On the Scaling and Burst Structure of Data Traffic

Sándor Molnár, András Gefferth[1]

Dept. of Telecommunications and Telematics
Technical University of Budapest
H–1117, Pázmány Péter sétány 1/D, Budapest, Hungary
Tel: (361) 463-3889, Fax: (361) 463-3107
E-mail: {molnar,gefferth}@ttt-atm.ttt.bme.hu

Abstract

Two important and related factors in data network traffic are the burstiness and the scaling phenomena. We use several analysis techniques to capture them including also a new burstiness measure. We apply several tests for checking traffic stationarity and several methods for testing the scaling behavior. In particular, we focus on detecting and characterizing long-range dependence properties. The methods are shown with applications based on real traffic measurements.

1 Introduction

Understanding and modeling the characteristics of the traffic in packet networks are especially relevant for the design and dimensioning of our growing multimedia networks (e.g. Internet or ATM WANs and LANs) [2, 8, 12]. We can see a number of new applications on these networks which dynamically change the traffic structure being carried.

Traffic *burstiness* is one of the key properties observed in data traffic [8, 13]. Intuitively, burstiness is present in data traffic when the arrivals (e.g. packets) appear to form clusters in time. Another indication of burstiness is that several relatively short interarrival times are surrounded by relatively long ones. In spite of the simple intuitive understanding the mathematical definition of burstiness is more complex and not so obvious. There are several different burstiness measures introduced in the teletraffic literature [8, 13]. The two important factors that influence burstiness are the marginal distribution and the autocorrelation of the arrival process. For example, strong short-range correlations induce burstiness. We should note that in several practical cases the first- and even the second-order characteristics of the arrival process are not adequate for a complete characterization of burstiness and higher-order description is also required.

Scaling behavior is another important key characteristic of traffic in today's data networks [7]. It has a strong connection to the above mentioned burstiness because burstiness should be defined in terms of time scales over which clustering activities occur. The surprising scaling phenomenon observed in data traffic is that these clustering activities are present over several time scales. It is in contrast to traditional telephone traffic characteristics where the traffic burstiness is smoothed out by increasing the time scale. In other words, there is an absence of characteristic time scale and a bursty traffic over many time scales with a complex "burst within burst" structure is observed. This phenomenon triggered a new modeling approach, called *fractal modeling*, which can offer parsimonious models for this behavior [12, 24]. Even if scaling is identified, it is not simple to describe how different statistical characteristics are scaled on different time scales. In many cases there are different scaling rules on different time scales and the traffic has a *multifractal* structure [6, 12].

In this paper we investigate both scaling and burstiness characteristics of actual measured data traffic.

We report some results concerning the traffic structure of measured IP traffic on an ATM link. These measurement have been made on our University ATM campus network connected to the Internet. The traffic generated by several hundred workstations represents a typical aggregated University traffic to and from the Internet. The details of our measurements are described in Section 2.

There are four targets of our traffic analysis reported in the paper. First, we have analyzed the traffic contributed to the aggregation concerning their volumes, bandwidth shares and application types. In Section 3 we can see the dominant applications and their properties.

Second, prior to any statistical analysis we performed several tests for checking the stationarity of the measured traffic. With our methods we can select stationary periods of the traffic which are the only candidates for traffic characterization and modeling. Therefore we avoid pitfalls and misinterpretations due to the nonstationarity effects present in real traffic. These results are shown in Section 4.

Third, the scaling structure of the traffic has been investigated in Section 5. We have carried out a scaling analysis study including different statistical methods (variance-time plot, periodogram, R/S analysis, wavelet analysis, etc.) to reveal the scaling and correlation structure of the traffic. Among the scaling properties we focus on the long-range dependence (LRD) analysis of the traffic because LRD has a significant impact on performance issues.

[1] The research was supported by the Inter-University Centre of Telecommunications and Informatics

Fourth, we introduce a new burst structure characterization method in Section 6. We analyze the bursts and idle periods concerning their distribution and correlation characteristics.

The contribution of this paper is to report traffic characteristics with analysis techniques considered to be significant in packet switched network traffic.

2 Traffic Measurements

The purpose of this section is to present our measurements analyzed in this paper. The traffic was measured in a University environment.

2.1 Network Topology

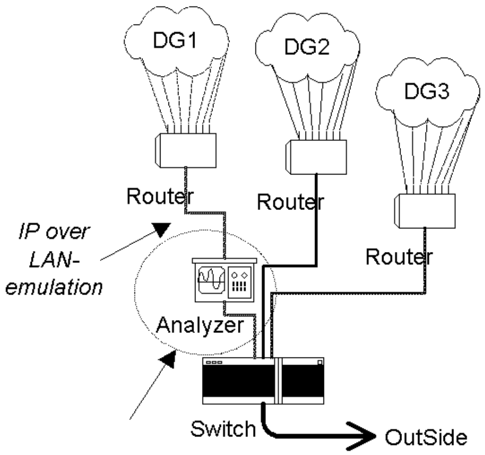


Figure 1: Network and measurement setup

Figure 1 shows the network and measurement setup. The network is located in the new University building of the Technical University of Budapest. The building is connected to the outside world by a 100MB FDDI and a 155MB ATM link. An ATM backbone is installed inside the building connecting Ethernet based local networks. These local networks are referred to as Department groups (DG). Each DG consists of about 100 workstations. Ethernet frames are transmitted over the ATM backbone using LAN emulation.

In our measurement a specific department group (DG1) was selected and the traffic between DG1 and the rest of the world was measured. DG1 contains about one third of all the workstation of the Department of Telecommunications and Telematics and the Department of Telecommunications. These workstations belong to department staff, PhD students, and student laboratories with a variety of operating systems. The network interfaces range from 10Base2 (BNC) through 100BaseT (UTP) to 100VGAnyLan. Some bridges and routers are also located in DG1. Users mainly generate IP and IPX packets.

2.2 Measurement tools

A traffic monitoring tool, called CapTie [22], was located on the 155 MB/s ATM link between DG1 and the switch. At the time of the measurement the analyzer had limited capabilities. It recorded the preselected data objects (IP packet, ATM cell, bytes) passing the link. It could measure a few different basic statistics at the same time. To get data from the analyzer an SNMP [19] request had to be sent requesting the value of the desired counter. In order to get detailed information a periodical polling of the counters was needed. This polling was carried out by an SNMP client program that ran on a workstation which was connected to the analyzer via the Ethernet network. The client program was based upon the cmu-snmp package.

2.3 Measured data

During our measurement the number of bytes in IP packets crossing the link in both directions on three different days: 14th, 16th and 19th of April 1999, from 8am to 6pm was logged. We investigated different layer protocols and also counted the number of non-IP packets. These protocols include transport layer protocols such as ICMP, TCP, UDP, and OSPF, as well as application layer protocols, FTP, SSH, Telnet, SMTP and HTTP. The granularity of our measurement was one second.

2.4 Measurement errors

The SNMP request packages from the client to the analyzer had to cross over an Ethernet segment which introduced some jitter in the request arrival time. This resulted in inaccuracy of the measured data. The workstation which collected the data ran a multitasking operating system with only a limited real-time support. Exact scheduling of the processes was not possible which was another source of

Error Source	Min. Delay	Max. Delay	Std. Dev
Network	3.4	910.2	93.8
Scheduling	0	152.3	1.3

Table 1: Measurement errors resulting form the two error sources in a typical dataset. The values are measured in msec. The minimum and maximum delays observed during the measurement and the standard deviation of the delays are reported.

inaccuracy. Table 1 shows the errors from these error sources in a typical dataset. It should be mentioned that the difference of the maximum and minimum delay is high in the case of the first source of error. However, it is only approximately 1% of the values where the delay was larger than 2msec plus the minimum delay, in this dataset, but in other datasets, this ratio is even smaller.

As can be read from Table 1 in our measurement the first source of inaccuracy was dominant. It is described in Section3.2 how these errors were handled.

3 Basic traffic analysis

The calculations presented in this section deal with the amount of data passing the ATM link. We focus on the contribution of different applications to the bandwidth usage and the total amount of data on different days.

3.1 Traffic data

Our data which serves as input for the calculations consists of several time series, also called traces.

We classified IP packets by their transport layer protocol identifier, as well as by their well-known TCP port number. Using these pieces of information one can decide which transport and application layer protocol data unit (PDU) is transmitted in the packet. There was no distinction between request and response PDUs in case of client-server services, such as FTP or WWW. The given trace contains the bytes belonging to both.

The following protocols have been observed:

- Transport Layer Protocols: UDP, TCP, ICMP and OSPF.
- Application Layer Protocols: FTP, HTTP, SMTP, Telnet and SSH.

The total number of bytes belonging to IP packets was also measured. These traces are referred to as *total-IP* traces and are used in the correlation and burst analysis in Sections 4, 5 and 6.

3.2 Preprocessing raw measurement data

Preprocessing data consisted of the following steps:

- correcting errors caused by jitter and
- producing the per second data from the values of the aggregated counters.

The correction was done by dropping values where the round trip delay from the SNMP agent to the analyzer and back was bigger than a given tolerance. The lost and dropped values were replaced by linear interpolation using the neighboring values. A typical dataset contained 36000 data, from which about 1-200 had to be dropped. The timeout was set to 1msec which compared to the 1 sec granularity of the measurement is not more than 0.1%. This inaccuracy is acceptable for our purposes.

3.3 Traffic intensities

Before doing any calculations on measured dataset a visual investigation is always useful. In Figure 2 the total-IP traces are depicted in direction out on all the three days. We can conclude that the traffic is bursty on many scales. A few huge peaks can be observed, which are much bigger than any of the other peaks. Removing these peaks we get a similar result: the remaining high peaks are still much bigger than the others, and so on.

In Figure 3 the same three datasets are depicted, where the values were aggregated over 10 minutes interval. In this picture we see that a daily variation similar to those observed by PSTN networks can not be seen clearly.

Figure 4 and Table 2 show the total amount of IP-traffic passing through the link during our measurement in both directions on each of the three days. We see that the volume of data coming in was always more than the volume of data going out.

We explain this observation by the fact that FTP and web access is asymmetric because outside world servers are more popular than the servers in the university. We found (see Section 3.4) that the majority of data is carried using FTP and HTTP protocols. We can also see that the amount of traffic on different days is significantly different.

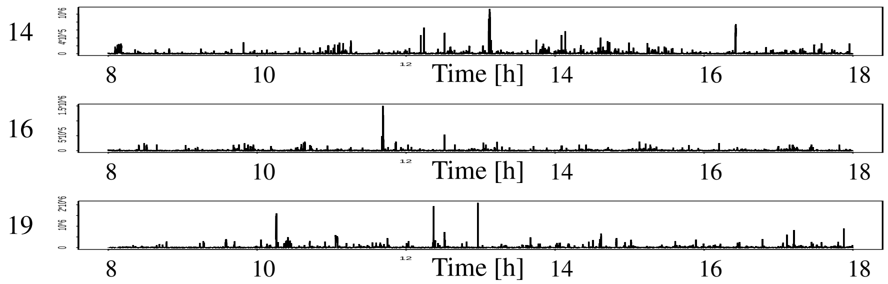


Figure 2: Intensity of the traffic (time unit: 1 sec)

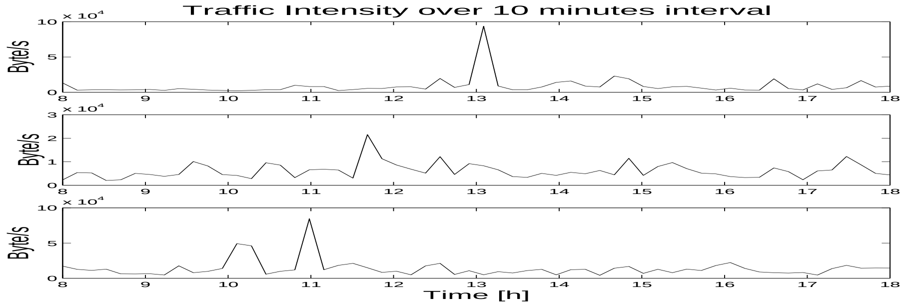


Figure 3: Intensity of the traffic (time unit: 10 min)

Day	14th Apr.	16th Apr.	19th Apr.
Out	325	221	498
In	534	835	869

Table 2: Total traffic on different days measured in Megabytes

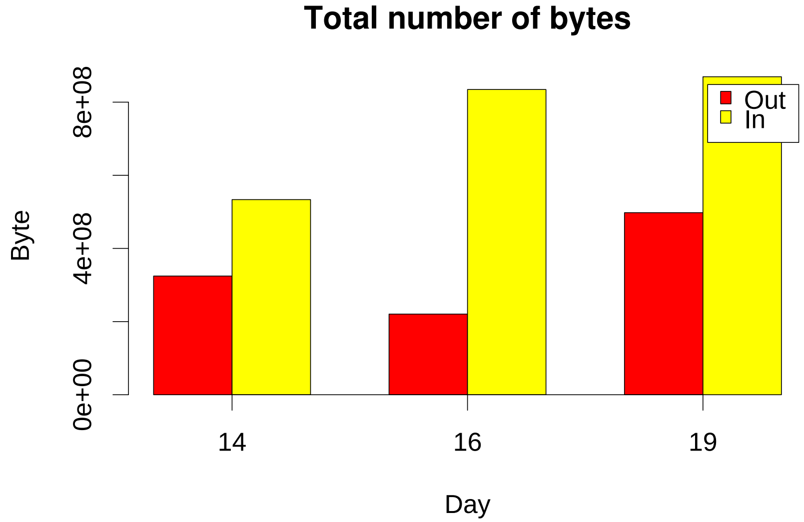


Figure 4: Total traffic on different days

3.4 Bandwidth share

The ratio of different protocols contributing to the overall load was calculated. Figure 5 shows the bandwidth share of different protocols averaged over all of our measurement data. Figure 5(a) shows the transport layer and Figure 5(b) the application layer protocols.

The numerical values are shown in the middle rows of Tables 3(a) and 3(b).

Besides averaging over all the data, the computation was performed separately for individual days and directions. The top and bottom rows of Tables 3(a) and 3(b) show the minimum and maximum bandwidth share of the protocols.

	TCP	UDP	ICMP	OSPF
Minimum:	85%	5%	1%	0%
Average:	90%	8%	1%	1%
Maximum:	93%	12%	2%	1%

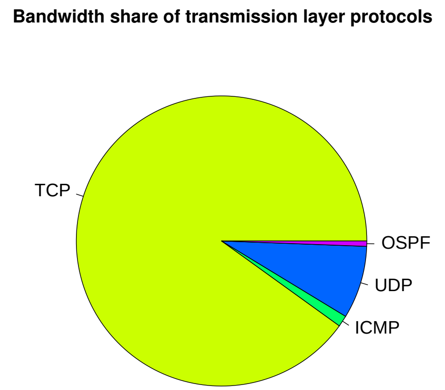
(a) Transport Layer Protocols

	FTP	HTTP	SMTP	TN	SSH	Other
Minimum:	18%	21%	3%	2%	0%	-
Average:	35%	29%	6%	2%	1%	27%
Maximum:	44%	40%	9%	3%	1%	-

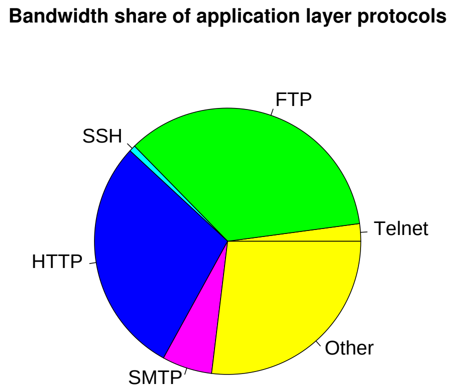
(b) Application Layer Protocols

Table 3: Bandwidth share of protocols at two different layers. The middle row shows the overall average. The top and bottom rows show the minimum and the maximum ratios measured on a daily basis.

We conclude, that the majority of data is carried by FTP and HTTP protocols. Although the bandwidth share of the applications changes from day to day the dominance of TCP, especially FTP and HTTP could be observed in all datasets. These sources of data do not need real-time transmission but a low response time is desirable.



(a) Transport Layer Protocols



(b) Application Layer Protocols

Figure 5: Ratio of the different protocols

4 Stationarity analysis

In the previous section the contribution of different protocols to the network traffic was investigated. In this and the next sections the aggregated traffic is investigated. We will use the total-IP traces introduced in Section 3.1.

Most of the existing tools for analyzing real traffic assume the stationarity of the traffic. Therefore before using these tools the assumption of stationarity should be validated.

It is usually not realistic to assume the stationarity of the whole traffic trace since traffic intensity varies during the time. A straightforward approach to overcome this problem is to select intervals, where traffic can be handled as stationary, and analyze them independently. In this section our method of selecting stationary intervals is presented. In Section 5 the results of analyzing these intervals is discussed.

Without any a priori information about the generation of the process the determination of stationarity based on a measured finite data set is mathematically impossible. Even if a given family of models exists, where the process is assumed to be a realization of one of them, only a statistical test can be performed. However, this test always has a probability of false decision.

Two heuristic mathematical methods have been used and the results have been combined to select stationary intervals.

4.1 Stationarity of distributions

This method is based on the changepoint detection method described in [18]. A window is slid along the data and the distribution of the samples in the two halves of the window are compared, see Fig. 6. Should the two distributions be significantly different we reject the assumption of stationarity and assume a changepoint near the middle of the window. It should be noted that (1) this method was developed for uncorrelated dataset, and (2) the equivalence of the two distributions is necessary but not sufficient for stationarity. Because of these limitations we also use another method described in the next subsection.

We choose a half-window size denoted by N. In the first stage we would like to determine from all neighboring intervals of length N whether the samples in the two intervals have the same distribution. Therefore we slide our window along the process, and at each point we perform a two sample Kolmogorov-Smirnov [9, 18] test to decide the equivalence of the two distributions.

We select those intervals

1. which are at least as long as the window,

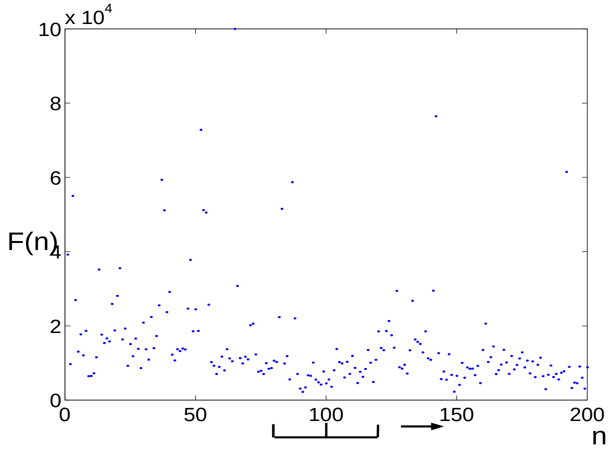


Figure 6: Sliding a double-window along the interval

2. all neighboring intervals proved to have the same distribution, and

3. the interval can not be extended to have the same property.

The value of N is determined by a trade-off between having enough data for a reliable statistical calculation, and at the same time have a fine picture of the data. The value of 600 was used in our studies.

Next we describe the applied Kolmogorov-Smirnov test. The purpose of the test is to decide whether two sets of samples are generated by the same distribution or not. A general description about statistical tests can be found in [10]. The test statistics is $D = \max|F(x) - G(x)|$, where $F(x)$ and $G(x)$ are the empirical distribution functions of the two samples. (See Figure 7.) $D(n)$ is the value of the test statistics when the middle-point of the window is at sample n.

If the two samples come from the same distribution then the following equation holds:

$$\lim_{N\to\infty} P\left(\sqrt{\frac{N}{2}}D < y\right) = K(y), \quad 0 < y < \infty,$$

Where $K(y)$ is the Kolmogorov distribution function, defined as

$$K(y) = 1 - 2\sum_{j=1}^{\infty}(-1)^j e^{-2j^2y^2}$$

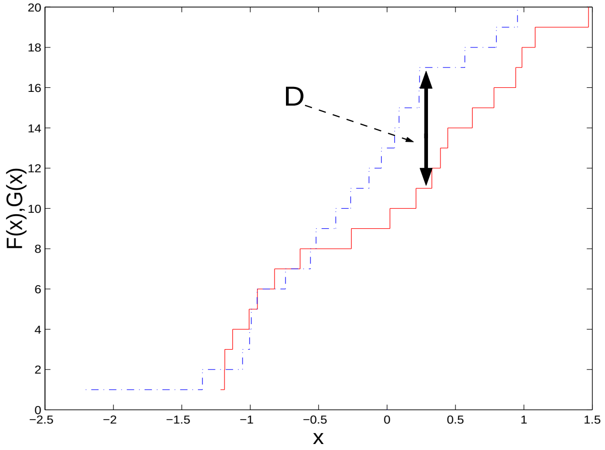


Figure 7: Calculating the Kolmogorov-Smirnov statistics from the empirical distributions

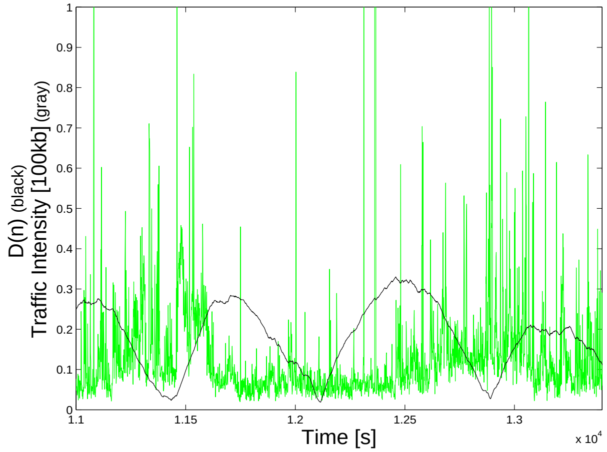


Figure 8: Kolmogorov-Smirnov statistics and the original time series (scaled)

We set the significance level of the test to 0.1.

Figure 8 shows a part of the input dataset together with the value of the statistic $(D(n))$, where the middle point of the interval is aligned with the data. (The original dataset is scaled in order to fit in the figure.)

4.2 Smoothness of the process

The purpose of this method to identify stationary intervals of the sample by capturing the periods with the same burstiness (smoothness) characteristics.

The original process (X_i) is averaged over neighboring intervals of length M to form a new process $X_k^{(M)}$.

$$X_k^{(M)} = \frac{1}{M} \sum_{i=(k-1)M+1}^{kM} X_i$$

The sample variance of this new process is denoted by σ^2.

$$\sigma^2 = \frac{1}{L-1} \sum_{i=1}^{L} \left(X_i^{(m)} - \widehat{X^{(m)}} \right)^2,$$

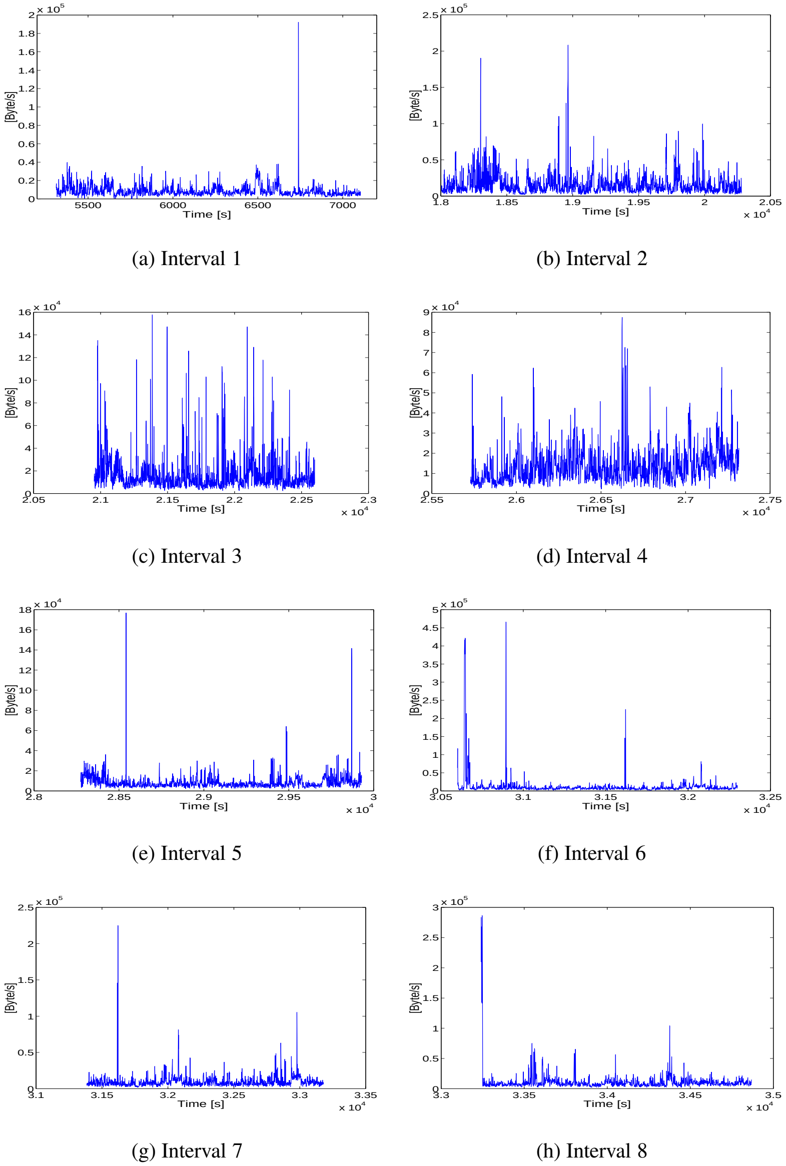
where $\widehat{X^{(m)}}$ is the average of the process $X_k^{(M)}$, and L is the length of the same sample.

The longest possible intervals have been selected where the difference of adjacent values was smaller than $\alpha\sigma$. For our analysis α was set to one, and M was set to 60.

4.3 Selecting the intervals

Both previous methods were applied to select the stationary intervals. The intersections of the selected intervals were used for further study. To have enough data for statistical analysis only intervals with a minimum length of 1500 samples were used.

For further study only one of the measured traces was chosen, the traffic trace measured on the 14th of April, direction in. The selected intervals are listed in Table 4 and depicted in Figure 9. The selected intervals might overlap since the changepoint based test might select overlapping windows. The selected 8 intervals cover 13167 sec, which is 37% of the total trace.



(a) Interval 1 (b) Interval 2

(c) Interval 3 (d) Interval 4

(e) Interval 5 (f) Interval 6

(g) Interval 7 (h) Interval 8

Figure 9: Intensity plot of stationary intervals, time unit: 1sec

Int.	1	2	3	4	5	6	7	8
Start	5312	18006	20952	25728	28274	30601	31383	33241
End	7104	20280	22598	27315	29929	32300	33180	34868
Len.	1793	2275	1647	1588	1656	1700	1798	1628

Table 4: Endpoints of the stationary intervals measured in seconds elapsed from the beginning of the trace

5 Correlation Studies

In the previous section it was described how the intervals to be analyzed have been selected. In this section the correlation structure of the process is analyzed, separately for the long- and short-term correlations. A so-called scaling behavior is reported to be present in network traffic traces. Scaling behavior means that fundamental properties of the traffic are left unchanged when viewed at different time-scales.

Different types of scaling behavior was detected in computer network traffic, including long-range dependence, self-similarity, and fractality [11, 15, 16, 17]. This paper focuses on the long-range dependent (LRD) structure of the process.

5.1 Autocorrelation function

The autocorrelation function of a process X is defined as

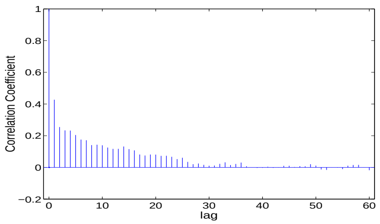
$$\rho_X(i,j) = \frac{E(X_i - \bar{X}_i)(X_j - \bar{X}_j)}{\sigma_{X_i}\sigma_{X_j}},$$

where $\bar{X}_s \overset{def}{=} E(X_s)$, and $\sigma^2_{X_s}$ is the variance of X_s. In case of weakly stationary processes this definition reduces to

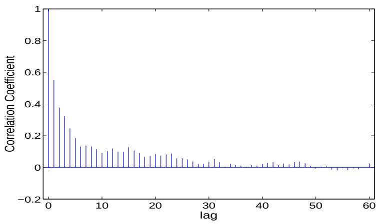
$$\rho(n) = \frac{E(X_i - \bar{X})(X_{(i+n)} - \bar{X})}{\sigma^2_X},$$

where $\bar{X}$ and σ^2_X denote the common mean and variance of the values, respectively. In practice the expected values are replaced by time averages.

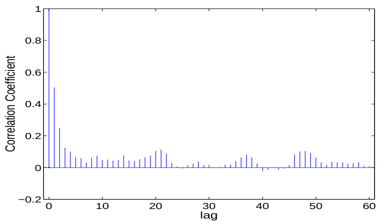
The calculated values for the stationary intervals for the first few lags (time unit: 1sec) are shown in Figure 10. Most of the results show strong short-term positive correlations. As we increase the lag the correlations become small but in most of the cases the decay of the autocorrelation is slow (except in Figure 10 (h)). In some cases the signs of periodicity can also be observed (cases of (c) and (d)).



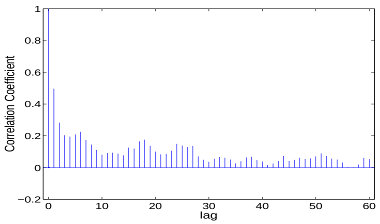
(a) Interval 1



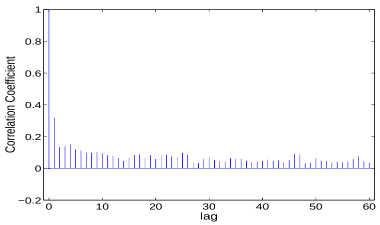
(b) Interval 2



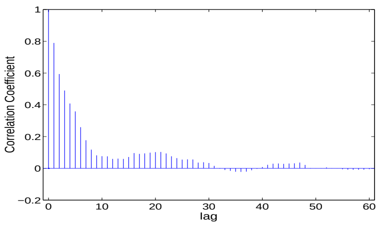
(c) Interval 3



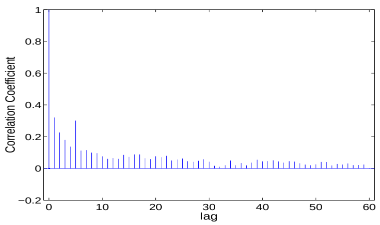
(d) Interval 4



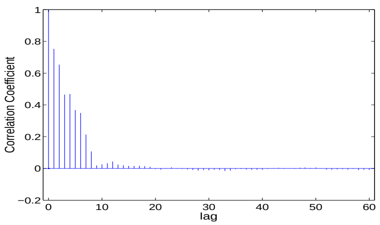
(e) Interval 5



(f) Interval 6



(g) Interval 7



(h) Interval 8

Figure 10: Autocorrelation function

5.2 Long-range dependence

In Section 5.1 the autocorrelation function was introduced and presented for small lags. In this section we focus on the asymptotic behavior of the autocorrelation coefficients at large lags.

The framework of LRD processes allows us to capture important properties of such traffic with only a few parameters [3]. The most important of these parameters is the Hurst parameter, denoted by H. Its value is between 0.5 and 1 for LRD processes. The higher the H the stronger the correlation. In this section our definitions and notations are given.

A weakly stationary stochastic process is called long-range dependent if its autocorrelation function ($\rho_X()$)decreases according to a power-law at large lags.

That is

$$\rho_X(k) \sim c_\rho |k|^{2H-2},\ k \to \infty, H \in (0.5, 1) \tag{1}$$

The equivalent of this definition in frequency domain is:

$$|f_x(\nu)| \sim c_f |\nu|^{1-2H},\ \text{as } \nu \to 0, \tag{2}$$

where $f_x()$ is the spectral density function [4]. For a given value of H, c_f and c_ρ differ in a constant factor only. The scaling property that LRD sources exhibit is that the autocorrelation function converges to a fixed function when the process is averaged over neighboring blocks. Let $X_k^{(m)}$ denote the new series:

$$X_k^{(m)} = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i,$$

where X_i is the original dataset. If the autocorrelation function of $X_k^{(m)}$ is denoted by $\rho^{(m)}(.)$ then

$$\lim_{m\to\infty} \rho^{(m)}(.) = \rho_\infty(.),$$

where $\rho_\infty(.)$ is a fixed function determined by H.

The Hurst parameter is very important since it defines the existence and strength of the LRD. However, it only determines one of the two parameters used in Eq. 1 and Eq. 2. The other parameter, c_ρ or c_f - depending on the content - is called the *second parameter* following the terminology of [23]. It is important to note, that both H and the second parameter describe the asymptotic properties at large lags and do not give any information about the short term correlation structure. Most of the methods found in the literature concentrate on estimating the value of H and little effort is made to give an estimate for the second parameter.

5.3 Parameter estimation methods

Several methods exist in the literature to estimate H. Some of these are heuristic methods. These methods are adequate to indicate the presence of long-range dependence but give no confidence interval, so the estimated values should be handled with caution.

The methods are not described here in detail. They can be found in a number of papers, including [1, 3, 16, 17, 23]. A more advanced estimation method is the Abry-Veitch estimator [23], which estimates both c_ρ and H (see Eq. 1) and gives confidence intervals. We show this method in the next subsection.

The **variance time plot** (VTP) is based on estimating the decay of variance when the samples are averaged over increasing windows. If X_k is a stationary LRD dataset, then [3]

$$\lim_{n\to\infty} \mathrm{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{c_\rho}{H(2H-1)} n^{(2H-2)} \tag{3}$$

For short-range dependent dataset ($H = 0.5$) the autocorrelation is summable and the variance of the time-average decreases as n^{-1}, where n is the block size. For LRD data this decrease is slower, $n^{-\beta}$, with $\beta = 2 - 2H$, as can be read from Eq. 3. In practice the the variance is estimated by the sample variance. The estimated variance is plotted against the block size (n) in a log-log plot and a linear regression is performed to fit a line on the values. H can be calculated from the slope of the line. In theory c_ρ could be calculated from the intersection but in practice this estimation method is not used.

The **periodogram** [3] estimation is based upon property of Eq. 2. The spectral density of the process is estimated by the periodogram which gave the name of this estimation technique. Plotting the estimated spectral density function on a log-log scale should produce a linear line at low frequencies with a slope of $1 - 2H$.

The **rescaled adjusted range statistics** (R/S) [3] is another estimator of the Hurst parameter. For a discrete time stochastic process, X_j, the rescaled range of X over a time interval n is defined as

$$\frac{R}{S} = \frac{\max\{W_i : i = 1, 2, \ldots, n\} - \min\{W_i : i = 1, 2, \ldots, n\}}{\sqrt{\mathrm{var}(X)}}$$

where $W_i = \sum_{k=1}^{i}(X_k - \bar{X}), i = 1, 2, \ldots, n$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. It can be proven that in case of LRD processes

$$\frac{R}{S} \sim \left(\frac{n}{2}\right)^H$$

The value of H is estimated by regression. In practice R/S is plotted against n on a log-log graph and a straight line is fitted to the values. The slope of the line gives H. This is the original method used by Hurst.

Since LRD describes the low-frequency behavior, only the low frequencies should be taken into account during the estimates. This means large window sizes for the variance-time plot, and low-frequencies for the periodogram method.

5.4 The Abry-Veitch wavelet analysis

Abry and Veitch gave a more sophisticated **wavelet based estimator** (AV) for the Hurst and the second parameter [23]. The advantage of this method is that the long-range dependent process in the time domain is converted to weakly correlated coefficients in the wavelet domain, which makes the estimation of the parameters more reliable. This method proved to be robust again some changes in the nature of traffic [21].

Wavelet transforms and the log-scale diagram will be introduced without the need of rigorous mathematical precision. The details can be found in [1, 23]. Discrete wavelet transformation (DWT) transforms a discrete time process from the time domain into a two dimensional discrete wavelet domain. The wavelet coefficients of a process X are denoted by $d_x(j, k)$. Wavelet transformation is based on a family of functions which are all transformed versions of the so called mother wavelet denoted by $\Psi_0(t)$. The choice of mother wavelet depends on the actual task. These functions have a common property that they are bounded in a sense in both the time and frequency domain. As a result these coefficients hold information about both time and frequency domain behavior.

Wavelets have a number $N \geq 1$ of vanishing moments, which means that if a mother wavelet with $N = n$ is chosen all polynomial parts of the signal of order less than n vanish after the wavelet transform. Formally $\int t^z \Psi_{j,k}(t)dt = 0, z = 0, 1, \ldots, N-1$, where $\Psi_{j,k}$ can be any wavelet base function, i.e. transformed version of the mother wavelet. The value of N can be freely chosen. The greater the value of N the more insensitive is the transformation to 'smooth' (polynomial) parts of the signal.

Coefficients $d_x(j, k)$, where $j = j_0$ is fixed $k = 1, 2, \ldots$ belong to the same frequency referred to as *scale* in wavelet framework. Parameter j is called *octave* and is equal to $\log_2(\text{scale})$. However, the term 'scale' is also used to refer to the logarithm of scale.

It is known (see, e.g. [23]) that in case of LRD processes

$$E\left[d_x(j, .)^2\right] = 2^{j\alpha} c_f C$$

where

$$C = \int |\nu|^{-\alpha} |\Psi_0(\nu)|^2 d\nu$$

is a constant that depends on α and the mother wavelet, $\Psi_0()$. We note that $\alpha = 2H - 1$. $Ed_x(j,.)^2$ is estimated from the values of $d_x(j,k), k = 1, 2, \ldots$ and plotted against j on a log-lin scale.[2] The values should produce a straight line. The value of α can be calculated from the slope of the line, while the intersection gives the second parameter. The calculated values do not closely align a straight line due to the lack of any strong correlation between them. This is an advantage of this method since strong correlation could make the line look more straight but the estimated slope will be biased, which is the case, for example, at the variance time plot method. The alignment of the values therefore should not be judged by eye. A weighted linear regression is performed to fit a line over the estimated data.

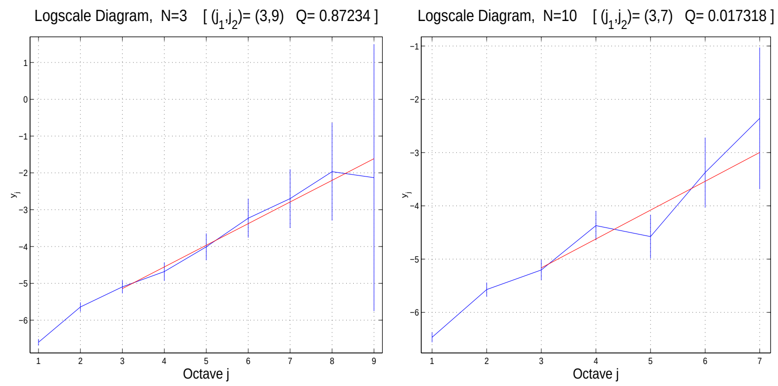
The results depend on the mother wavelet used and on the choice of scales over which the linear regression is performed.

As an example, we investigate the fractional Gaussian noise, which is the only stationary, zero mean LRD process with Gaussian marginal distribution. Figures 11(a) and 11(b) show the log-scale diagram of the same fractional Gaussian noise dataset generated with Hurst parameter 0.8. The difference is the number of vanishing moments of the mother wavelet. (With the increase of the vanishing moment, the maximum available scale decreases. This is a technical limitation of the method.)

On the x-axis the octave j is plotted, and the y-axis shows the estimated value of $Ed_x(j,.)^2$ with the confidence intervals. Note, that with the increase of scale, which is equivalent to the decrease of frequency, less and less details are available which makes the estimation less reliable and the corresponding confidence intervals larger. The number of vanishing moments (N), the scales over which the straight line is fitted (j_1, j_2) and the goodness of fit (Q) can be read from the title line of the figures. Since long-range dependence is a property of the low-frequency behavior, in case of a LRD process the values align from a lower cut-off scale, (j_1) to the maximum scale available in the dataset. If this is not the case it does not simply mean that the data is not LRD. Even short-range dependent processes with summable autocorrelation coefficients have a 'trivial scaling' behavior with $H = 0.5$. In this case the slope of the fitted line is 0. If the values do not align it can be because of two main reasons. Either the scaling behavior starts at a lower frequency which is not present in the given data with finite length, or the data has a more complicated structure, possibly due to non-stationarity.

[2] This, in fact, is a log-log scale since j is a logarithmic scale.

These figures show that the change of the number of the vanishing moments does not result in drastical change of the estimated parameters, nor in drastical change of the shape of the graphs. Although not detailed here one can believe that the slope and intersection of the fitted line does not strongly depend on the values of j_1 and j_2. This is in contrast with real data, where the calculated values do depend on the fitting region, and might also depend on the number of the vanishing moments.



(a) $H = 0.794, c_f = 0.0067$ (b) $H = 0.771, c_f = 0.0073$

Figure 11: Log-scale diagram of fractional Gaussian noise (fGn) dataset. The value of N is the number of vanishing moments. The j_1 and j_2 give the range over the least square line was fitted and Q gives the goodness of fit. The critical level of rejecting the null-hypothesis that the values form a straight line. The value is between 0 and 1, the higher the better. (A value of about 0.1 is generally sufficient.) Vertical lines give the 95% confidence intervals)

5.5 Results of LRD tests

The sequences of arrived bytes in every 1sec interval were analyzed in each test.

Variance-Time Plot

The results of this method for all the selected intervals can be seen in Figure 12. The estimated values, which can be read form the picture, are very close to each other suggesting that the Hurst parameter did not change during time. We can observe that the values in the subfigures 12(h) and 12(g) are far from a straight line, 12(b), 12(c), 12(d), 12(e), 12(f) are close to a straight line, and the values in subfigure 12(h) almost perfectly fit, as far as the low-frequency (large aggregation level) is concerned.

As we will later see this has nothing to do with the accuracy of the estimate, since, e.g. the variance-time plot estimate of interval 8 is very far from values estimated with other methods but the estimate for interval 1 corresponds well to the other estimates.

R/S plot

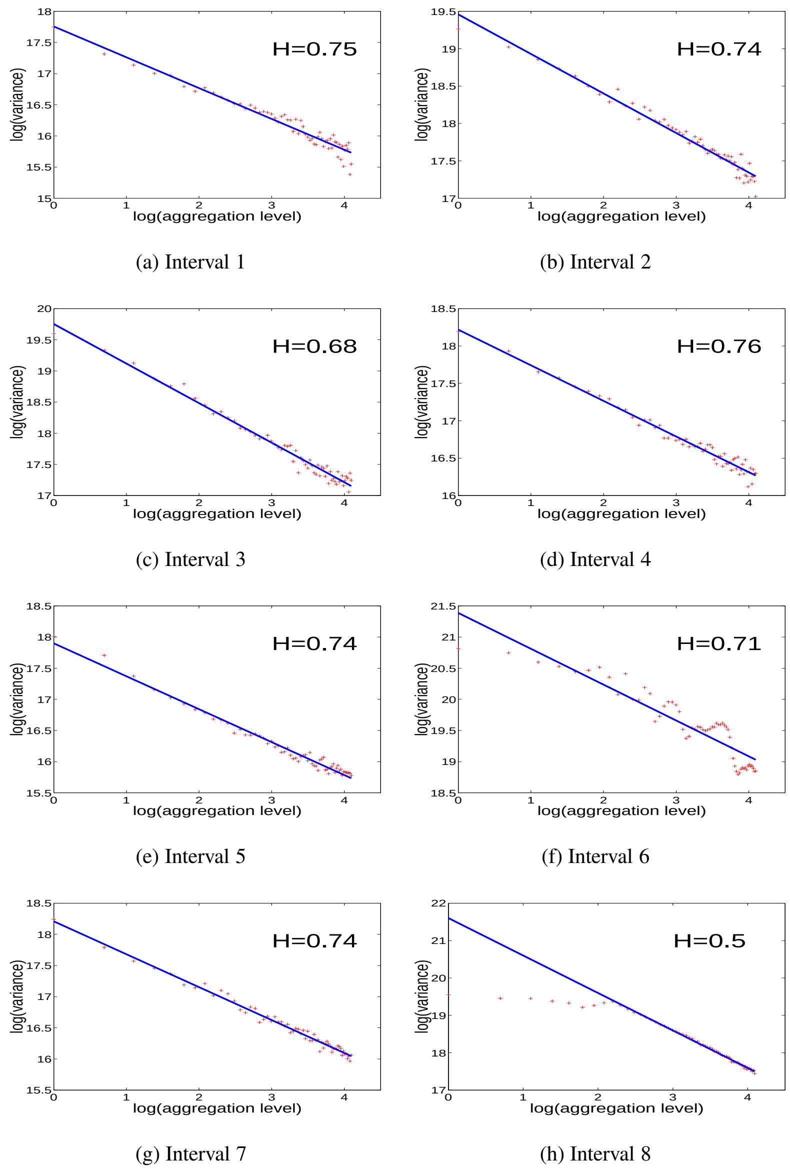
The results of this method can be seen in Figure 13. In all of the subfigures of Figure 13 the linear alignment of the values can be observed. This is not a characteristic of the specific dataset, the RS analysis produces straight looking graphs for most datasets measured in practice.

Periodogram

The results of this method can be seen in Figure 14. In case of the periodogram method the estimated H value depends on the region over which the line was fitted. This region should start from the leftmost value available and be sufficiently long to be able to perform a linear regression but it should not be too long since only the low-frequency values are desired to have an impact on the estimate. As in the case of the variance time plot estimate interval 8 shows significantly different behavior at low frequencies.

Wavelet analysis

Figure 15 shows the log-scale diagram for the second dataset for different number of vanishing moments (N). It can be observed that although in all cases the values align a straight line but the calculated values and the region where the values align (j_1, j_2) depend on N.



(a) Interval 1
(b) Interval 2
(c) Interval 3
(d) Interval 4
(e) Interval 5
(f) Interval 6
(g) Interval 7
(h) Interval 8

Figure 12: Variance time plot

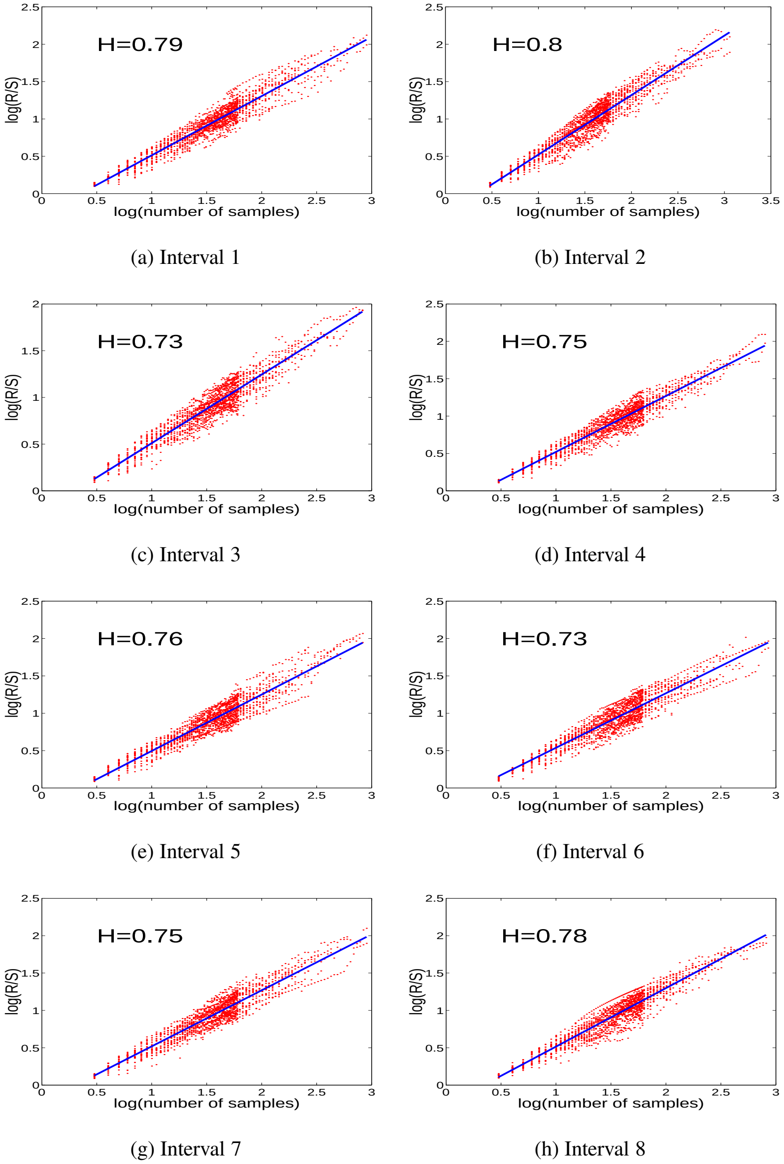


Figure 13: R/S plot

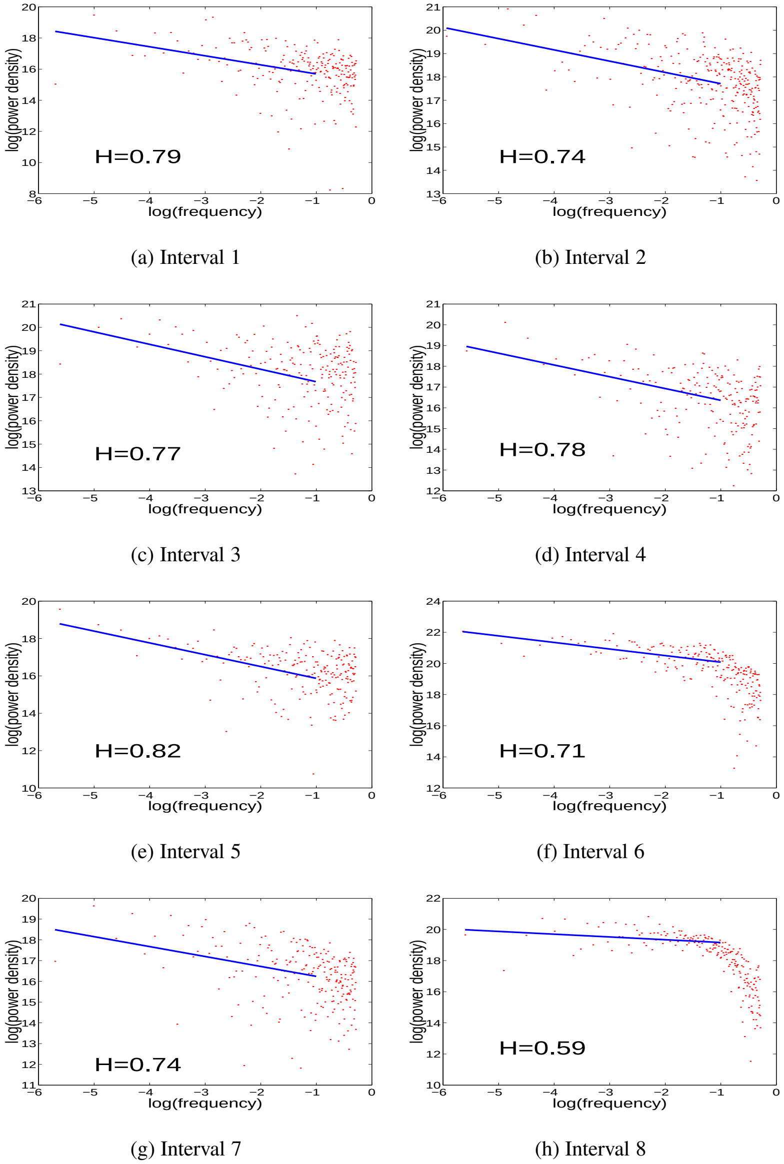
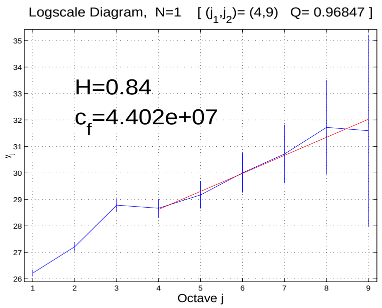
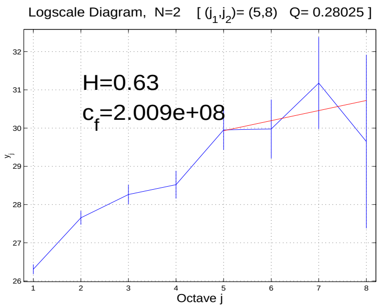


Figure 14: Periodogram

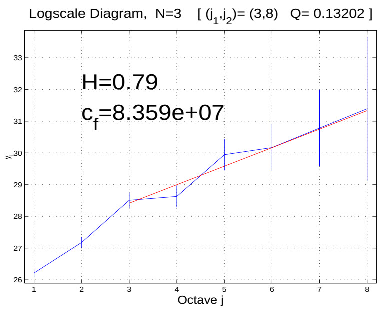
We have already seen in Figure 11 that in case of fractional Gaussian noise (fGn), which is a pure stationary and zero mean self-similar process, that the calculated values do not really depend on N. It is due to the fact that fGn is an exact self-similar process having the same scaling parameter at each time scale. The results show that the measured traffic has a more complicated structure than an fGn. The values estimated at different values of N should be compared with the corresponding confidence intervals. It is done later in this section.



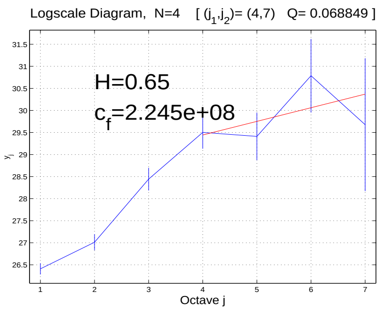
(a) Number of vanishing moments= 1



(b) Number of vanishing moments= 2



(c) Number of vanishing moments= 3



(d) Number of vanishing moments= 4

Figure 15: Logscale diagram for the second interval with different number of vanishing moments

Comparison of results

All the results calculated with the different methods are depicted on the same figure (see Figures 16(a) to 16(h)). The first few values give the results of the AV estimation with different number of vanishing moments with the 95% confidence intervals. The last values are the values from the variance time plot, R/S plot and periodogram plot, respectively. These methods do not give any confidence intervals. The horizontal line is the average of the values. Since the last three methods do not give any confidence interval a weighted averaging would be pointless. The fitting region (j_1, j_2) of the wavelet estimates was selected by manually investigating the goodness of fit for different values of j_1 and the largest scale j_2 available in the dataset.

For the next discussion exclude the last three values and concentrate only on the wavelet estimates. In case of intervals 1 and 3, the estimated values, correspond well to each other, the confidence intervals overlap and the estimations are close to each other.

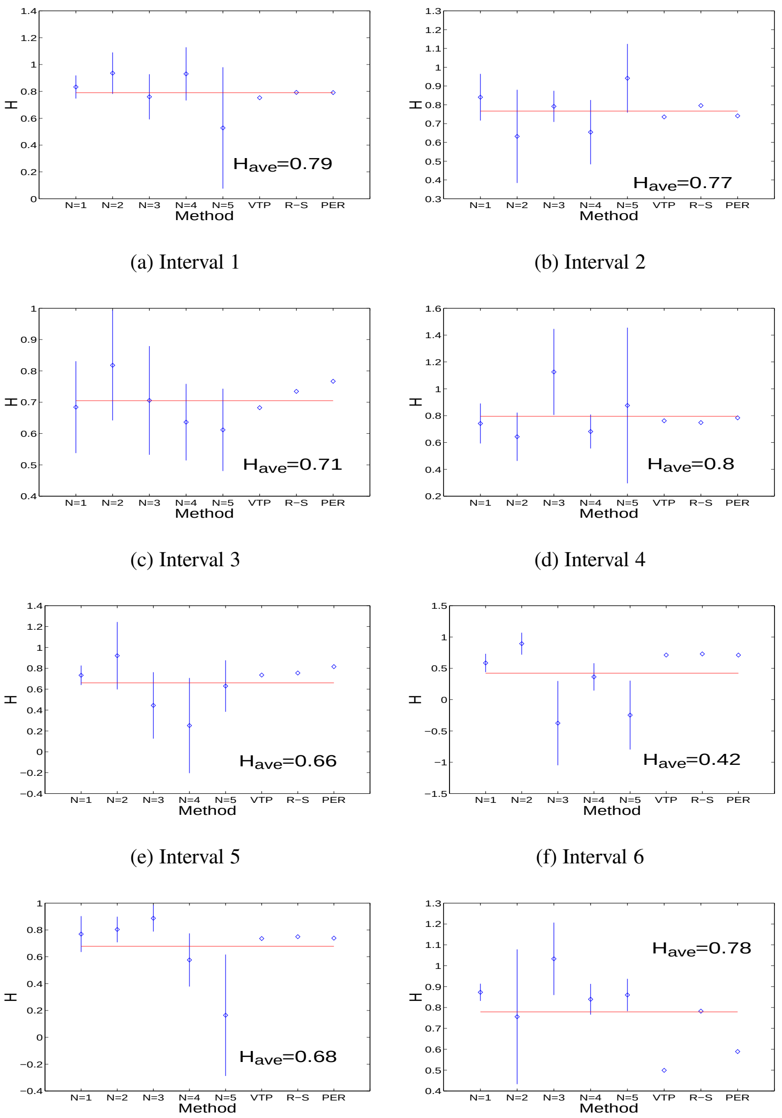
In case of interval 2 the results are not that good but still acceptable. In Figure 16(h) we can see that estimates corresponding to $N = 2$ and $N = 3$ are quite different form the other estimates but the corresponding confidence intervals are much larger. This shows that these values are less reliable estimates and therefore an average of the other three estimates fits in all the confidence intervals.

Although at the other intervals we might find a value that fits into all the confidence intervals, the estimated values are far from each other and it is very unlikely that the process has a simple LRD structure which can be characterized by two parameters.

Now we can discuss the results of the heuristic methods. In the case of the first three intervals, the heuristic estimates correspond well to the wavelet estimates.

In the case of the 4th, 5th, 6th and 7th interval the values are close to each other and to the average of the wavelet estimates but these are to be handled with caution since the wavelet analysis suggests that traffic is not generated by a pure self-similar process.

In case of interval 8, we have seen that the results from the AV analysis do not contradict each other but the heuristic estimates are far from the intersection of the confidence interval. So again, here we might suspect non-stationarity or some other complicated structure of the data. This shows a weakness of the heuristic methods that they are not adequate for detecting any deviance from simple scaling. It also shows that one should never rely on the results of only one analysis. Several methods have to be performed to have a realistic picture of the data.



(a) Interval 1 (b) Interval 2

(c) Interval 3 (d) Interval 4

(e) Interval 5 (f) Interval 6

(g) Interval 7 (h) Interval 8

Figure 16: All estimated H values

6 Characterization of the burst structure

The purpose of this section is to give an alternative approach to traffic characterization which could complement existing methods.

The fact that computer network traffic is very bursty was confirmed by our study. The bursty nature of traffic was also pointed out and investigated in previous studies [2, 8, 12, 13, 14]. Burstiness means that sudden jumps occur frequently in the link utilization. These sudden jumps make it very hard to interpret results from correlation studies. Furthermore, it is very unlikely that our models for long-range dependent traffic such as fGn or fARIMA would produce such high jumps. On the other hand the dimensioning of computer networks should be based upon the traffic characteristics during the highly utilized periods, as well as the characteristics of the occurrence of highly utilized periods in the traffic. However, the Hurst parameter describing the long-range dependence in the traffic does not adequate for such a characterization. For this reason a characterization of the traffic was carried out concentrating on the bursts themselves.

An interval of traffic was called a *burst* or *busy period* if it was the largest possible interval where the bandwidth usage was constantly bigger than a given value δ viewed on a one second time-scale. For each value of δ busy and non-busy (silent) intervals can be separated.

In this section the behavior of these intervals will be examined for different values of δ.

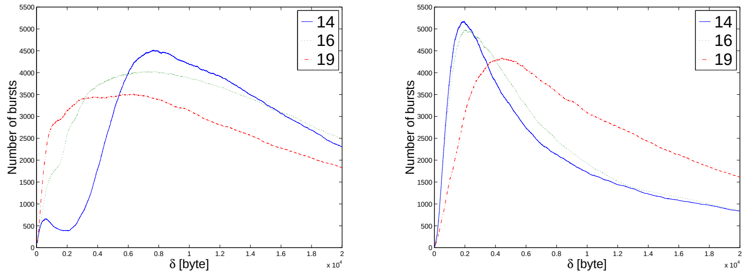
As in Section 5 we use the total-IP traces introduced in Section 3.1.

6.1 Number of bursts

The number of bursts as a function of δ was calculated and plotted in Figure 17, separately for the different directions. For each input data we define δ_{max} which gives the value of δ where the maximum number of bursts can be observed. δ_{max} is a natural separating level between busy and silent intervals since the more burst occur the more useful is to observe their behavior. It can be seen that δ_{max} differs from trace to trace that shows that there is no characteristic level which could be selected as the separator between busy and silent periods based upon this characterization.

6.2 Correlation structure of bursts

In this and the remaining subsections only one trace, the total-IP trace measured on the 16th of April, direction in, is investigated in more detail.



(a) Direction In

(b) Direction Out

Figure 17: Number of bursts as a function of threshold (δ)

Int.	1	2
Start	1771	2788
End	2871	3650
Len.	1101	863

Table 5: Endpoints of the selected stationary intervals from the burst size series. The values give the sequential numbers of the bursts.

Using $\delta = \delta_{max}$, the autocorrelation structure of the busy periods was calculated. Two time series were produced,

- the *duration* of the bursts, measured in seconds, and
- the *size* of the bursts, measured in bytes, giving the total amount of data transmitted during the burst.

The burst-size series are depicted in Figure 18. In this figure the series were averaged over neighboring windows of size 50 and shows that this sample is not stationary. For this reason shorter stationary intervals were selected with the methods described in Section 4.1 (with $N = 50$) and Section 4.2 ($M = 10$). The selected stationary intervals are described in Table 5. The intervals were selected using the burst-size series but the same intervals were used to investigate the burst-duration series.

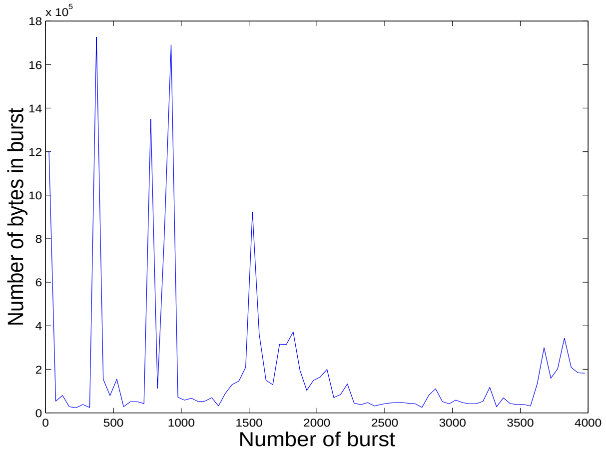


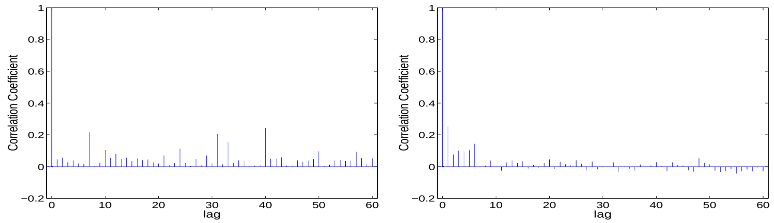
Figure 18: Burst sizes

Autocorrelation function

The autocorrelation function of the two intervals of the selected series are depicted in Figures 19 and 20. We can see that the two intervals have different correlation structure.

6.3 Correlation between busy and silent intervals

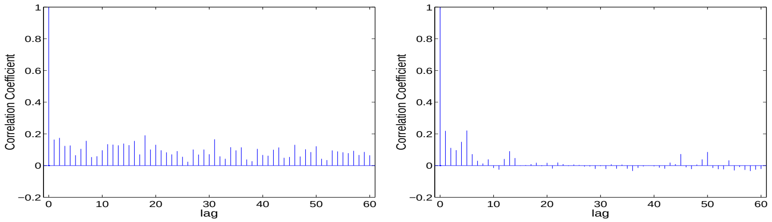
Besides calculating the self correlation structure of the busy periods the relation of busy and silent intervals is also relevant for traffic characterization purposes.



(a) Interval 1

(b) Interval 2

Figure 19: Autocorrelation function of the burst sizes



(a) Interval 1

(b) Interval 2

Figure 20: Autocorrelation function of the burst durations

For each value of δ the correlation between the duration of busy and the duration of subsequent silent interval was calculated and plotted as a function of δ in Figure 21(a). The correlation between a pair of random variables, X and Y is defined as

$$\rho = \frac{E(X - E(X)(Y - E(Y))}{\sigma_X \sigma_Y},$$

where σ_X and σ_Y are the standard deviations of the two variables. The value is estimated by calculating the time average. The correlation between busy and preceding silent intervals was also calculated. This is shown in Figure 21(b). Here the whole trace was used for this calculation.

These figures show that there is no significant correlation between these two series. Similar results have been found by analyzing other input data.

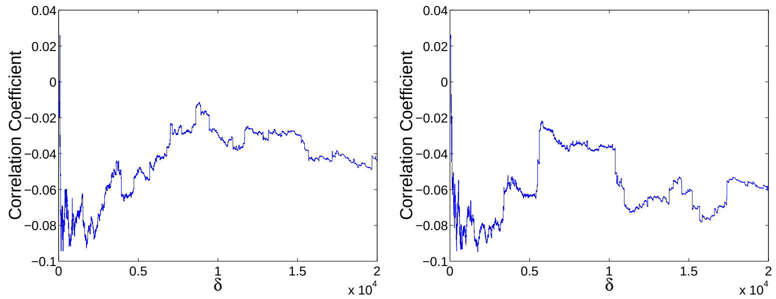
6.4 Tail distribution of burst sizes

The file sizes transmitted on the Internet are reported to have a so-called *heavy-tailed* distribution. Since the previously defined bursts are likely to be caused by transmission of large files the burst series was analyzed to check whether its distribution is heavy-tailed or not.

Let X be a non-negative random variable with cumulative distribution function F. F is said to be heavy-tailed with parameter α [20] if

$$1 - F(x) = x^{-\alpha} L(x), x \geq x_0, \tag{4}$$

where L is slowly varying at infinity, i.e. $\lim_{x\to\infty} L(tx)/L(x) = 1, t > 0$, and x_0 depends on $L(x)$.



(a) Cross-correlation between busy and **subsequent** silent intervals

(b) Cross-correlation between busy and **preceding** silent intervals

Figure 21: Cross-correlation between busy and silent intervals as a function of δ. Note the small scale on the y axis.

For example, in the simplest case $L(x) \equiv 1$, and the distribution with $F(x) = 1 - x^{-\alpha}, x \geq 1$ with $\alpha > 0$ is the so-called Pareto distribution.

To test whether the distribution is heavy-tailed or not the complementary distribution function $(1 - F(x))$ is plotted on a log-log scale and a straight line is fitted to the data. If the values fit the line closely we can conclude that the distribution is heavy-tailed and the parameter α can be read from the slope of the line.

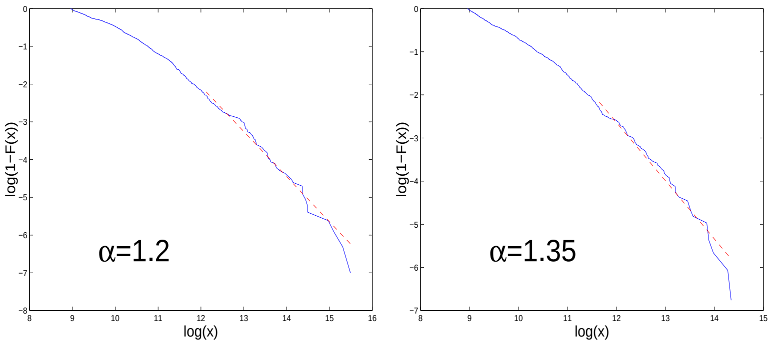
This test is depicted in Figure 22 for the burst size series for the two selected intervals. We can see that the values form a straight line which shows that the distribution is heavy-tailed. The value of the parameter α is different showing that the statistical properties of the two traces differ.

This heuristic parameter estimation method is just one of the existing tools. Other methods [17, 20] (QQ plots, Hill estimator, De Haan's moment method, etc.) are also recommended to use for reliable analysis results.

7 Conclusion

We have reported a burstiness and scaling analysis study based on real measured data.

Our statistics show that the most dominant application protocols in our University environment are the HTTP and the FTP protocols approximately with the same ratio of the total bandwidth. It is also not surprising that the most frequently



(a) Interval 1

(b) Interval 2

Figure 22: Distribution of the burst sizes

used transfer protocol is the TCP.

We have carried out a stationarity analysis on the measured traffic traces and identified stationary periods of the traffic.

Our correlation analysis shows that the traffic is bursty with positive short-term correlations and also exhibit long-range dependence. The burst structure was investigated in detail and a new burstiness characterization method was also presented.

Our results also demonstrate that the traffic structure is rather complex with non-stationary changes and the estimation of long-range dependence should not be based only a single method in order to avoid pitfalls of statistical analysis.

We have also found that the correlation structure of traffic changes during the time highlighting the importance of identifying stationary intervals of the traffic.

Acknowledgement

The authors would like to thank Zoltán Pusztai for his help during the measurements, Trang Dinh Dang for his help during the analysis, Darryl Veitch for the code of the wavelet based estimator and Attila Vidács for his comments on the paper.

References

[1] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Wavelets for the Analysis, Estimation and Synthesis of Scaling Data", in: **Self Similar Network Traffic Analysis and Performance Evaluation**. (1999), K. Park and W. Willinger, Eds.

[2] R. G. Addie, M. Zukerman, and T. D. Neame, "Broadband Traffic Modeling: Simple Solutions to Hard Problems", *IEEE Communication Magazine*, vol. 36,no. 8, 88–95, (August 1998).

[3] J. Beran, "Statistics for Long-Memory Processes", Chapman & Hall, (1994).

[4] P. J. Brockwell and R. A. Davis, "Time Series: Theory and Methods", Springer, (1996).

[5] P.J. Brockwell and R.A.Davis, "Introduction to Time Series and Forecasting", Springer, (1996).

[6] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data Networks as Cascades: Investigating the Multifractal Nature of Internet WAN Traffic", *ACM Computer Communication Review*, vol. 28, 42–55, (September 1998).

[7] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz, "The Changing Nature of Network Traffic: Scaling Phenomena", *ACM Computer Communication Review*, vol. 28, 5–29, (April 1998).

[8] D. L. Jagerman, B. Melamed, and W. Willinger, "Stochastic Modeling of Traffic Processes", in: **Frontiers in Queueing**, pages 271–370. CRC Press, (1997).

[9] M. Kendall and A. Stewart, "The Advanced Theory of Statistics", Charles Griffin, London, (1966).

[10] R. J. Larsen and M. L. Marx, "An Introduction to Mathematical Statistics and its Applications", Englewood Cliffs, N.J : Prentice-Hall, (1986).

[11] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (extended version)", *IEEE/ACM Transaction on Networking*, vol. 2,no. 1, 1–15, (February 1994).

[12] S. Molnár and I. Maricza eds., "Source Characterization in Broadband Networks", Interim report, COST 257, Vilamoura, Portugal, (January 1999).

[13] S. Molnár and Gy. Miklós, "On Burst and Correlation Structure of Teletraffic Models", *5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, West Yorkshire, UK., (July 1997).

[14] S. Molnár and Gy. Miklós, "Peakedness Characterization in Teletraffic", *IFIP International Conference on Performance of Information and Communication Systems*, Lund, Sweden, (May 1998).

[15] S. Molnár and A. Vidács, "On Modeling and Shaping Self-Similar ATM Traffic", *15th International Teletraffic Congress*, Washington, DC, USA, (June 1997).

[16] S. Molnár, A. Vidács, and A. A. Nilsson, "Bottlenecks on the Way Towards Fractal Characterization of Network Traffic: Estimation and Interpretation of the Hurst Parameter", *International Conference of the Performance and Management of Complex Communication Networks*, (November 1997).

[17] S. Molnár, A. Vidács, and D.D. Trang, "Heavy-Tailedness, Long-Range Dependence and Self-Similarity in Data Traffic", *7th International Conference on Telecommunication Systems Modeling and Analysis*, Nashville, TN, USA, (March 1999).

[18] T. Éltető and S. Molnár, "On the Distribution of Round-trip Delays in TCP/IP Networks", *The 24th Annual Conference on Local Computer Networks (LCN'99)*, Lowell/Boston, MA, (October 1999).

[19] D. T. Perkins, "Understanding SNMP MIBs", Technical report, Synoptics, (September 1993), Revision 1.1.7.

[20] S.I. Resnick, "Heavy Tail Modeling and Teletraffic Data", *The Annals of Statistics*, vol. 25, 1805–1869, (1997).

[21] M. Roughan and D. Veitch, "Measuring Long-Range Dependence under Changing Traffic Conditions, Extended Version", *Proceeding Infocom*, Manhattan, (April 1999).

[22] H. Saito, T. Tsuchia, Gy. Marosi, Gy. Horváth, P. Tatai, and S. Asano, "Real-time Simulation with Traffic Monitoring Tool", *ICCCN99*, Natick-Boston, MA, USA, (October 11-13, 1999).

[23] D. Veitch and P. Abry, "A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence", *Special issue on Multiscale Statistical Signal Analysis and its Applications, IEEE Trans. Info. Th.*, vol. 45,no. 3, 878–898, (April 1999).

[24] W. Willinger, M. Taqqu, and A. Erramilli, "A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High-Speed Networks Stochastic Networks: Theory and Applications", *Royal Statistical Society Lecture Notes Series*, vol. 4. Oxford University Press, (1996).