments existing in real trace better than do the other models, leading to queue lengths closer to reality. The queue length values produced by the 4-state model are closer than those furnished by the 8-state model, especially for high levels of utilization.

One potential reason for the good performance of the model proposed in this paper is the characterization of the packet size distribution as a bimodal distribution and the construction of the model based on this pattern. MWM and MAP models, on the other hand, use unimodal distributions. Even if a U-shaped Beta distribution were used for the MWM modeling, it would be difficult to fit such model to the traces since a U-shaped Beta distribution shows symmetry around its mean value, a pattern not identified in the traces.

Figs. 9–11 plot the loss rate as a function of the buffer size for utilization levels of 0.7 and 0.9. The buffer size was varied from 10^{+2} to 10^{+5} bytes. For small buffer sizes, all the models considered are equally good since higher order moments are not considered. As the buffer size increases, however, the models perform quite different. The MWM and 2-state models underestimate the loss rate as a consequence of queue length underestimation, and perform poorly for large buffer sizes due to the lack of ability to capture the high order correlations existing in real traffic streams. Conversely, the models with four and with eight states overestimate the loss rate which provides a more conservative estimation. The overestimation of bandwidth needed to provide loss rate guarantees with these two models would not be very high since the overestimation of the loss rate is not particularly high. Moreover, loss rate results from the 4-state model are much closer to those of the real network trace than are those of the 8-state model. No significant difference was observed in the precision of these two models as a function of network utilization. However, the capacity for estimating the loss rate of the MWM model and that of the two state one decrease for lower loads.

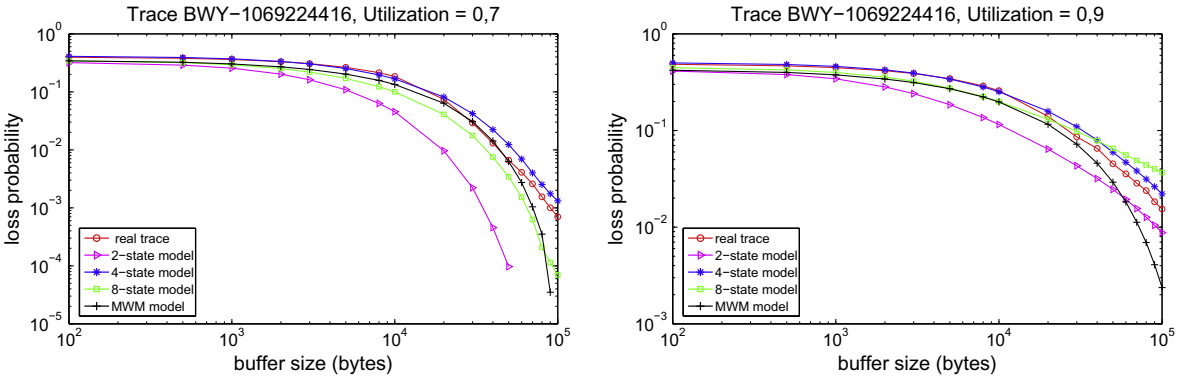
To evaluate the use of the efficiency of the four state model, the execution time to generate traces of different sizes was measured and the time compared to that required by the MWM. Experiments were conducted in a computer with an 1.84 GHz AMD Athlon XP 2500 processor with and 1 Gb RAM memory. The execution time of the models with two and eight states are not shown since the model with four states is the one of interest given its accuracy. Table 7 shows the execution



(a) Trace BWY-1069224416 (b) Trace BWY-1069762448

(c) Trace UFL-1095079334

Fig. 8. Queue length as a function of the utilization level.



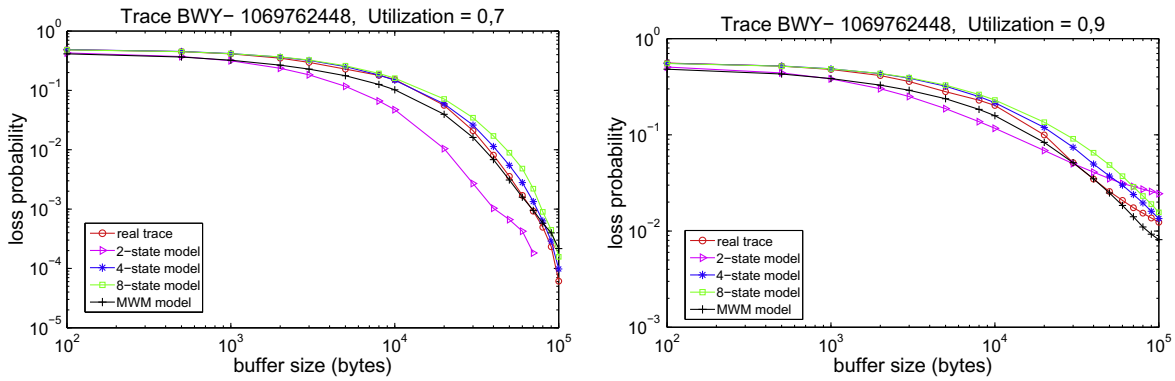
(a) Utilization level equal to 0.7 (b) Utilization level equal to 0.9

Fig. 9. Loss probability evaluation for trace BWY-1069224416.

time for the two models. The time taken by the generator of the 4-state model is typically one order of magnitude larger than that required by the MWM model. Nonetheless, the time required by the 4-state model is still quite reasonable for simulation studies, and its accuracy clearly compensates for the computational demands.

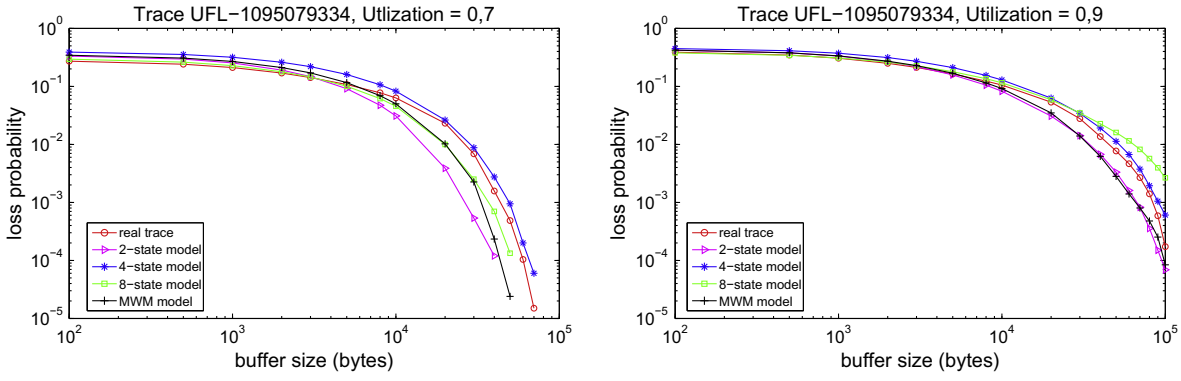
Another aspect which impacts on the usability of a model is the number of parameters involved. Table 8 shows the number of states and parameters required for each of the models studied. Note that although the MAP model requires few parameters, this may be the potential reason for its lack of accuracy, especially in the reproduction of the correlations existing in real traffic streams.

An evaluation of the capability of the proposed model to mimic the scaling existing in real traffic traces was conducted. The scaling pattern found in synthetic traces, generated by the 4-state model setting up according to the parameters



(a) Utilization level equal to 0.7 (b) Utilization level equal to 0.9

Fig. 10. Loss probability evaluation for trace BWY-1069762448.



(a) Utilization level equal to 0.7 (b) Utilization level equal to 0.9

Fig. 11. Loss probability evaluation for trace UFL-1095079334.

Table 7
Execution time as a function of the trace size.

Number of packets	Execution time 4-state model	Execution time MWM model
10^4	2.5730 s	0.2510 s
10^5	9.9060 s	0.8320 s
10^6	95.0150 s	3.8440 s
10^7	891.0620 s	45.6560 s

Table 8
Usability of the investigated models.

Model	Number of states	Number of parameters
4-state	4	12
MAP 16	16	6
MAP 32	32	7
MWM	–	$2 + log_2 N$

collected from real traces, was studied using the Linear Multiscale Diagram (LMD). Fig. 12 provides LMD diagrams for the synthetic counterparts of the traces BWY-1069224416, BWY-1069762448 and UFL-1095079334, evaluated from their break-up time-scales up to their biscaling knee time-scales, i.e. the multifractal cut-off time-scale, see Table 1 for more details.

The main conclusion is that the scaling function h_q of the synthetic counterparts, generated by the 4-state model, of all studied traces show a non-horizontal behavior as would be for LMD diagrams of any multifractal measure. The tied

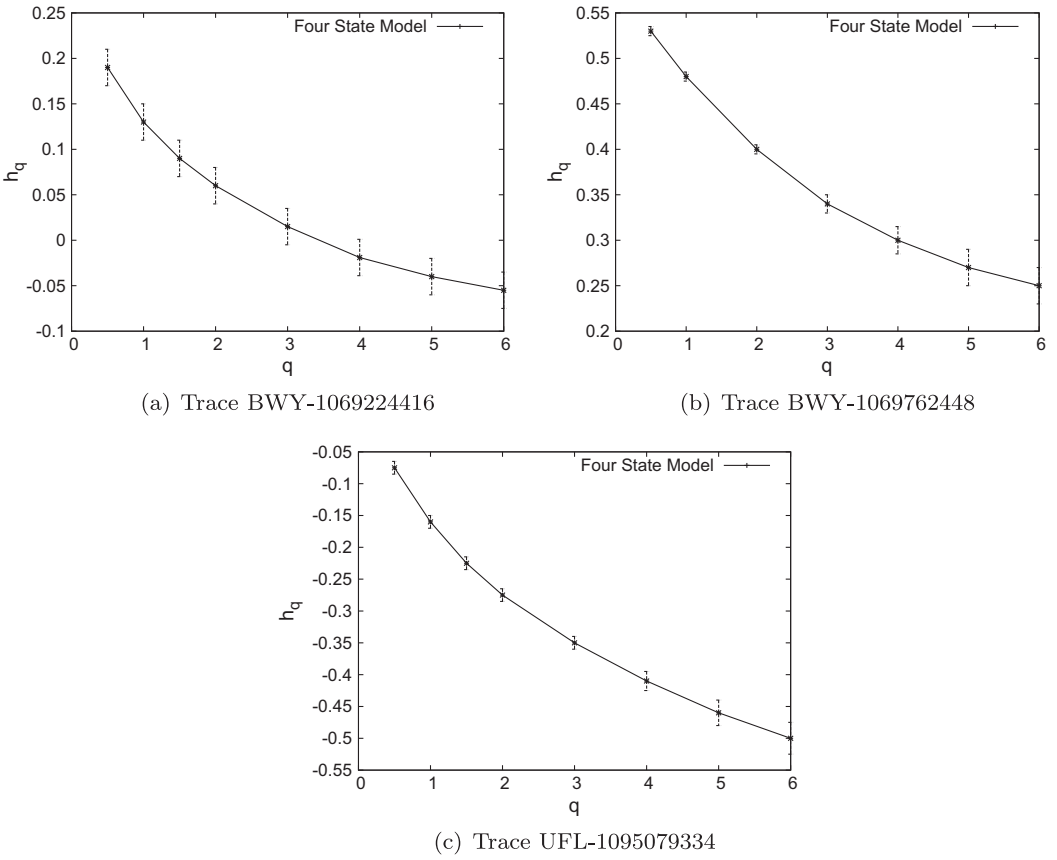


Fig. 12. Multifractal analysis of synthetic traces generated by the 4-state model.

confidence intervals reinforce that and, hence, consider other possible pattern, i.e. a horizontal behavior and monofractal measure, could emerge from the synthetic traces would be a mistaken.

7. Related work

There has been great interest in developing simple models for aggregated UDP traffic that can be used for traffic generation in simulations. Some of these works are briefly surveyed in this section.

In [18], a Hidden Markov model for Internet traffic sources at packet level was introduced. Experimental results show that the model is able to estimate statistical parameters and produce synthetic traces. By exploiting temporal dependencies, the model is able to perform short-term prediction. In [19], a Hidden Markov model was proposed for joint modeling of losses and delays. The authors also analyzed the significance of hidden states automatically found by training algorithm to congestion levels of the network.

In [20], a methodology to identify common traffic profiles from largely unstructured data files was introduced. The methodology combines data mining and information theory to automatically find useful information. Results indicate the good effectiveness of the methodology and its usefulness to network operators and security analysts.

In [21], it was shown that TCP self-clocking coupled with queuing in the bottleneck of the connection's forward path can create ON/OFF inter-arrival structures, and thus, strong correlations and burstiness. It is concluded that to reduce the sub-RTT burstiness of Internet traffic pacing at the sources should be adopted.

In [22], two set of models were proposed. The first models the client behavior for TCP or UDP port and the second models the aggregated traffic from TCP or UDP clients. The dataset used to create the models includes over 36 million TCP flows and 93 million UDP flows collected over a 1 year period. Empirical distributions were used to model five aspects. Data was collected the open-source software NETI@home. Simulation results demonstrate the effectiveness of the proposed models.

Envelope processes [10,23] have been proposed to model aggregated traffic, however, they do not reproduce the traffic behavior on a packet level time-scale.

The present paper differ from the surveyed work [18–22] by the generality of the data set collected in different points of the Internet as well as by the simplicity of the proposed model.

8. Conclusions

Although there has been an increasing use of the UDP protocols by real-time applications over the Internet, the great majority of Internet traffic is carried by the TCP protocol. Nonetheless, it is highly important to have an accurate model for UDP streams since the UDP protocol does not react to network congestion. This paper introduced a simple model composed of four states that accurately reproduces the patterns existing in real UDP flows. The approach used was to characterize the marginal distributions of UDP flows so that the distributions and range of parameters used in the model can be defined. The model was validated using simulation and compared to trace driven simulation. The model was also compared to other models with different number of states as well as established models in the literature. It was shown that the proposed model is quite accurate and its use is recommended to generate synthetic data in packet level simulation experiments. Although the mix of Internet applications changes from time to time, the proposed model aggregated captures the two main characteristics of UDP streams (packet size and burst duration). Therefore, its use is recommended for different network scenario.

Acknowledgements

This work was partially sponsored by FAPESP and by CNPq.

References

[1] R. Riedi, M. Crouse, V. Ribeiro, R. Baraniuk, A multifractal wavelet model with application to network traffic, IEEE Trans. Inform. Theory 45 (3) (1999) 992–1018.
[2] J.L. Vhel, B. Sikdar, A multiplicative multifractal model for TCP traffic, Computers and Communications, IEEE Symposium on 0 2001, p. 0714.
[3] V. Ribeiro, R. Riedi, M. Crouse, R. Baraniuk, Multiscale queuing analysis of long-range-dependent network traffic, in: INFOCOM 2000, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings IEEE, vol. 2, 2000, pp. 1026–1035.
[4] C. Hollot, Y. Liu, V. Misra, D. Towsley, Unresponsive flows and aqm performance, in: INFOCOM 2003, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, IEEE Societies, vol. 1, 2003, pp. 85–95.
[5] Y. Liu, F. Lo Presti, V. Misra, D. Towsley, Y. Gu, Fluid models and solutions for large-scale ip networks, in: SIGMETRICS '03: Proceedings of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, ACM, New York, NY, USA, 2003, pp. 91–101. doi:http://doi.acm.org/10.1145/781027.781039.
[6] L. Ostrowsky, Modelos de tráfego para fluxos gerados pelo protocolo UDP, Master's thesis, Campinas State University, December 2005.
[7] L. Ostrowsky, N. da Fonseca, C. Melo, A traffic model for udp flows, in: Communications, 2007, ICC '07 IEEE International Conference on, 2007, pp. 217–222. doi:10.1109/ICC.2007.44.
[8] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, On the self-similar nature of Ethernet traffic (extended version), IEEE/ACM Trans. Netw. 2 (1) (1994) 1–15.
[9] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, The multiscale nature of network traffic: discovery, analysis, and modelling, IEEE Signal Process. Mag. 19 (3) (2002) 28–46.
[10] N.L.S. Fonseca, G.S. Mayor, C.A.V. Neto, On the equivalent bandwidth of self-similar sources, ACM Trans. Model. Comput. Simul. 10 (2) (2000) 104–124. doi:http://doi.acm.org/10.1145/364996.365003.
[11] D. Veitch, N. Hohn, P. Abry, Multifractality in TCP/IP traffic: the case against, Comput. Networks 48 (3) (2005) 293–313.
[12] Coralreef website, last access on May, 2011.
[13] T. Soong, Fundamentals of Probability and Statistics for Engineers, Wiley, 2004.
[14] A. Horvth, M. Telek, A markovian point process exhibiting multifractal behavior and its application to traffic modeling, in: 4th International Conference on Matrix-Analytic Methods in Stochastic models, Adelaide, Australia, 2002.
[15] J.A. Nelder, R. Mead, A simplex method for function minimization, The Comput. J. 7 (4) (1965) 308–313.
[16] http://www.stat.rice.edu/riedi/soft/mwm/ (accessed 07.09).
[17] TANGRAM II website, last access on May, 2011.
[18] A. Dainotti, A. Pescape, P.S. Rossi, F. Palmieri, G. Ventre, Internet traffic modeling by means of hidden markov models, Comput. Networks 52 (14) (2008) 2645–2662.
[19] P.S. Rossi, G. Romano, F. Palmieri, G. Iannello, Joint end-to-end lossdelay hidden markov model for periodic udp traffic over the internet, IEEE Trans. Signal Process. 54 (2) (2006) 530–541.
[20] K. Xu, Z.-L. Zhang, S. Bhattacharyya, Profiling internet backbonetraffic: behavior models and applications, ACM SIGCOMM Comput. Commun. Rev. 32 (4) (2005) 169–180.
[21] H. Jiang, C. Dovrolis, Why is the internet trafc bursty in short time scales? in: ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, 2005, pp. 241–252.
[22] J. Charles R. Simpson, D. Reddy, G.F. Riley, Empirical models of tcp and udp enduser network traffic from neti@home data analysis, in: 20th Workshop on Principles of Advanced and Distributed Simulation, 2006, pp. 166–174.
[23] C.A.V. Melo, N.L.S. da Fonseca, Envelope process and computation of the equivalent bandwidth of multifractal flows, Comput. Networks 48 (3) (2005) 351–375.