Self-Similar Traffic Modeling Techniques in ATM Networks

Master's Thesis

Attila Vidács

Technical University of Budapest
Faculty of Electrical Engineering
Dept. of Telecommunications and Telematics
High-Speed Networks Laboratory

May 23, 1996

Acknowledgments

A number of people have directly or indirectly contributed to this thesis. To all these people I express my deepest gratitude.

The official supervisor of this Master's Thesis has been Dr. Sándor Molnár. First of all, I wish to thank him for his encouraging support and guidance.

The present thesis summarizes the results of a research work started three years ago at the High-Speed Networks Laboratory at Technical University of Budapest. I also would like to express my appreciation to my tutor, Dr. Tamás Henk at HSNLab.

Also, I wish to thank Prof. Kauko Rahko for the opportunity to finalize this study at his Laboratory of Telecommunications Technology at Helsinki University of Technology. I also would like to thank Dr. Kalevi Kilkki for his encouraging help. Special thanks to all those people at HUT who helped me during my stay in Finland.

I would also like to acknowledge my unofficial advisor Dr. Ilkka Norros from VTT for useful discussions on many issues related to this thesis.

My colleagues and friends, Zsolt Kenesi and András Hegedűs deserve special thanks for all the help they have provided during the three years.

Last but not least, especial thanks are due all those people who helped me and inspired me during this work either in profession or in private life.

....

Espoo, May 22, 1996

Contents

Chapter 1

Introduction

The title of the thesis raises three questions at once. At first, “What is the meaning of the notion self-similarity?” Second, “What is the connection between self-similarity and traffic modeling in teletraffic theory?” And finally, “Why this topic is discussed in case of ATM networks?”. Next, I attempt to answer these questions to put the topic of the present work in its place .

The evolving Broadband Integrated Services Digital Network (B-ISDN) nowadays dominates the research of major telecommunication industries, and is designed in such a way that its evolution will rapidly form a worldwide high speed network. The CCITT has supported this activity in recent years through work performed by its standardization bodies, and has adopted the ATM concept as the vehicle for providing bearer services for all present and future teleservices [33]. Today, it is few who doubt that the coming generation of high capacity networks and switching systems, both in the public and private domain, will be based on the ATM (Asynchronous Transfer Mode) principle. However, ATM is new and unproven—also with respect to traffic engineering [15].

Understanding the nature of traffic in high-speed, high-bandwidth communications systems such as B-ISDN is essential for engineering, operations, and performance evaluation of these networks. In a first step toward this goal, it is important to know the traffic behaviour of some of the expected major contributors to future high-speed network traffic. But traffic characteristics in high speed packet based networks differ substantially from those in telephone networks. The differences can be attributed to the *complexity* inherent in packet traffic. The complexity of traffic in packet networks arises from the wide range of applications and services provided by these networks, as well as the inherent burstiness of traffic generated by these applications. The differences between traffic characteristics in telephone networks and packet networks are so great, that despite the well known robustness of teletraffic theory results, care should be taken applying traditional models and methods to packet networks [31]. Therefore, for the performance evaluation and sizing of packet based networks, the problem of obtaining an accurate and tractable characterization of packet traffic is of particular importance. As such, this topic has received considerable attention in the literature [29].

Careful statistical analysis of a large number of traffic traces from a wide variety of working packet-based networks consistently indicate traffic variations over a wide range of time scales. These variations can be parsimoniously described using the concept of self-similarity [9]. Intuitively, the attractive feature of self-similar processes is that they look the same at different time scales. One of the major implications of the self-similarity finding for traffic modeling is that it suggest an appealing alternative to the popular belief that since actual network traffic is extremely complex in nature, only complicated and highly parameterized models are likely to result in an accurate approximation of reality. By accepting the self-similarity hypothesis, the principle of parsimony can be applied and results in simple models that exhibit the same intricate features as actual network traffic.

1.1 Outline of the Thesis

The thesis is organized as follows:

- After this short introduction, Chapter 2 gives an overview about the history of high-speed network measurements and their results.

- Chapter 3 gives the full mathematical background of self-similar stochastic processes. The definitions for the notions of self-similarity and long-range dependence are given, and the second part of the chapter summarizes the properties of self-similar processes.

- In Chapter 4, different techniques are introduced to analyze the stochastic behaviour of self-similar processes and to estimate the degree of self-similarity. Four different graphical statistical test are discussed.

- In Chapter 5 different modeling approaches are summarized, based on very different views of data analysis. The purpose of these models are to generate self-similar synthetic traffic traces.

- Chapter 6 describes the traffic measurements made on the FUNET network. A brief description of the measurement unit used and the measuring configuration is given, together with the detailed description of the measured data sets.

- Chapter 7 describes the statistical analysis of the measured data, with emphasis on testing for self-similarity. Besides the self-similarity tests, the basic statistical analysis of the measured data is presented.

- Chapter 8 demonstrates the applicability of three different self-similar traffic models based on different approaches chosen from the models described in Chapter 5. After generating synthetic traffic traces with the models under discussion, the resulting processes are also analyzed and a comparison is made between the results obtained from real and synthetic traffic traces.

- Finally, Chapter 9 concludes this paper with a summary of main results.

Part I

Analysis and Modeling Techniques of Self-Similar Processes

Chapter 2

History

"640K ought to be enough for anybody."

Bill Gates (1955–), in 1981

When we speak about history in Computer Science we do not have to travel back in time so far. But care should be taken when we evaluate the results of the recent past. The evolution of that scientific and technological area is extremely fast—even the most recent results and products could turn into 'history' within a few months. As it is said in a free style, the goal of Computer Science is to build something that will last at least until we have finished building it. As a consequence, every result and statement should be considered together with its date and the place when and where it was achieved.

The present section tries to give an overview about the history of high-speed network measurements and their results. But instead of listing all the measurements done and presented in the literature, I shall focus on the trends and emerging challenges an up-to-date measurement should cope with.

2.1 High-Speed Network Measurements

Over the last twenty years, the collection of traffic measurements from "live" packet networks has underground a major change. Around 1980, an Ethernet LAN running at 10 Mbps represented the state-of-the-art high-speed network, traffic recording equipment that was able to record every complete Ethernet packet at that speed was non-existent, and the real time storage requirements of such a hypothetical traffic recorder were prohibitive. At present, packet networks running at 150 Mbps, or even at Gigabit-speeds, are becoming more and more common, and technological advances in the areas of data recording and data storage have lead to the development of traffic monitoring equipment that is capable of recording every message/frame/cell at those speeds for multiple hours, and of writing the entire data onto tapes.

With the availability of high-precision recording equipment, traffic data collection from packet networks results in data sets that are unique in terms of volume and quality. For example, an hour worth of ATM traffic (on an OC3 link, lightly loaded) is easily seen to give rise to data sets extending beyond the 100 million observations range. Exploiting the resulting volume of data poses a significant challenge for data analysts and traffic modelers, and the complexity of the measured data only adds to these difficulties.

Finally, for the data analyst as well as for the traffic modeler, it is worthwhile to keep in mind that networks are highly dynamic entities and undergo constant changes (e.g., network topology, user population, services and applications, protocols). This makes life for the data analyst and traffic modeler even more difficult; the task is comparable to trying to hit a moving target! That is, the traffic analyst is asked to detect some unknown characteristic features in enormous-sized empirical records, where the records are constantly augmented by new traffic measurements from new networks carrying new services and applications [9].

2.2 Complexity of Packet Traffic

In the recent past, the area of traffic modeling for fast packet networks has benefited from a "flood" of high-resolution, high-quality and high-volume traffic measurements from a wide range of "live" packet networks. These measurements provide overwhelming evidence that, in contrast to commonly made assumptions, actual traffic has characteristics that span many time-scales. Examples of these characteristics are long-range dependence (or correlations that decay slowly) and heavy-tailed densities (e.g., for inter-arrival times, holding times) [9]. (The corresponding mathematical theory is presented in Section 3.)

It is notable that these properties are observed in traffic from a wide range of packet applications from low speed ISDN to Variable Bit Rate video; from single sources to multiplexing of hundreds of sources; intranetwork as well as internetwork traffic; for all times of day. These measurement studies reveal that the critical characteristic of the self-similar traffic is that there is no natural length of a burst: at every time scale ranging from a few milliseconds to minutes and hours, similar-looking traffic bursts are evident [29]. Closer examination of data shows that an apparent busy period resolves into smaller busy periods separated by idle intervals. In contrast, traditional models assume the presence of one dominant time scale, and these assumptions lead to Markovian models of packet interarrival times and resource holding times. On the other hand, the "burst within burst" (or self-similar) structure observed in packet traffic, first studied by Mandelbrot [22], is characteristic of fractals. (Appendix A gives the intuitive notion of self-similarity by presenting point sets which are *fractals*. Those pictures are very attractive but a 'snowflake-curve' is a little bit far from a sentence of cells in a computer network. That is why I shall introduce self-similarity through definitions and theorems instead of drawing nice pictures which look pretty but are useless in exact mathematical discussions.) Conventional models do not capture this self-similar structure. This is not to say that traditional models cannot be extended in principle to capture these properties over a wide range of time scales, but this adds to the model complexity, which is undesirable from the viewpoint of tractability and in that potentially large number of inputs may be required by the model. For all the complexity of the "burst within burst" structure, there are often simple rules that underlie the different time scales of this structure. It is conceivable then to quantify the burst structure by a small set of parameters [29].

2.3 The Bellcore Measurements

One particular high-speed network measurement and analysis study referred to as the 'Bellcore measurements' played a main role in the history of self-similar traffic modeling. The results from that study form the basis of most of the publications in this subject until now, so it is worth to give a short summary here about those path-finder results.

Around 1990, high resolution traffic measurement studies at the Bellcore Morris Research Engineering Center have demonstrated that, over time scales of engineering interest, packet traffic exhibits many of the defining properties of fractal processes. The studies based on the analysis of LAN traffic data collected by Leland and Wilson [20] who were able to record hundreds of millions of Ethernet packets without loss and with recorded time-stamp accurate to within 100 μs. The data were collected between August 1989 and February 1992 on several Ethernet LAN's at Bellcore. The main findings of their statistical analysis are briefly:

- Ethernet LAN traffic is statistically self-similar, irrespective of when the data were collected and where they were collected in the network.
- The degree of self-similarity measured in terms of the Hurst-parameter H is typically a function of the overall utilization of the Ethernet and can be used for measuring the "burstiness" of the traffic.
- Major components of Ethernet LAN traffic such as external LAN traffic or external TCP traffic share the same self-similar characteristics as the overall LAN traffic.
- The packet traffic models currently considered in the literature are not able to capture the self-similarity property and can therefore be clearly distinguished from the measured data.

The indications of the impact of the self-similar nature of packet traffic for engineering, operations, and performance evaluation of high-speed networks are already ample [20]:

- Source models for Ethernet users are expected to show extreme variability in terms of interarrival times of packets (i.e., the infinite variance syndrome).
- Commonly used measures for "burstiness" such as the index of dispersion (for counts), the peak-to-mean ratio, or the coefficient of variation (for interarrival times) are no longer meaningful for self-similar traffic but can be replaced by the Hurst-parameter.
- The nature of congestion produced by self-similar network traffic models differs drastically from that predicted by standard formal models and displays a far more complicated picture than has been typically assumed in the past.
- First analytic results show a clear distinction between predicted performance of certain queueing models with traditional input streams and the same queueing models with self-similar inputs.

Following these studies, numerous researchers worldwide have concluded and reported measurement studies during the last five years that essentially confirm the presence of fractal features in measured network traffic.

Chapter 3

Mathematical Background

"There is no branch of mathematics, however abstract, which may not be applied to phenomena of the real word."

Nikolai Lobatchevsky

The present section gives the mathematical background of self-similar stochastic processes. First, the mathematical definitions for the notions of self-similarity and long-range dependence are given. Next, the second part of the chapter summarizes the properties of self-similar processes and gives the definition of an empirical law that is referred to as 'Hurst's law' or the 'Hurst effect' in the literature.

At first, for the exact mathematical discussion we need to create the appropriate mathematical model for real packet traffic processes. In case of packetized data traffic this abstraction is quite simple. Let $X = (X_1, X_2, \ldots)$ be an integer valued stochastic process in discrete time where X_i is the number of packets arriving at the ith time interval. (For example, in case of a traffic measurement on an Ethernet LAN, X_1 is equal to the number of frames captured in the first 10 milliseconds of the measurement time, X_2 is the number of frames arriving in the time interval 10-20 milliseconds, and so on.) In case of variable length packets, X_i can represent the number of bytes received in the given interval. If we choose the time interval equal to the packet transmission time, we get the process X', where the value of X'_i can be 0 or 1, depending on whether a packet is received in time slot i or not.

Sometimes it is useful to introduce the process $Y = \{Y_i : i \geq 1\}$, where Y_i is equal to the elapsed time between the arrival instants of packets ith and

$(i-1)$th. The distribution of these inter-arrival times can be the peculiarity of the measured traffic process.

To make the list complete, one can define the cumulative arrival process $A = \{A_t : t > 0\}$, where A_t is the total number of packets arrived in time interval $(0, t]$. Figure (3.1) shows all these processes in case of a simple traffic pattern.

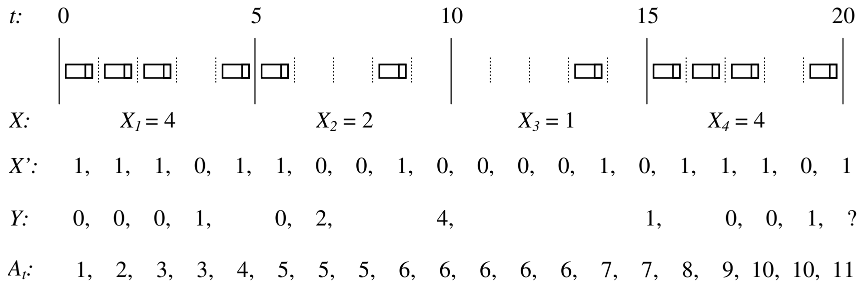


Figure 3.1: Mathematical representations of packet traffic flows

With the help of the above defined stochastic processes all features of packet traffic can be described in exact mathematical language.

3.1 Definitions

In this section the mathematical definitions for long-range dependence and self-similarity are presented.

In order to develop definitions, consider a wide-sense stationary (sometimes called covariance stationary) process $X = (X_1, X_2, X_3, \ldots)$ with constant mean, finite variance and autocorrelation function

$$r(k) = \frac{E\Big\{\big(X_n - E\{X\}\big)\big(X_{n+k} - E\{X\}\big)\Big\}}{E\Big\{\big(X_n - E\{X\}\big)^2\Big\}},$$

that depends only on k, and power spectrum

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r(k) e^{-ik\omega}$$

Further, let $X^{(m)}$ denote a new time series obtained by averaging X over non-overlapping blocks of size m, and replacing each block by its arithmetic mean:

$$X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + \cdots + X_{km}), \qquad k = 1, 2, 3 \ldots \tag{3.1}$$

Thus $X^{(m)}$ corresponds to viewing the process X using a time scale that is a factor of m coarser than that used to view X itself [27]. Note, that for each m, the aggregated time series $X^{(m)}$ defines a covariance stationary process.

A stationary time series is called *long-range dependent*, if the correlation between neighbouring exclusive blocks does not asymptotically vanish when the block size increased. By definition,

Definition 3.1 *The process X is said to exhibit* long-range dependence *of index H if it has an autocorrelation function of the form*

$$r(k) \sim A(k)k^{-(2-2H)}, \qquad (k \to \infty) \tag{3.2}$$

where $A(k)$ is a slowly varying function at infinity, and H, referred to as the Hurst parameter, is in the range $(0.5, 1)$ [29].

Thus, processes with long-range dependence are characterized by an autocorrelation function that decays hyperbolically as the lag increases [20]. Note that this implies that the sum over all lags of the autocorrelation function is divergent.

A process is called *self-similar*, if it behaves, up to a scaling factor, in exactly the same way at all time scales. The definition is:

Definition 3.2 *Consider the process X, and the aggregated process $X^{(m)}$. Let $r^{(m)}(k)$ be the autocorrelation function of the aggregated process. Then X is said to be*

- exactly self-similar *if $X \equiv m^{-H}X^{(m)}$, i.e., if X is identical to $X^{(m)}$ to within a scale factor in a finite dimensional distribution sense.*
- exactly second-order self-similar *if $r^{(m)}(k) = r(k), \quad (k \geq 0)$.*
- asymptotically second-order self-similar *if $r^{(m)}(k) = r(k), \ (k, m \to \infty)$*

where the Hurst parameter H is also referred to as the self-similarity parameter [29].

A self-similar process is long-range dependent if it has strictly positive correlations [3].

3.2 Properties

In this section I summarize the properties of self-similar processes. Many of these features are equivalent and can be derived straight from the definitions stated above. Trying to keep the exact mathematical description I focus on traffic modeling based aspect—a comparison is made between self-similar processes and processes generated by short-range dependent "traditional" traffic models.

The following is only a brief summary based on [29] and [3].

3.2.1 Long-range dependence

A self-similar process with self-similarity parameter $H > 0.5$ is long-range dependent. Thus, processes with long-range dependence are characterized by an autocorrelation function that decays hyperbolically as the lag increases. Moreover, Equation (3.2) implies $\sum_k r(k) = \infty$. This non-summability of the correlations captures the intuition behind long-range dependence, namely that while high-lag correlations are all individually small, their cumulative effect is of importance and gives rise to features which are drastically different from those of the more conventional, i.e., short-range dependent processes. The latter are characterized by an exponential decay of the autocorrelations, i.e., $r(k) \sim a^k$, $(0 < a < 1)$ as $k \to \infty$, resulting in a summable autocorrelation function $0 < \sum_k r(k) < \infty$ [20].

Practical aspects of long-range dependence in traffic modeling are [3]:

- Long-rage dependence implies that traffic behaves worse than would be expected on the basis of short-range dependent models. The fluctuations are larger, and congestion builds up more easily.

- The estimation of traffic parameters is an entirely different problem compared with traditional modeling. For example, such an ordinary notion as the mean rate of traffic has to be interpreted in a different way. (I shall discuss this problem in more details in Section 7.2.5 on page 49.)

3.2.2 1/f-noise

Long-range dependence manifests itself in the frequency domain as *1/f-noise*, which is the term used to refer to a sharp divergence in the power spectrum near the origin [29]. The exact definition for 1/f-noise is the following:

Definition 3.3 *The process X is said to exhibit* 1/f-noise *if the power spectrum $S(\cdot)$ is of the form*

$$S(\omega) \sim B(1/\omega)\omega^{-(2H-1)}, \qquad (\omega \to 0) \tag{3.3}$$

where $B(\cdot)$ is slowly varying at infinity, and Hurst parameter H is in the range $(0.5, 1)$ [29].

This is in contrast to widely used theoretical models where the power spectra is flat at low frequencies.

3.2.3 Non-degenerate correlation structure

Intuitively, the most striking feature of self-similar processes is that their aggregated processes $X^{(m)}$ possess a non-degenerate correlation structure as $m \to \infty$ [20]. For typical stochastic processes, as m increases the autocorrelation of $X^{(m)}$ *decreases* until in the limit the elements of $X^{(m)}$ are uncorrelated. (In other words, their aggregated processes $X^{(m)}$ tend to second order pure noise as $m \to \infty$. For a self-similar process, on the other hand, the process X have the *same* autocorrelation function.[27]

3.2.4 Slowly decaying variances

From statistical point of view, the most salient feature of self-similar processes is that the variance of the arithmetic mean decreases more slowly than the reciprocal of the sample size [20]:

$$var\{X^{(m)}\} \sim am^{-(2-2H)}, \quad as \quad m \to \infty, \tag{3.4}$$

where a is a finite positive constant independent of m, and $0 < H < 1$. On the other hand, for covariance stationary processes whose aggregated series $X^{(m)}$ tend to second-order pure noise, the sequence $(var\{X^{(m)}\} : m \geq 1)$ satisfies $var\{X^{(m)}\} \sim bm^{-1}$, as $m \to \infty$, where b is a finite positive constant independent of m [20].

Cox (1984) showed, that a specification of the autocorrelation function satisfying (3.2) (or equivalently, of the spectral density function satisfying (3.3)) is the same as the specification of the sequence $(var\{X^{(m)}\} : m \geq 1)$ with the property (3.4) [20].

3.2.5 Hurst effect

Historically, the importance of self-similar processes lies in the fact that they provide an elegant explanation and interpretation of an empirical law that is commonly referred to as *Hurst's law* or the *Hurst effect*. Briefly, for a given set of observations $(X_k : k = 1, 2, \dots, n)$ with sample mean $\overline{X}(n)$ and sample variance $S^2(n)$, the *rescaled-adjusted range* or the *R/S statistic* is given by

$$\begin{aligned} R(n)/S(n) &= \frac{1}{S(n)}\Big(\max(0, W_1, W_2, \dots, W_n) - \\ &\quad - \min(0, W_1, W_2, \dots, W_n)\Big), \end{aligned} \tag{3.5}$$

with

$$W_k = (X_1 + X_2 + \cdots + X_k) - k\overline{X}(n), \quad k = 1, 2, \dots, n) \tag{3.6}$$

Hurst (1951) found that many naturally occurring time series appear to be well represented by the relation

$$E\{R(n)/S(n)\} \sim cn^H, \quad as \quad n \to \infty, \tag{3.7}$$

with *Hurst parameter* H typically about 0.73, and c a finite positive constant that does not depend on n. On the other hand, if the observations X_k come from a short-range dependent model, then Mandelbrot and Van Ness (1968) showed that

$$E\{R(n)/S(n)\} \sim dn^{0.5}, \quad as \quad n \to \infty, \tag{3.8}$$

(d a finite positive constant, independent of n). The discrepancy between (3.7) and (3.8) is generally referred to as the *Hurst effect* or *Hurst phenomenon* [20].

Chapter 4

Analysis Tools

"There is not just one analysis for a substantial problem."

F. Mosteller [24]

This chapter tries to give an overview about the statistical methods used testing for and estimating the degree of self-similarity. Four different graphical statistical tests are discussed, the so-called IDC-plot, the variance-time plot, the R/S statistic, and an analysis method based on periodograms.

4.1 Testing for Self-Similarity

Since we are always dealing with finite data sets, it is in principle not possible to decide whether the asymptotic relationships in Definition (3.1) and (3.2) hold or not. For processes that are not self-similar in the sense that their aggregated series converge to second-order pure noise, the correlations will eventually decrease exponentially, continuity of the spectral density function at the origin will eventually show up, the variances of the aggregated processes will eventually decrease as m^{-1}, and the rescaled adjusted range will eventually increase as $n^{0.5}$. For finite sample sizes, distinguishing between these asymptotics and the ones corresponding to self-similar processes is, in general, problematic [20].

While we cannot check whether by definition a traffic trace is self-similar or not, we can check for different features of self-similarity present in actual packet traffic. From a theoretical point of view, slowly decaying variances, long-range dependence, a spectral density of the form (3.3), and fractal dimensions are different manifestations of one and the same property of the underlying covariance stationary process X, namely that X is (asymptotically or exactly) self-similar. Subsequently, we can approach the problem of testing for and estimating the degree of self-similarity from at least four different angles: (1) analysis of the variances of the aggregated processes $X^{(m)}$, (2) time-domain analysis based on the R/S statistic, (3) periodogram-based analysis in the frequency domain, and (4) calculating the dimension of the given data set. This section provides a brief description of the corresponding statistical and graphical methods (for more details, see [1]) [20].

4.1.1 IDC—Index of Dispersion for Counts

A commonly used measure for capturing the variability of traffic over different time scales is provided by the *index of dispersion (for counts)*. For a given time interval of length L, the index of dispersion for counts (IDC) is given by the variance of the number of arrivals during the interval of length L divided by the expected value of that same quantity:

$$IDC(L) = var\{\sum_{j=1}^{L} X_j\}/E\{\sum_{j=1}^{L} X_j\} \tag{4.1}$$

In case of conventional traffic models such as Poisson or Poisson-like processes and the popular Markov-modulated Poisson processes (MMPPs), the IDC is either constant or converges to a fixed value quite rapidly. On the other hand, self-similar traffic models are easily shown to produce a monotonically increasing IDC. Note that when plotting log(IDC(L)) against log(L), this property results in an asymptotic straight line with slope $2H - 1$ [20].

Compared to the closely related variance-time plot method, plotting the IDC curve and estimating its slope provides a quick and simple engineering-based approach to testing for self-similarity of a traffic measurement set [20].

4.1.2 Variance-time plot

A *variance-time plot* should show that, if the sample is aggregated by a factor of m (corresponding to Equation (3.1)), then, asymptotically, the variance of the aggregated versions falls off by a factor of $m^{-\beta}$, where $\beta = 2(1 - H)$. This means that the variances of the aggregated processes decrease linearly (for large m) in log-log plots against m with the slopes arbitrarily flatter than -1 (see Equation (3.4)). This is a heuristic test in the sense that the statistical properties of these plots are not known, but it is valuable because of the accompanying physical intuition: it indicates how "bursty" the sample is when viewed over progressively larger time scales [27][20].

The so-called variance-time plots are obtained by plotting $\log(var(X^{(m)}))$ against $\log(m)$ and by fitting a simple least squares line through the resulting points in the plane, ignoring the small values of m. Values of the estimate $\hat{\beta}$ of the asymptotic slope between -1 and 0 suggest self-similarity, and the estimate for the degree of self-similarity is given by $\hat{H} = 1 - \hat{\beta}/2$ [20].

Clearly, variance-time plots are not reliable for empirical records with small sample sizes. But in case of huge data sets, the variance-time plots become highly useful and give a rather accurate picture about the self-similar nature of the underlying time series and about the degree of self-similarity [20].

4.1.3 R/S-analysis

The feature that makes R/S statistics particularly attractive is its relative robustness against changes in the marginal distribution, even for long-tailed distributions. On the other hand, for marginal distributions which are close to normality a dramatic loss in efficiency is reported and no detailed analysis of robustness of R/S statistics has taken place yet.

Given an empirical time series of length N ($X_k : k = 1, \ldots, N$), the whole series is subdivided into K non-overlapping blocks. Now, we compute the rescaled adjusted range $R(t_i, d)/S(t_i, d)$ for a number of values d, where $t_i = \lfloor N/K \rfloor (i-1) + 1$ are the starting points of the blocks which satisfy $(t_i - 1) + d \leq N$.

$$
\begin{aligned}
R(t_i, d) &= \max\{0, W(t_i, 1), \ldots, W(t_i, d)\} - \\
&\quad - \min\{0, W(t_i, 1), \ldots, W(t_i, d)\},
\end{aligned}
$$

where

$$
W(t_i, k) = \sum_{j=1}^{k} X_{t_i+j-1} - k \cdot \Big(\frac{1}{d} \sum_{j=1}^{d} X_{t_i+j-1}\Big), \quad k = 1, \ldots, d.
$$

$S^2(t_i, d)$ denotes the sample variance of $X_{t_i}, \ldots, X_{t_i+d-1}$. For each value of d one obtains a number of R/S samples, which decreases from K for larger values of d. One computes these samples for logarithmically spaced values of d, i.e., $d_{l+1} = m \cdot d_l$ with $m > 1$, starting with d_0 of about 10. Plotting $\log R(t_i, d)/S(t_i, d)$ vs. $\log d$ results in the R/S plot, also known as *pox diagram*.

Next, a least squares line is fitted to the points of the R/S plot, where both the R/S samples of the smallest values of d are not considered (because they are dominated by short-range correlations) and those R/S samples of large values of d where the number of samples are less than say 5 (because they are statistically insignificant). The slope of the regression line for these R/S samples is an estimate for the Hurst parameter H. Both the number of blocks K and the number of values d should not be chosen to small. In addition, some care has to be taken when deciding about the end of the transient, i.e., which of the small values of d should not be taken into consideration for the regression line. In practice, it should be checked whether different parameter settings lead to consistent H estimates for $X^{(m)}$ with different aggregation levels m.[3]

4.1.4 Periodogram-based analysis

While variance-time plots and R/S-statistics are very useful tools for identifying self-similarity, the results for the limit laws of the corresponding statistics are absent. When a more refined data analysis is required (e.g., confidence intervals for the degree of self-similarity parameter H), then periodogram-based estimators are used.

The main idea of these methods is to assume a certain self-similar process type, and to fit the parameters of this process to the given empirical sample. The fitting should be optimal in the sense that the periodogram of the sample and the spectral density of the process are minimizing a given goodness-of-fit function. (Note, that because the basic assumption mentioned above, this method is rather an *estimation procedure* for the self-similarity parameter H and *not a test* for whether a sample of a time series is consistent with long-range dependence [27]!)

As mentioned in Section 3.2 on page 16, the spectral density of self-similar processes obeys a power law near the origin. Thus, the first idea to determine the Hurst parameter H is simply to plot the periodogram in a log-log grid, and to compute the slope of a regression line which is fitted to a number of low frequencies. This should be an estimate of $1 - 2H$. In most of the cases this will lead to a wrong estimate of H since the periodogram is *not appropriate* to estimate the spectral density. Thus, more sophisticated methods have to be applied to obtain useful estimates of H [3].

Several periodogram-based estimators can be found in the literature. In the following, an MLE (Maximum Likelihood Estimator) is presented briefly according to [3]. The method originally presented in [2] is based on Whittle's approximate MLE for Gaussian processes [37].

The spectral density of the self-similar process is denoted by $f(\lambda; \theta)$, where the parameter vector $\theta = (\theta_1, \ldots, \theta_M)$ of the process is structured as follows. θ_2 denotes the Hurst parameter H. $\theta_1 = \sigma_\varepsilon^2$ is a scale parameter, where σ_ε^2 is the variance of the innovation ε of the infinite AR-representation of the process, i.e., $X_j = \sum_{i=1}^{\infty} \alpha_i X_{j-i} + \varepsilon_j$. This implies $\int_{-\pi}^{\pi} \log(f(\lambda; (1, \theta_2, \ldots, \theta_M))) d\lambda = 0$. If necessary, the parameters θ_3 to θ_M describe the short-range behaviour of the process. For FGN [1] , only σ_ε^2 and H have to be considered. With $\eta = (\theta_2, \ldots, \theta_M)$, the Whittle estimator $\hat{\eta}$ of η minimizes the quality-of-fit function

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; (1, \eta))} d\lambda$$

where $I(\cdot)$ denotes the sample periodogram (i.e., power spectrum as estimated using a Fourier transform) defined by

$$I(\lambda) = \frac{1}{2\pi N} |\sum_{j=1}^{N} X_j e^{ij\lambda}|^2.$$

$\hat{H}$ is given by $\hat{\theta}_2$ and the estimate of σ_ε^2 by

$$\hat{\sigma}_\varepsilon^2 = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; (1, \hat{\eta}))} d\lambda.$$

The approximate 95%-confidence interval of $\hat{H}$ is given by

$$\hat{H} \pm 1.96 \sqrt{\frac{V_{11}}{N}}$$

where $V = 2D^{-1}$ and the matrix D is defined by

$$D_{ij} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(\lambda) \frac{\partial}{\partial \theta_j} \log f(\lambda) d\lambda.$$

In practice, there are two problems which may have an effect on the robustness of the estimator [32]:

- Deviations from the assumed model spectrum: i.e., deviations at higher frequencies lead to a bias in the estimate of H. One possible solution is to estimate H only from periodogram ordinates at low frequencies. For large data sets, one can also aggregate the data over non-overlapping blocks of length m and compute several $\hat{H}^{(m)}$ for the $X^{(m)}$ (see [20]).
- Deviation from Gaussianity: Gaussianity can often be achieved by transforming the data but then it has to be proven that the estimate of H for the original and the transformed data sets are identical.

[1] Fractional Gaussian Noise (FGN) is an exactly self-similar stochastic process.

4.1.5 Fractal dimensions

Erramilli et al. [7] propose using *fractal dimensions* (see Mandelbrot [22]) in order to characterize the fractal-like nature of packet traffic.

Intuitively, a dimension is an indication of the extent to which a set or an object (e.g., arrival times) fills the space in which it is embedded. For example, the widely used *correlation dimension* has been shown to be computationally feasible and it is associated with a measure known as the *correlation integral* $C(t)$, defined as the number of pairs of points i, j whose distance is less than t, i.e.,

$$C(t) = \sum_i \sum_j \frac{C_{ij}}{N^2},$$

where C_{ij} is 1 or 0 depending on whether the points indexed i, j in the set are separated by a distance less than t, or greater than t. (In case of traffic measurements, N is the number of registered packets in a given time interval and C_{ij} can be interpreted as is equal to 1 if the time between packet arrivals i. and j. is smaller than t, otherwise 0. In this case $C(t)$ is the probability that a pair of arrivals occur within a time interval t.) The correlation dimension is then defined as

$$D_c = -\lim_{t \to 0} \frac{\log C(t)}{\log t}$$

if this limit exists. For self-similar sets, the correlation integral scales as $C(t) \sim t^{D_c}$ over a range of time scales, and D_c is found by estimating the slope of a plot of $\log C(t)$ against $\log t$ [8].

The estimation of the correlation dimension is an active research topic of statistics. The optimal method for a set and the determination of the confidence interval are not known in this moment. The estimation error of the dimension is commonly taken to be the error estimate of the linear regression procedure—but it is likely optimistic.

(The statistical method presented above is not used during the analysis I made. The reason why the algorithm was mentioned here is to make the list of statistical methods testing for self-similarity more complete.)

Chapter 5

Self-Similar Traffic Models

"Virtually no model is ever correct—
all are but better or worse approximations for reality."

F. R. Hampel [12]

"Which one is the best traffic model to describe real packet traffic in ATM networks?" The question is simple but the answer is not so! Simply speaking, the task for traffic modeling is to find the "appropriate" traffic models to describe actual data traffic. But the word "appropriate" here is very context sensitive. We can only say that a 'good' model should possess the following features [33]:

1. *Proximity to real sources*
   A model to be selected should be fitting to actual sources. The importance of this criterion is more than obvious. (Note, that from a statistical point of view, a number of different models can be consistent with a given set of data.)

2. *Generality*
   In some cases, the model adopted should be general enough. It means that the model should fit to a variety of different services by appropriately varying the values of the parameters defining it. If a model is general enough, then the analytical tools applied to the analysis would be unified accordingly.

3. *Simplicity*
   A model should be described by means of a small number of parameters. It would also be preferable that the defining parameters make good intuitive sense, by being closely related to actual physical quantities.

4. *Ease of fitting to real sources*
   The model can fit to an actual source by selecting appropriate values for the parameters defining it. This is usually done by expressing the moments of certain random variables related to the model in terms of its parameters, and then by setting the parameter values so that the resulting moment-values match the corresponding experimental results. It is desirable that the moments used in this process be easily measurable experimentally. Of course, the parameter values are derived by solving an appropriate system of equations, which should be simple and numerically stable.

5. *Analytical tractability and accuracy*
   Almost always, exact analysis is next to impossible, since it requires a tremendous amount of computation. Thus, the model should give rise to tractable as well as accurate approximate analysis.

6. *Ease of implementation*
   The model should be easily implementable in experiments, by means of both computer simulations and hardware-based traffic generators. In fact, the complexity of the algorithm describing cell generation can be used as a measure of ease in implementation.

7. *Appropriateness for modeling aggregate traffic and/or output traffic*
   It is often of interest that a model fits the aggregate traffic stream formed by a multitude of sources. This feature would render the model appropriate for the analysis of statistical multiplexers, as well as for the output of other links. In fact, a model would be particularly suitable for network-wide analysis if it were to fit both aggregate and output traffic streams.

The discussion above implies that the 'ideal traffic model' does not exist, at least for the time being. Hence, instead of selecting one of the models to be presented as the 'best', different modeling approaches are summarized in this section based on very different views of data analysis.

One can choose between two approaches in order to deal with self-similar properties of packet traffic. The two approaches are (1) stochastic models related to self-similar processes and (2) deterministic models using nonlinear chaotic maps. Note that the motivation for both approaches is the desire for a relatively simple description of the complex packet traffic generation process. Both modeling approaches yield a single parameter that describes the fractal nature of traffic (i.e., the Hurst parameter for self-similar models, and the fractal dimension for chaotic map models) and seems to capture the intuitive notion of "burstiness" [20].

Next, four different modeling techniques are examined in more details. The first one is based on exactly self-similar stochastic process called fractional Brownian motion. The second and third modeling approaches use asymptotically self-similar stochastic processes, and the final method is based on deterministic chaotic maps. Finally, two alternative modeling approaches are presented and neither of them based on self-similar processes. The first one is based on short-range dependent traffic models, and the second one is attacking the assumption of stationarity in case of actual traffic traces.

5.1 Fractional Brownian Traffic

The model based on *fractional Brownian motion* (FBM) is an exactly self-similar model that appears to be a reasonable presentation of aggregate packet network traffic.

We call the process

$$A_t = mt + \sqrt{am}Z_t, \qquad t \in (-\infty, \infty) \tag{5.1}$$

fractional Brownian traffic, where Z_t is a normalized fractional Brownian motion (FBM), i.e., a stochastic process $Z_t,\ t \in (-\infty, \infty)$, characterized by the following properties [25]:

1. Z_t has stationary increments;
2. $Z_0 = 0$, and $E\{Z_t\} = 0$ for all t;
3. $E\{Z_t^2\} = |t|^{2H}$ for all t, where $H \in [1/2, 1)$;
4. Z_t has continuous paths;
5. Z_t is Gaussian, i.e., all its finite-dimensional marginal distributions are Gaussian.

Z_t is strictly self-similar process with Hurst parameter H. In the special case $H = 1/2$, Z_t is the standard Brownian motion and the traffic model (5.1) reduces to a usual diffusion model [25].

The process has three parameters m, a and H with the following interpretations: $m > 0$ is the mean input rate, $a > 0$ is a variance coefficient, and $H \in (1/2, 1)$ is the Hurst parameter of Z_t. In the special case $H = 1/2$ we call A_t *Brownian traffic*.

The factor $\sqrt{m}$ in (5.1) is motivated by the following superposition property: The sum $A_t = \sum_{i=1}^{K} A_t^{(i)}$ of K independent fractional Brownian traffics with common parameters a and H but individual mean rates m_i can be written as $A_t = mt + \sqrt{ma}Z_t$, where $m = \sum_{i=1}^{k} m_i$ and Z_t is a FBM with parameter H. This shows that the roles of the three parameters of the traffic model can be separated so that H and a characterize the "quality" of the traffic in contrast to the long run mean rate m which characterizes its "quantity" alone [26].

Besides self-similarity, the other important feature of fractional Brownian traffic is its Gaussian character. The Gaussian distribution appears when a large number of independent random variables are summed up together. The output of a single source is usually very far from a Gaussian process. Fractional Gaussian traffic is a rather generally applicable model for traffic aggregated from a large number of independent sources with peak rates are sufficiently low in comparison to the aggregated mean rate [25].

An obvious weakness of (5.1) as a traffic model is that the amount of traffic arriving within an interval can be negative. In case of a traffic process the notion 'negative traffic' is meaningless (or at least in most cases). To solve this problem, a simple conversion method is presented in Section 8.1.2.

5.2 Fractional ARIMA processes

The class of *Fractional Autoregressive Integrated Moving-Average* (FARIMA) processes was introduced by Granger and Joyeux (1980) in [11] and Hosking (1984) [16] as an extension to the traditional ARMA and ARIMA classes, which are only capable to model the short-range portion of the correlations of empirical data sets [3].

A fractional ARIMA(p, d, q) process, where p and q are non-negative integers and d is real, is defined to be a stochastic process $X = (X_k : k \geq 0)$ with a representation given by

$$\Phi(B)\Delta^d X_k = \Theta(B)\varepsilon_k,$$

where $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in the backward-shift operator $B\{X_k\} = X_{k-1}$, $\Delta = 1 - B$ denotes the differencing operator, and Δ^d is the fractional differencing operator defined by

$$\Delta^d = (1-B)^d = \sum_k \binom{d}{k}(-B)^k, \quad \text{with} \quad \binom{d}{k}(-1)^k = \frac{\Gamma(-d+k)}{\Gamma(-d)\Gamma(k+1)}$$

and $(\varepsilon_k : k \geq 0)$ is a white noise process. It has been shown in [11] that for $d \in (-1/2, 1/2)$, X is stationary and invertible, and its autocorrelations satisfy $r(k) \sim ak^{2d-1}$ as $k \to \infty$, where a is a finite positive constant independent of k. Moreover, Cox (1984) [4] showed that the aggregated time series $X^{(m)}$ of a fractional ARIMA(p, d, q) process satisfy (3.2) and X is asymptotically second-order self-similar with self-similarity parameter $H = d + 1/2$, for all $0 < d < 1/2$ [20].

Fractional ARIMA processes are much more flexible with regard to the simultaneous modeling of the short-term and long-term behaviour of a time series than fractional Gaussian noise, mainly because the latter, having only the three parameter m, a and H, is not capable of capturing the wide range of low-lag correlation structures encountered in practice [20].

5.3 Superposed ON/OFF Sources

The model presented here was first introduced by Mandelbrot [23] and later extended by Taqqu and Levy [34]. The basic idea of the model is the construction of self-similar processes based on aggregating many simple renewal reward processes exhibiting inter-renewal times with infinite variances.

Following Mandelbrot's original work, an idealized[1] ON/OFF source model is characterized by a reward sequence $\{W_l : l \geq 0\}$, i.e., a 0/1 valued discrete time stochastic process, with $W(l) = 1$ or 0, depending on whether or not there is a packet at time l. Thus, the reward sequence $\{W(l)\}$ consists of a sequence of 1's ("ON-periods") and 0's ("OFF-periods"). Let $p \equiv P(\text{a given period is an ON-period}) = 1/2$, and assume that the lengths of the ON- and OFF-periods are independent and identically distributed (i.i.d.) integer-valued positive random variables, denoted $U_k,\ k \geq 1$. Let $S_k = S_0 + U_1 + U_2 + \cdots + U_k,\ k \geq 0$ be the corresponding renewal times. We assume that $\{S_k,\ k \geq 0\}$ is stationary. This can be achieved by choosing the distribution of S_0 in the following special way:

$$P(S_0 = u) = \frac{1}{E\{U\}} P(U \geq u + 1), \quad u = 0, 1, 2, \ldots$$

To ensure the stationarity of the reward sequence $\{W(l),\ l \geq 0\}$, let $l = 0$ be in an ON-period with probability 1/2.

Next suppose that here are M i.i.d. sources, where the m^{th} source $(m = 1, \ldots, M)$ has its own reward sequence $\{W^{(m)}(l),\ l \geq 0\}$. Then the superposition of cumulative reward ("packet load") at time l is $\sum_{m=1}^{M} W^{(m)}(l)$.

[1] We consider here *idealized* ON/OFF models where an ON-period can be followed by an ON-period and an OFF-period can succeed another OFF-period.

Aggregating this load through (non-overlapping) time blocks of size b, we get

$$W^*_{M,b}(j) = \sum_{l=bj+1}^{b(j+1)} \sum_{m=1}^{M} W^{(m)}(l), \quad j = 0, 1, 2, \ldots,$$

where j denotes the aggregation block number. We are interested in the statistical behaviour of the sequence $\{W^*_{M,b}\}$ for large M and b. This behaviour can only depend on the distribution of U in such a way that, as $M \to \infty$ and $b \to \infty$, $\{W^*_{M,b}\}$ adequately normalized is *fractional Gaussian noise* $\{G_{H,\sigma}(t), t \geq 0\}$, the only Gaussian sequence which is self-similar. To obtain fractional Gaussian noise we suppose that U has a hyperbolic tail distribution, that is, it satisfies

$$P(U > u) \sim cu^{-\alpha} \quad \text{as} \quad u \to \infty, \quad 1 < \alpha < 2, \tag{5.2}$$

where c is a positive finite constant, independent of u. For example, U may have a discrete Pareto-type [2] distribution. Under the conditions stated above the following holds [34]:

Theorem 1 *For large enough source number M and block aggregation size b, the cumulative load $\{W^*_{M,b}(j),\ j \geq 0\}$ behaves statistically as*

$$\mathcal{L} \lim_{b\to\infty} \mathcal{L} \lim_{M\to\infty} b^{-H} M^{-1/2} \Big(W^*_{M,b}(j) - \frac{bM}{2} \Big) = G_{H,\sigma}(j),$$

where

$$H = (3 - \alpha)/2, \quad \sigma^2 = c/(4E\{U\}2(\alpha - 1)(2 - \alpha)(3 - \alpha))$$

and $\mathcal{L}\lim$ means convergence in the sense of the finite-dimensional distributions (convergence in law).

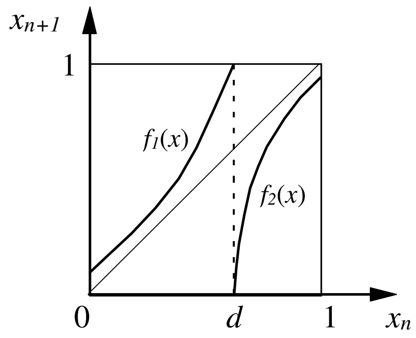
Heuristically, Theorem 1 states that the mean level $bM/2$ provides the main contribution for large M and b; fluctuations from that level are given by the fractional Gaussian noise $G_{H,\sigma}(j)$ scaled by a low order factor $b^H M^{1/2}$.

Note that $1 < \alpha < 2$ implies $1/2 < H < 1$. Thus, the main ingredient that is needed for the limiting result to hold is the hyperbolic tail behaviour (5.2), which guarantees the infinite variance property (i.e., high variability) of the ON- and OFF-periods of a source [38].

[2] Appendix D.2 gives the defining properties of the Pareto distribution.

5.4 Chaotic Maps

Another promising approach to modeling packet traffic was recently proposed by Erramilli and Singh (1990) [6] and is based on deterministic *chaotic maps*. Chaos is a dynamical system phenomenon in which simple, low order, nonlinear deterministic equations can produce behaviour that mimics random processes. To illustrate the underlying idea, consider a nonlinear map $f(\cdot)$ that describes the evolution of a state variable $x_n \in (0,1)$ over discrete time as $x_{n+1} = f(x_n)$.



$$f(x) = \begin{cases} f_1(x) & \text{if } 0 < x \leq d \\ f_2(x) & \text{if } d < x < 1 \end{cases}$$

and

$$x_{n+1} = f(x_n)$$

Figure 5.1: Chaotic maps as models of packet traffic

For the map to be chaotic, $f_1(\cdot)$ and $f_2(\cdot)$ must satisfy the SIC (*sensitive dependence on initial conditions*) condition[3]. By definition,

Definition 5.1 *A map $f : J \to J$ has* sensitive dependence on initial conditions *(SIC) if there exists $\delta > 0$ such that, for any $x \in J$ and any neighborhood N of x, there exists $y \in N$ and $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$, where $f^n(x) = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}(x)$*

That is, a map displays sensitivity to initial conditions if there exist points arbitrarily close to x which eventually separate from x by at least δ under iteration of $f(\cdot)$. In other words, SIC implies that two trajectories starting with almost identical initial conditions will evolve along distinctly different trajectories. In case of deterministic dynamical systems if initial conditions are known, the future behaviour of the system can be computed for all time. In practice, initial conditions can be specified with only a finite accuracy.

[3]This statement is popularly taken to be the defining property of chaos, because it is associated with the unpredictability of chaotic systems. To give the exact definition for chaotic maps more assumptions are needed (see [29]).

These uncertainties in initial conditions are magnified at an exponential rate if SIC exist, making the behaviour unpredictable [29].

The packet generation process for an individual source can now be modeled by stipulating that the source generates one or no packet at time n depending on whether x_n is above or below an appropriately chosen threshold. If $f(\cdot)$ is a chaotic map, the resulting packet process can mimic complex packet traffic phenomena. (Note, that the defining formulae for $f(\cdot)$ is fully deterministic and a given initial value x_0 fully determines the series of x_n. It can be viewed as a particular realization of a stochastic process.)

In particular, Erramilli and Singh have shown that a simple, two parameter nonlinear chaotic map referred to as the *intermittency map* can capture many of the above-mentioned fractal properties in actual packet traffic measurements [20]. The defining equations for that map are the following:

$$x_{n+1} = \begin{cases} \varepsilon + x_n + c x_n^m & \text{if } 0 < x_n \leq d \\ \frac{x_n - d}{1-d} & \text{if } d < x_n < 1 \end{cases} \qquad \text{where} \quad c = \frac{1-\varepsilon-d}{d^m}$$

As can be seen from the equation the model has only three parameters: ε, c and m. (I shall use this model to generate synthetic traffic traces later on and the parameter estimation method for fitting the model to real traffic patterns will be explained in Chapter 8.1.4.)

The generation of synthetic traffic via nonlinear chaotic maps makes the dynamical system approach to packet traffic modeling particularly appealing. Once an appropriate chaotic map has been derived from a set of traffic measurements, generating a packet stream for an individual source is generally quick and easy. On the other hand deriving an appropriate nonlinear chaotic map based on a set of actual traffic measurements, currently requires considerable guessing and experimenting [20].

5.5 Alternative Modeling Techniques

In the present context, alternative modeling technique means that using self-similar traffic models to describe packet traffic exhibiting the so-called self-similar features is not the only way one can follow. In the following, two modeling approaches are presented and neither of them based on self-similar processes. The first approach is based on traditional short-range dependent traffic models. The second one is attacking the assumption of stationarity in case of actual traffic traces and tries to emphasize that the revealed self-similar properties may be due to non-stationarity in the collected data.

5.5.1 Pseudo long-range dependent models

Another modeling approach is based on traffic models which are based on Markov chains with finite state space but exhibit second order self-similar behaviour over several, for example 4–6 timescales. Since these models are, however, necessarily short-range dependent in the strict mathematical sense, one can call them *pseudo long-range dependent* [3].

5.5.2 Non-stationary models

The self-similarity hypothesis to traffic modeling has also resulted in discussions and arguments that put into question a basic tenet of traffic modeling, namely the assumption of *stationarity*. In particular, Duffield et al. [5] put forward the argument that the empirically observed large Hurst parameter of the data set in [38] may be due to non-stationarity in the data, i.e., apparent shifts in the mean rate [9].

While such alternate hypotheses are useful to keep in mind when dealing with self-similar phenomena, the following two points relate to the aggregate traffic traces:

- Apparent shifts in the mean are in fact characteristic of a trace of a stationary, long-range dependent process; the higher the Hurst parameter of the process, the more pronounced the "level shifts".
- From a more statistical viewpoint, there exist nowadays methods that can distinguish between long-range dependence and shifting means (e.g., see [35]).

Part II

Investigation on ATM Networks

Chapter 6

Measurements

"Measure what is measurable,
and make measurable what is not so."

Galileo Galilei (1564–1642)

The second part of the thesis gives the full description of my investigations on ATM networks. To present the findings, the methodology followed separates into three main parts under the names Measurements, Analysis and Modeling.

The present chapter gives an overview about the measurements made on the Finnish FUNET ATM network. But before the technical details the principles of ATM are summarized. Next, the description of the measurement unit and the measured network can be found, together with the description of the measured data sets.

6.1 The Very Basics of ATM

Asynchronous Transfer Mode (ATM) is a packet based, connection oriented transfer mode based on asynchronous time division multiplexing and the use of fixed length cells. Each cell consists of an 48-byte-long information field and additional 5 bytes of header information. The header is primarily used to identify cells belonging to the same virtual channel, and to perform the appropriate routing. Because of the connection oriented technique, cell sequence integrity is preserved per virtual channel.

The information field of ATM cells is carried transparently through the network. All services (i.e., voice, video, data) can be transported via ATM, including connectionless services [28].

From my point of view, the data traffic of the measured ATM network is only considered to be a continuous cell stream, independently from the cell header and information fields. I focus on the aggregated traffic in a particular point of the network so the measurement here means to register all the cell arrival times for each cell traveling through the observation point. Because of the limits of the measurement unit used, only a limited amount of cells can be registered in such a way. When a longer measurement period is used the registered data is the number of cells in a given time interval instead of the time stamps for every single cell.

Though the measurement unit was able to register more data from the traffic stream than only the time stamps or the number of arrivals, during the analysis only the above mentioned data is necessary. (As an exception, in case of the FUNET measurements I used the VPI and VCI fields extracted from the cell headers to identify the source of the cells to get more information about the traffic load from different parts of the network.)

6.2 The Measurement Unit

The measurements were made by an HP Broadband Series Test System equipment on the FUNET ATM network. The present section provides a brief description about the measurement unit and focuses on the main features of the measurement capabilities used.

6.2.1 HP Broadband Series Test System

The Hewlett-Packard Broadband Series Test System is a set of tools for ATM and Broadband ISDN development. Using a totally modular approach, this complete test system integrates physical layer testing with higher layer services protocol testing [40]. (During our work only the cell capture capability of the equipment was used.)

Modularity is one of the key advantages of the HP Broadband Series Test System. The three main parts of the system are the UNIX host controller and user interface, the Cell Protocol Processor (CPP) and the Line interface modules.

6.2.2 Controller and user interface

An embedded UNIX controller executes higher layer protocol and services test software, and controls the Cell Protocol Processor and the line interfaces. The X-Motif user interface simplifies the definition of PDU sequences, and provides quick access to complex simulation and load generation sequences. The UNIX/C user programming environment is also supported.

During the measurements only the lower layer measurements were used.

6.2.3 Cell Protocol Processor

The Cell Protocol Processor (CPP) provides protocol testing features for the test system. The CPP can be configured with several different line interfaces, including the SDH STM1 line interface. The main features of the CPP are:

- The CPP can generate up to eight channels of user-defined ATM traffic. It also provides both real-time and post-processed measurements on ATM traffic.
- Measurements can be displayed in real-time either numerically or graphically. These measurements can be recorded to disk for storage and later use.
- Live PDU data (131,072 cells) can be captured into the 8MB of capture RAM. All captured data is timestamped with the calendar time (resolution = 0.01 μs). Incoming cells are filtered using up to eight pattern matchers.
- Real-time statistics can be calculated on the incoming cell streams (i.e., total cells received).
- The CPP supports a UNIX/C user programming environment.

6.2.4 155 Mb/s Optical Line Interface

The 155 Mbps Optical Line Interface generates and analyzes ATM cells mapped into a 155 Mbps SONET/SDH or cell-based signal. This module provides test capability at the ATM layer and SONET/SDH or cell-based physical layers. It can also be used with the CPP module to provide further test capability at the ATM, AAL and Services layers.

Beside the cell capture possibility of the CPP, during the measurements the cell-based measurement was used provided by the line interface. The recorded data was the number of physical layer cells received per seconds.

6.3 The FUNET Measurements

The measurements were made on the FUNET ATM network. "FUNET" stands for "Finnish University and Research Network" which network provides to its members primarily Internet-network services based on TCP/IP-protocol. All these services are provided by CSC—Center for Scientific Computing which is a private, non-profit, government-sponsored organization. CSC is a national service center that specializes in scientific computing and data communications providing modeling, computing and information services for universities, research institutes and industry.

6.3.1 The FUNET network

The FUNET long-distance network is built on Telecom Finland's ATM-network. Telecom Finland is the country's largest telecommunications company, fully owned by the Finnish State.

All the Nordic national networks (FUNET, DENnet, ISnet, SUNET and UNINETT) are connected to the Nordic Backbone Network (NORDUnet) which has a connection point in Stockholm, Sweden. In many aspects the backbone NORDUnet operates as the international branch of these networks. NORDUnet has connections to the US backbones, the European backbones and to networks in central and eastern Europe [21].

The topology of the whole FUNET and NORDUnet networks are presented in Appendix B on page 80.

6.3.2 Measuring configuration

The measurement was made at the Center for Scientific Computing (CSC) in Espoo, Otaniemi. As can be seen from the network topology (see Appendix B) this location is in the logical center of the whole network. All the international links start from here, including the main crosslink to Stockholm, which link connects the FUNET network to NORDUnet.

Our measurement equipment was inserted between the network and the high-capacity ATM switch situated in Espoo. From that point all the ATM traffic from the FUNET network transported through that switch and the traffic generated at the CSC and transmitted to the rest of the world could be monitored. Figure (6.1) gives the logical measuring configuration used through the whole measurement.

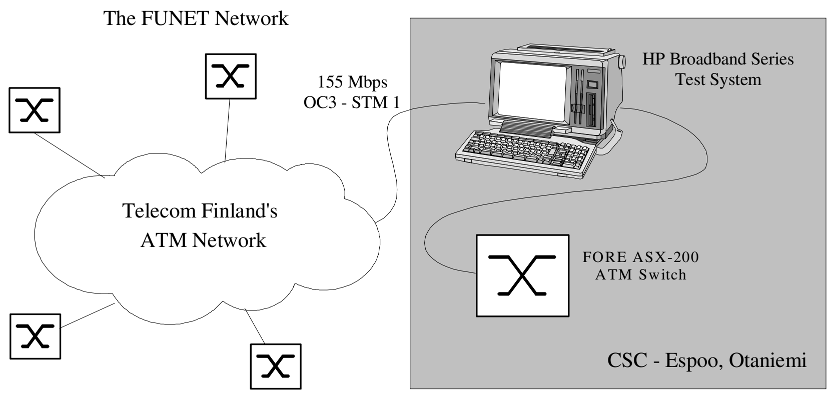


Figure 6.1: The FUNET measuring configuration

6.3.3 Measured traffic

As can be seen from the measuring configuration above, the measured traffic was the aggregated traffic at the most heavily loaded point of the whole FUNET network. As it is the usual case in traffic measurements focusing on the aggregated traffic stream in a particular point of a network, it is almost impossible to gather information about all the sources generating that aggregated traffic. In our case the only information available were the VPI and VCI addresses extracted from the registered ATM cells. Figure (6.2) shows the pie-chart for the captured ATM cells according to the source locations.

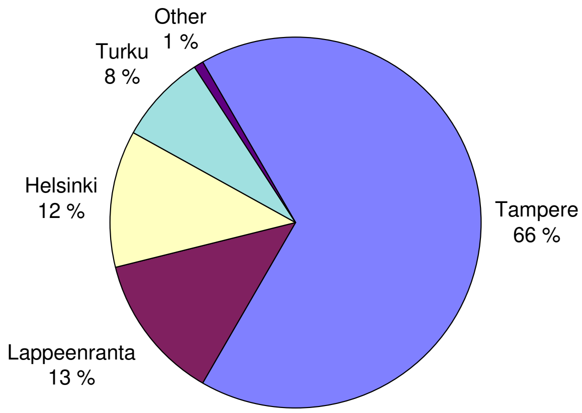


Figure 6.2: Distribution statistics of traffic sources by physical location

It can be seen from the figure, that the two-third of all the network traffic comes from Tampere, which is not surprising since the Technical University of Tampere is situated in that city. The network traffic from Helsinki, Turku and Lappeenranta gives the remaining part of all the network load.

During the measurements, two types of data collections were made. In the first four cases the measured data was the time stamp of the arrival time instant for every single cell on the link. Because of the upper limit for the number of captured cells each measured data file contains 131,072 time stamps only, which was about 3-5 seconds according to the network load. (Besides the time stamps the measurement unit registered the VPI and VCI fields from the cell headers also.)

For the long-term analysis longer measurement periods were needed, so in the next type of measurements the recorded data was the number of cells received in a one second interval. In this case the time interval of the observation could take several minutes long.

Filename	#packets	Total time (sec)
FUNET1	131,072	3.962,485
FUNET2	131,072	5.120,411,4
FUNET3	131,072	4.371,536
FUNET4	131,072	6.410,039,4
FUNETSTA.T3	14,807,546	425
FUNETSTA.T4	43,768,430	1964

Table 6.1: Qualitative description of the sets of the FUNET measurements

A summary description of these data sets is given in Table (6.1). The first four sets of traffic measurements belong to the first type, which again means, that the registered data was the arrival time stamp for every cell. The files FUNET1, FUNET2 and FUNET3 contain traffic data captured from the incoming traffic from the whole country to the CSC, and the FUNET4 measurement was made on the outgoing link. In case of the last two measurements, the registered data was the number of cells received in every second on the incoming link.

As the numbers imply, the average traffic load was about 14 Mbps for the first three measurements, and about 8 Mbps in case of the FUNET4 data. (The value 14 Mbps means that the measured ATM link was loaded at 10% of its full capacity on average.)

In the following, I refer to the above listed data as the 'FUNET measurements' and the filenames are used to identify each data set.

Chapter 7

Analysis

"Errors using inadequate data are much less than those using no data at all."

Charles Babbage (1792–1871)

"Is the ATM traffic self-similar?" The question is brief and goes straight to the point, but senseless in that form. The mathematical notion of self-similarity deals with infinite data sets with infinite structure, but in practice we are always dealing with finite data sets so it is in principle not possible to decide whether the asymptotic relationships in definition of self-similarity hold or not. The only thing we can talk about whether or not the data set under discussion possess features that can be modeled by self-similar processes rather than by classical traffic modeling methods.

The other weak point of the question is the definition of the expression 'ATM traffic'. Because of the wide range of services running on a working broadband ATM network, the so-called 'typical ATM traffic pattern' does not exist. Traffic data collected from different points of the same network or from the same point at different times can have totally different characteristics. During my work, the traffic measurements were made on the FUNET ATM network where the measured data was the aggregated traffic on an ATM link. Thus, all the results from the analysis described below are valid only for that measurement.

As a consequence, the question raised at the beginning of this chapter should be replaced with the following one: "Is there any self-similar feature present in the FUNET measurement data?" This chapter tries to answer this question. But before the self-similarity tests a basic statistical analysis is presented to reveal the fundamental characteristics of the measured traffic traces.

7.1 Tools

Before writing about the analysis performed on the measured data sets, it is better to give a brief description about the analysis tools used.

In the present context, analysis means to implement statistical methods and run the computations on a computer. The input data for the algorithms is the measured or generated traffic data file (see Figure (7.1)).

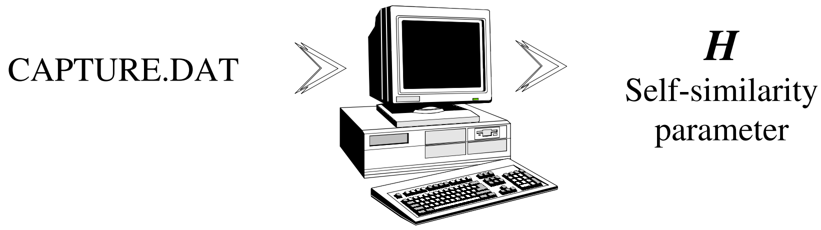


Figure 7.1: Methodology used for analyzing data sets

The routines used are based on a software package developed at the Technical University of Budapest within the scope of a diploma work in 1996 [14]. The source codes were written in ANSI C and C++ languages. (Appendix E gives the list and a brief description of the programs, but the descriptions for all the routines can be found in [14] in more details.)

The hardware platform was mainly an IBM-PC 486DX2 (12MB) computer which had enough computation power for most of the test algorithms. In case of the routines that required more CPU time (i.e., calculating periodograms) a SUN Workstation was used.

To present the results graphically and do the post-processing, the Microsoft Excel v7.0 was used.

7.2 Basic Statistical Analysis

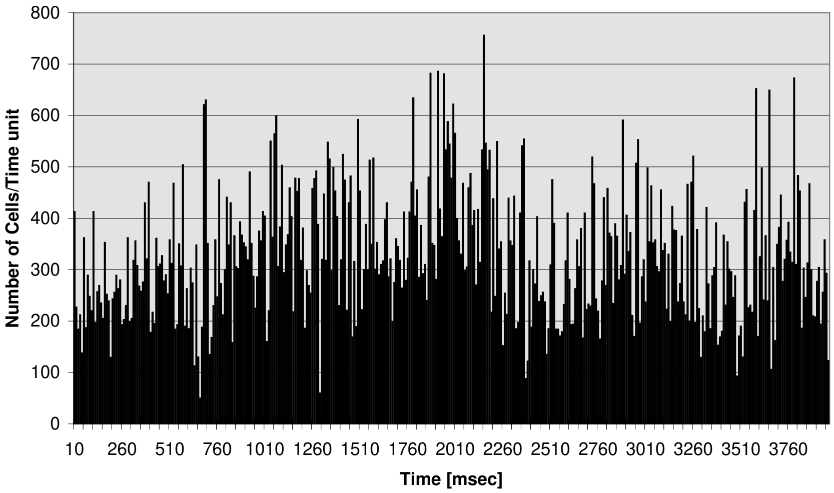
The basic statistical analysis here covers all the statistical test used to reveal the fundamental properties of a data set and not to test the self-similarity property directly. Of course, some of these methods are able to capture some traces of the presence of self-similarity but the calculated value is not the self-similarity parameter of the process. Instead, basic descriptors such as distributions and histograms are calculated.

7.2.1 Histograms

It is very useful to draw a histogram for traffic measurement data sets to get a qualitative picture about the load trend and burstiness of the traffic. In case of the following histograms the plotted value against a given time interval is the number of cells received in that interval. If we plot that diagram for different time scales, we get some information about the burstiness of the process.

At first, the data set FUNET1 was chosen for the analysis. The file contains about 130,000 cell time stamps and the length of the measurement was nearly 4 seconds. Figure (7.2) shows the histogram for the whole measurement, where the time unit is 10ms.

Figure 7.2: FUNET1 - Histogram (Time unit = 10 msec)



One can conclude at first glance that the traffic is highly bursty—the plot shows sharp peaks separated by narrow but deep dwells. It is worth to plot the same histogram for a Poisson process (see Figure (7.3)), and compare the two diagrams. (The parameter λ for the Poisson process was set to be the same as the mean rate calculated from the measured data.)

Compared to the real traffic trace, the Poisson process is much smoother in that time scale—there are no sharp peaks and deep dwells in the diagram.

If we reduce the length of the time interval to 1 milliseconds and after that to 100 microseconds, we can form a notion of the structure of the process at different time scales. (Figure (7.4) and (7.5) shows the resulting plots, where the analyzed data was the first 10% and 1% of the whole measurement.)

Figure 7.3: POIS1 - Histogram (Time unit = 10 msec)

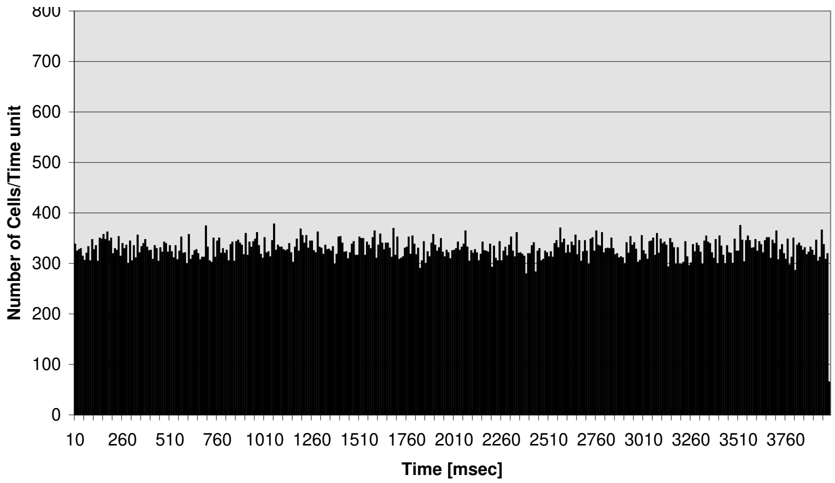


Figure 7.4: FUNET1 - Histogram (Time unit = 1 msec)

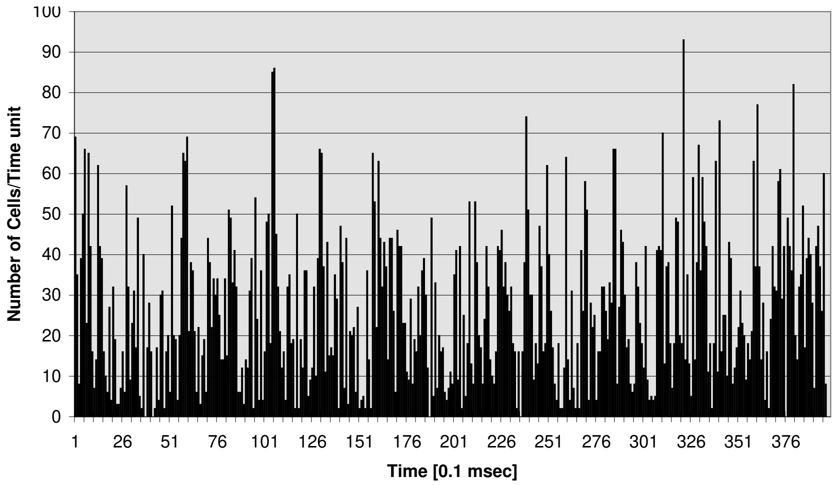
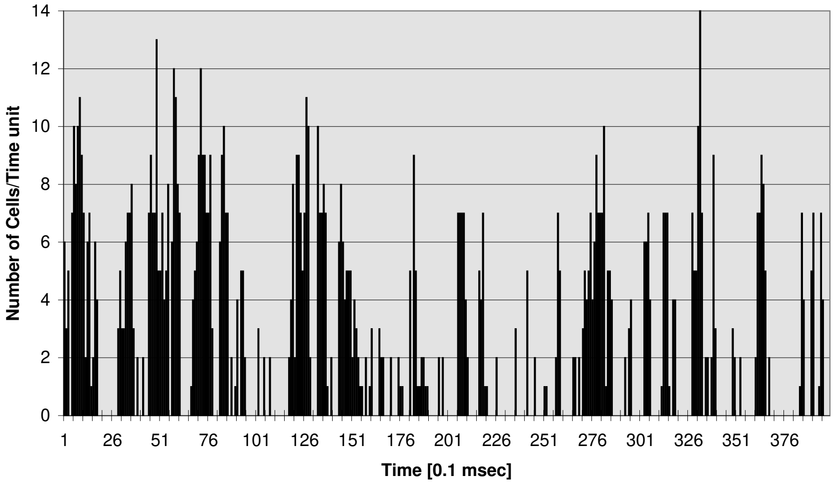
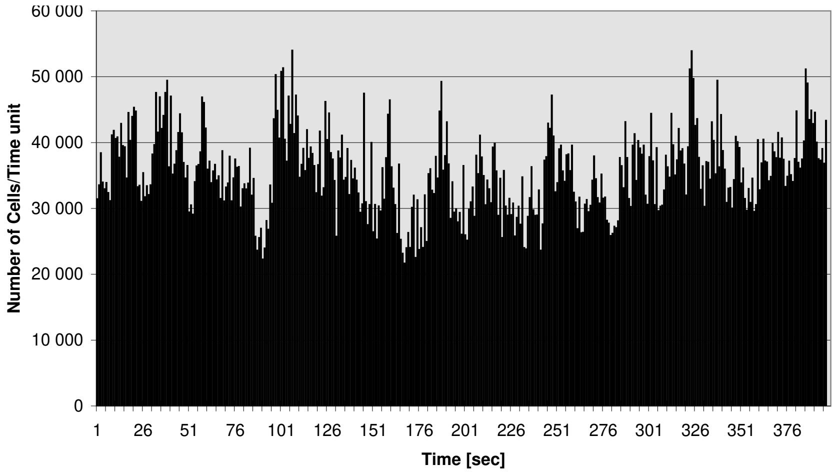


Figure 7.5: FUNET1 - Histogram (Time unit = 0.1 msec)



As can be seen, the traffic process is bursty at those two time scales also. If we want to analyze the process at higher time scales, we should turn to the FUNETSTA.T3 traffic trace where the registered data was the sequence of cell counts in every seconds. The length of the measurement was 425 seconds. (Figure (7.6) shows the whole measurement interval.)

Figure 7.6: FUNETSTA.T3 - Histogram (Time unit = 1 sec)

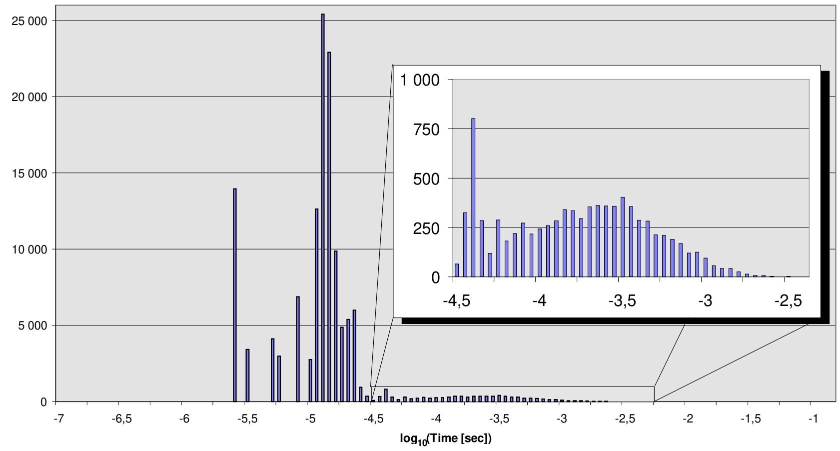


The conclusion derived from the plot is the same: in spite of the time interval is 2 orders of magnitude higher than it was in case of the FUNET1 data set, the process is bursty again. (Note, that the burst-within-burst (or self-similar) structure is based on this symmetricity across time—an exactly self-similar process has the same structure at every time scale.)

7.2.2 Inter-Arrival Time Distribution

The probability distribution function of cell inter-arrival times (IAT) is one of the main characteristics of a traffic process. (Although during the testing for self-similarity, I do not analyze the IAT process of the measured data sets, but in Section 7.2.5 the standard deviation of that process is calculated.) Otherwise it is worth to plot the distribution for the IAT process in case of the FUNET1 data set while it reveals a special feature of the process. Figure (7.7) presents the obtained diagram, where the time axis is scaled logarithmically ($\log T$) and the ordinates are the cell counts satisfying $T < \mathrm{IAT} < mT$, where $\log m = 0.05$.

Figure 7.7: IAT Distribution - FUNET1



Analyzing the histograms in the previous section, one can conclude that the traffic process under discussion is bursty at many time scales. This means that the apparently busy periods are separated by idle intervals even in higher time scales which results in high inter-arrival times with certain positive probability. From the IAT plot it can be seen that the inter-arrival times span across more than 3 orders of magnitude. (The tail of the distribution is magnified to reveal the important part of the plot at high time scales.)

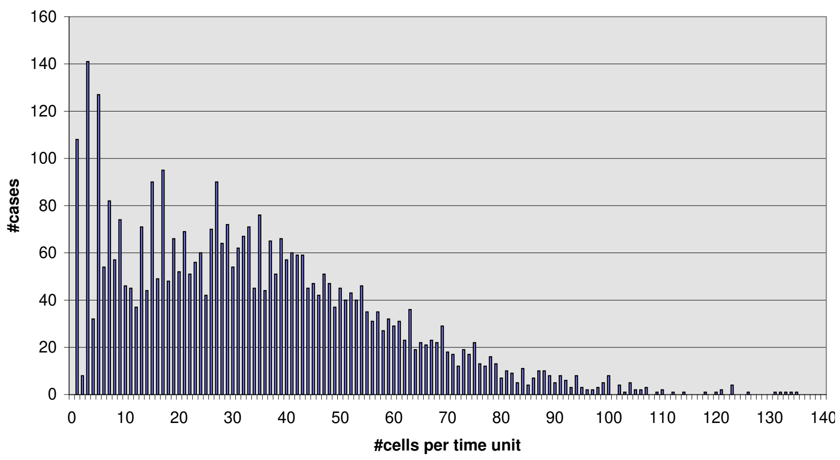
To analyze the distribution in more details, it would require the knowledge of additional information about the measured traffic, such as the characteristics of the running services on the network, the number of source-destination pairs spoke up during the measurement, the source characteristics for all the sources, and so on. To collect all these information in case of real-time, high-speed traffic measurements is completely impossible. But the IAT distribution plot is useful in the modeling phase in my work (see Section 8), where comparing the plot with the same plots obtained from the synthetic traffic traces generated by the models reveals the differences or similarities between the synthetic and real traffic processes.

7.2.3 Cell distribution

Beside the IAT process, the sequence of the number of cells arriving in contiguous fixed time intervals is the main characteristic of an arrival process. During the analysis, all the test algorithms used are based on this process in estimating the self-similarity parameter H. That is why it is worth to plot the distribution of the process.

As for the FUNET1 data set, the analyzed process was the sequence of cell counts in every 1 millisecond interval. The distribution of this process can be seen on Figure (7.8), where the number of cases are plotted against the number of cells received within the 1 msec time interval for the whole measurement.

Figure 7.8: FUNET1 - Cell distribution



Without any further calculation, one can only conclude that the outlook of the graph is steady and nearly bell-shaped (except only for small values next to the origin). Again, that diagram is also useful to compare the goodness of fit of the models to real traffic traces.

7.2.4 Autocorrelation function

The autocorrelation function $r(k)$ gives deep characterization of the correlation structure of a given time series. In the present context, the autocorrelation function is important for us because the shape of the function reveals whether or not the long-range dependence is present in the measured data sets. As it was mentioned in Section 3.2.1, processes with long-range dependence are characterized by autocorrelation function that decays hyperbolically as the lag increases. This behaviour is drastically different from those of the more conventional—i.e., short-range dependent—processes. The latter are characterized by an exponential decay of the autocorrelations, i.e., $r(k) \sim a^k$, $(0 < a < 1)$ as $k \to \infty$.

At first, I analyzed the FUNET1 data set where the process under examination was the sequence of cell counts received in every 100 microseconds. The values of the autocorrelation function $r(k)$ were calculated using the following equation:

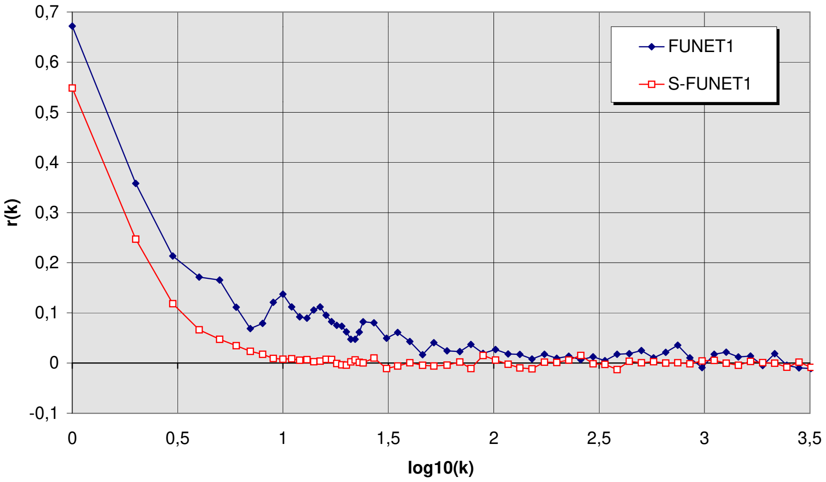
$$r(k) = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k}(X_i - \overline{X})(X_{i+k} - \overline{X})}{\frac{1}{N-k}\sum_{i=1}^{N-k}(X_i - \overline{X})^2}, \quad \text{where} \quad \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Here, N is the number of values in the data set ($N = 39,624$). The resulting plot can be seen on Figure (7.9).

But before starting to evaluate the obtained diagram, calculate the same function for the same data set after removing the long-term correlations. This can be done in an easy way by shuffling the data randomly. To re-shuffle a measured traffic trace one should only mix the sequence of cell inter-arrival times randomly and then rebuild the cell stream again. This shuffling was done in case of the FUNET1 file and the result is referred to as the 'Shuffled-FUNET1' or 'S-FUNET1' file. Figure (7.9) shows that diagram also.

In case of the shuffled FUNET1 data set, the short-term correlations are strong but as the value of k becomes higher and higher, the function drops toward zero exponentially fast. For the value of $k = 10$ ($\log k = 1$) the correlations disappear. But, on the contrary, the autocorrelation function of the FUNET1 set has the so-called long-tailed property, which means that the tail of the function tends to 0 much slower. The correlations remain positive for values of k much higher than 100. This property indicates the presence of long-range dependence in the traffic data set.

Figure 7.9: Autocorrelation function for sets FUNET1 and S-FUNET1



7.2.5 Parameter estimation from long-range dependent traffic data

Long-range dependence has important implications for parameter estimation. As will be shown, many estimators are more biased and converge more slowly than in the traditional short-range dependent case [3].

Estimation of the mean rate

The mean rate of a stationary traffic stream over time interval $[0, T]$ is estimated by the sample mean

$$\hat{m}_T = \frac{A_T}{T}$$

where A_T is the cumulating arrival process. With long-range dependence, the estimation of the mean rate may become so inaccurate that the whole notion of mean rate must be interpreted in a different way [3].

To illustrate this, assume that the true traffic obeys the fractional Brownian model presented in Section 5.1 on page 27. The variance of $\hat{m}$ is

$$var\{\hat{m}_T\} = maT^{2H-2},$$

so that bringing the relative standard deviation below ε requires that

$$T > \left(\frac{1}{\varepsilon}\sqrt{\frac{a}{m}}\right)^{1/(1-H)}.$$

The left hand side increases rapidly with respect to H. (For example, if we have $m = 2$Mbit/sec and $a = 200$kbit/s typical values in today's LAN interconnections, then the observation time required for the accuracy of a standard deviation 10% would be $10^{1/(2-2H)}$ seconds which is about five minutes with $H = 0.8$ but more than 27 hours with $H = 0.9$! In the latter case even the meaning of the mean rate parameter becomes problematic, since the deterministic day-time variation has to be taken into account [3].

In the following, I examine the above mentioned effect in case of the FUNET measurements. The analyzed data set was the FUNET1 traffic trace. For the easy interpretation of the results, in parallel to the FUNET1 file I analyzed a simple Poisson process exactly the same way. The parameter λ for the Poisson process was set to be the same as the mean rate calculated from the measured data. Since the FUNET1 file contains 131,072 cells the POIS1 file also has the same number of timestamps. (For more details in generating synthetic traffic trace according to a Poisson process, see Section 8.1 on page 65.)

The analysis method was the following: The mean rate was calculated for different window sizes using the equation

$$\hat{m}_T^{(i)} = \frac{A_{t_i+T} - A_{t_i}}{T},$$

where A_t is the cumulative arrival process, T is the window size satisfying $10t_{cell} < T < t_{total}$ and $0 < t_i < t_{total} - T$ where t_{cell} is the transmission time of a cell over a 155 Mbps link and t_{total} is the total time of the measurement interval. For a given size of T the windows are non-overlapping and evenly distributed over the whole measurement time interval. Further, let $i = \min(25, t_{total}/T)$ which means that in case of small window sizes, only 25 sample mean were calculated.) The values of T form a logarithmically spaced series over the whole measurement interval.

Plotting the values $\hat{m}_T^{(i)}$ against $\log T$ results the diagram of interest: Figure (7.10) shows the convergence of the calculated mean values for the FUNET1 file as the window size increases.

Before drawing the conclusions, compare the diagram with the plot obtained from the Poisson process showed in Figure (7.11).

If we compare the two diagrams, the difference is distinctly visible. In case of the Poisson process the estimate of the mean rate converges much more faster to a given value as the window size increases. The diagram is smooth enough, the points are not scattered so much even in case of small values of T. On the contrary, in case of the FUNET1 data set the diagram is more entangled. Calculating the mean rate from about a thousand observations

Figure 7.10: FUNET1 - Convergence of the estimated mean rate

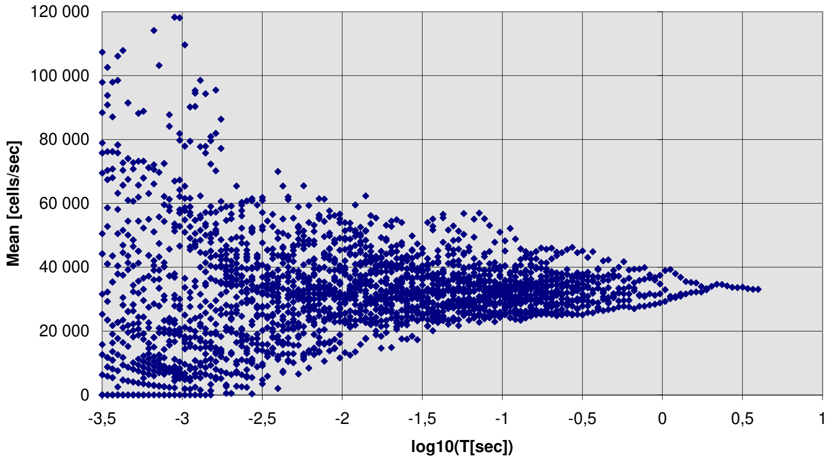
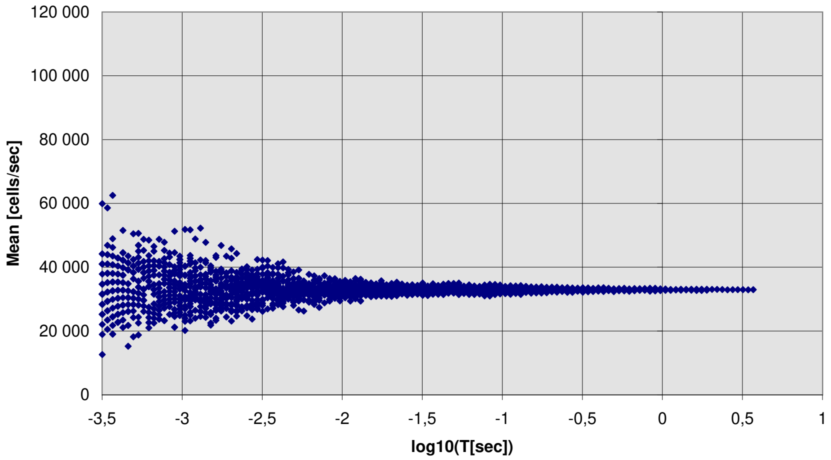


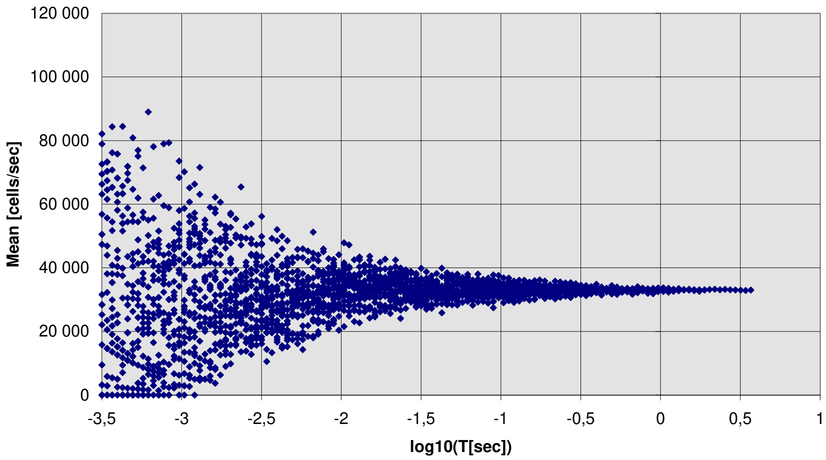
Figure 7.11: POIS1 - Convergence of the estimated mean rate



the result still has considerable uncertainty and even the tail of the plot do not smooth out for large values of T.

The presence of long-range dependence could cause that behaviour. This statement can be verified by removing the dependencies from the data and redoing the analysis. To do this, I re-shuffled the FUNET1 data set the same way as in the previous section to destroy the given correlation structure. For short-range dependent data sets the estimates of the mean rate for the original data set and the re-shuffled data set show only small difference, whereas for long-range dependent data sets the difference is more obvious. (But, of course, this is rather an indication of long-range dependence than a proof.)

Figure 7.12: Shuffled-FUNET1 - Convergence of the estimated mean rate



The obtained plot is similar to the plot of the Poisson process rather than to the FUNET1 plot by the nature of the case. The left-hand side of the broom-shape plot is wider than in case of the POIS1 file because of the distribution of cell inter-arrival times, but the tail of the diagram for large values of T is smooth enough and converges strictly to a given mean value.

To summarize the findings, we can say that the analyzed data set possess a particular feature of long-range dependent processes as far as the parameter estimation from measured data sets concerned.

Estimation of the variances

However, the estimation of the variance function $v(t) = var\{A_t\}$ of a stationary traffic stream with the presence of long-range dependence is difficult for large H, since the sample variances are then both strongly biased and very slowly convergent [3].

Let us again restrict to the simplest case, fractional Brownian traffic. Then consider the convergence of the sample variance to the true variance. We simplify the problem further by assuming that the mean rate is known exactly, and consider the unbiased variance estimators

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(Z(i) - Z(i-1))^2.$$

(Note, that the assumption above does not hold in case of real traffic measurement. The exact mean rate m is not known and has to be estimated by the sample mean as above. This step introduces additional error into the calculations.)

After straightforward calculations and approximations for large values of n (for more details, see [3]), we have:

$$\sqrt{var\{\hat{S}_n^2\}} \sim \begin{cases} n^{-0.5} & H < 3/4 \\ n^{-2(1-H)} & H > 3/4 \end{cases}$$

As can be seen from the formulae, the value $H = 3/4$ separates the region where the relative standard deviation goes to zero as $n^{-0.5}$ from a region where the convergence is still slower and depends on H. Again, there is a big difference between the cases $H = 0.8$ and $H = 0.9$: the sample size required for a standard deviation of 10% is less than 10^4 for the former but greater than 10^6 for the latter [3]!

Next—as in the previous subsection—I shall make a test to reveal that property in case of our FUNET measurements. But instead of calculating the variance function $v(t)$ of the traffic stream, I focus on the standard deviation of the inter-arrival time (IAT) process which is fundamentally come close to that process.

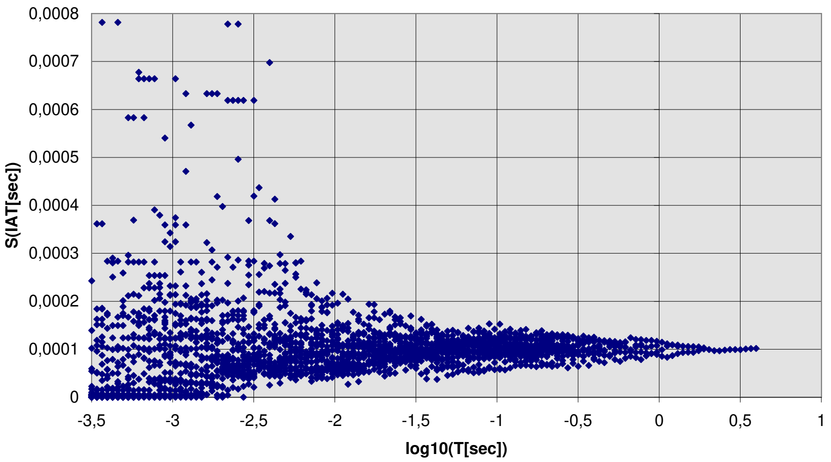
Let X_j denote the arrival time of the jth cell captured ($X_1 \equiv 0$) and $Y_j = X_{j+1} - X_j$ is the elapsed time between the jth and $(j+1)$th cell arrivals. The standard deviation for the inter-arrival time process now can be calculated as follows:

$$\hat{S}_T^{(i)} = \sqrt{E\{{Y_T^{(i)}}^2\} - E^2\{Y_T^{(i)}\}},$$

where $Y_T^{(i)} = \{Y_j : X_j \in [t_i, t_i + T)\}$ for every T and $i = \min(25, t_{total}/T)$. As in the case of the mean rate estimate method discussed earlier, let $10t_{cell} < T < t_{total}$ and $0 < t_i < t_{total} - T$ which again means that for a given size of T the windows are non-overlapping and evenly distributed over the whole measurement time interval.

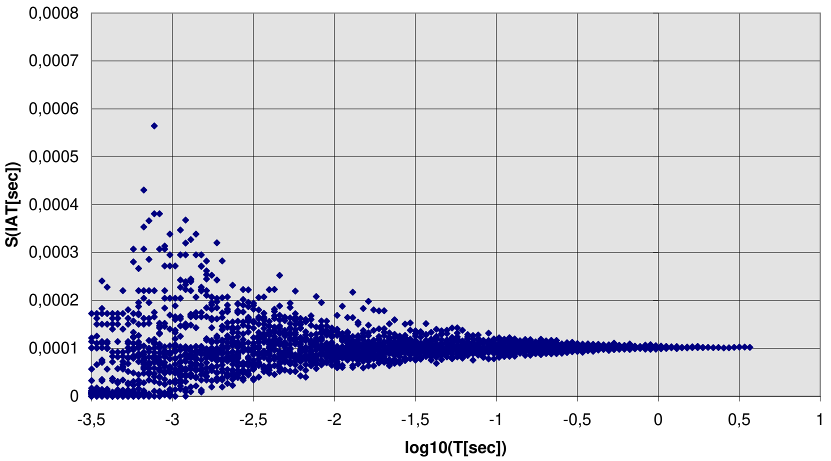
The analysis method applied here is similar to the previous one. The values of the standard deviation $\hat{S}_T$ for the IAT process are calculated using different window sizes T and plotting $\hat{S}_T$ against $\log T$. The convergence of the values of the standard deviation can be seen from the resulting plot as the sample size increases. Figure (7.13) shows the diagram for the FUNET1 data set.

Figure 7.13: FUNET1 - Convergence of $\hat{S}_T$ for the IAT process



In this case it is no use to analyze the POIS1 data set since the variance of the Poisson process is equal to λ which differs from the variance of the FUNET1 inter-arrival time process. But it is useful to analyze the shuffled version of the FUNET1 data file and compare the results (see Figure (7.14)).

Figure 7.14: Shuffled FUNET1 - Convergence of $\hat{S}_T$ for the IAT process



Comparing the two diagrams the difference is clear again. Shuffling the data means to destroy the long-range dependence in the data set which results in smoother and more convergent plot. We can state once again, that the analyzed measured data set possess such a statistical feature which is the characteristic of long-range dependent processes.

7.3 Testing for Self-Similarity

Using the graphical tools described in Section 4.1, namely, IDC-plots, variance-time plots, R/S diagrams, and periodogram plots, I shall analyze the FUNET measurement sets. Each sequence contains 132,072 cells, so the corresponding estimates for the degree of self-similarity based on the above methods can be expected to be accurate.

7.3.1 IDC—Index of Dispersion for Counts

Plotting IDC curves and estimating its slope provides fast graphical method for estimating the Hurst parameter of a given time series.

Again, for a given interval of length L, the index of dispersion for counts (IDC) is given by the variance of the number of arrivals during the interval of length L divided by the expected value of that same quantity:

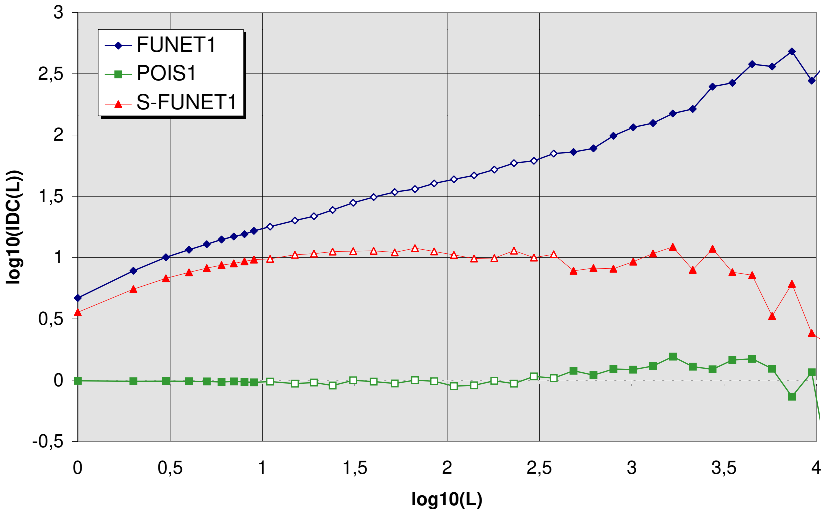
$$IDC(L) = var\{\sum_{j=1}^{L} X_j\}/E\{\sum_{j=1}^{L} X_j\}$$

Note that when plotting log(IDC(L)) against $\log L$, the result is an asymptotic straight line with slope $\gamma = 2H - 1$. Figure (7.15) depicts the IDC curves corresponding to the traces FUNET1, POIS1, and the shuffled version of FUNET1. (The bigger white markers indicate the section of the curve to which the regression line was fitted.) In each case the analyzed process is the sequence of cell counts in every 100μs interval.

The IDC curve for the FUNET1 file increases monotonically throughout a time span that covers 3-4 orders of magnitude and shows an asymptotic slope that is strictly different from the horizontal line and is estimated to be about 0.38, resulting in an estimate $\hat{H}$ of the Hurst-parameter H of about $\hat{H} \approx 0.69$. This behaviour is in stark contrast to conventional traffic models such as Poisson processes, where the IDC is constant.

To verify the result, shuffle the FUNET1 data set once again and plot the IDC diagram for that shuffled traffic trace. The curve starts as in the case of the FUNET1 plot but soon it stops increasing and remain constant

Figure 7.15: IDC Plot



for values of $\log L$ greater than 1. This means that removing the long-range dependencies from the data set, the self-similar feature disappears. But the resulting process is still different from the Poisson process. For the values of L lower that 10 the slope of the curve is positive which indicates the presence of short-range dependence. For the values of L greater that 10 the slope of the diagram is zero but the IDC value is about 10. This means that the traffic is bursty but not self-similar. The result is identical with the conclusions obtained from the autocorrelation function of the process.

The same analysis was made for all of the data sets from the FUNET measurements. Table (7.1) shows the results: the slopes (γ_{idc}) of the fitted lines and the values of the estimated Hurst-parameter $\hat{H}$.

Filename	slope γ_{idc}	Estimate of $\hat{H}_{idc}$
FUNET1	0.38	0.69
FUNET2	0.33	0.67
FUNET3	0.31	0.66
FUNET4	0.44	0.72
FUNETSTA.T3	0.40	0.70
FUNETSTA.T4	0.33	0.67

Table 7.1: Estimated values of H from IDC plots

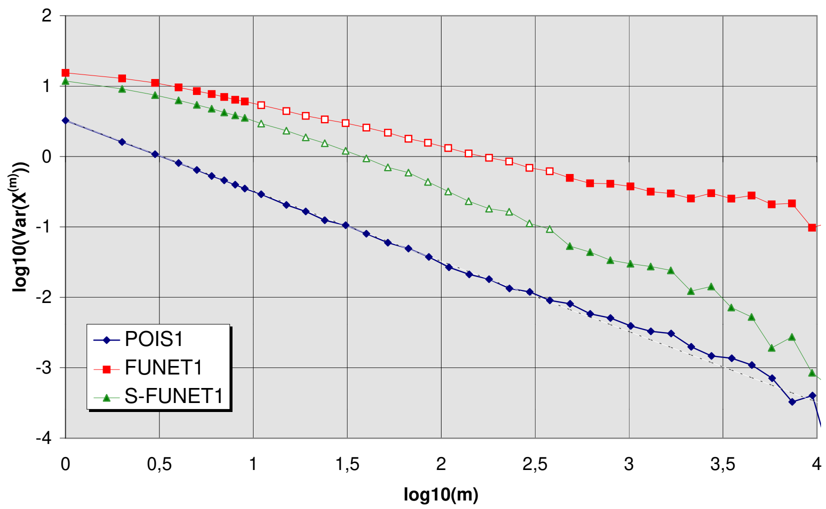
As can be seen from the table, the values of $\hat{H}$ are pretty much the same for all the data sets. It is remarkable, that in case of the last two data sets the analyzed process was the sequence of cell counts in each seconds instead of 100 μs as in the case of the first four sets. In spite of the fact that the time scale was four orders of magnitude higher the Hurst-parameter remained the same.

7.3.2 Variance-time plots

Again, a variance-time plot shows that, if the sample is aggregated by a factor of m (corresponding to Equation 3.1), then the variance of the aggregated versions falls off by a factor of m^{β}, where $\beta = 2(1-H)$. The variance-time plots are obtained by plotting $\log(var(X^{(m)}))$ against $\log m$ and by fitting a line through the resulting points in the plane, ignoring the small values of m. The estimate for the degree of self-similarity is given by $\hat{H} = 1 - \hat{\beta}/2$ where $\hat{\beta}$ is the estimate of the asymptotic slope of the fitted line.

The variance-time plot for the FUNET1, POIS1 and the shuffled version of FUNET1 data sets can be seen on Figure (7.16) and the estimated values of $\hat{\beta}$ and $\hat{H}$ are listed in Table (7.2). (The bigger white markers indicate the section of the curve to which the regression line was fitted.)

Figure 7.16: Variance-Time Plot



Since the variance-time plots and the IDC diagrams are closely related statistical methods, the results obtained from this method are the same as in the previous subsection. The interpretation of the results is also the same:

the estimated slope of the variance-time plot is equal to -0.5 for short-range dependent processes and clearly flatter for traffic traces exhibiting long-range dependent structure.

Filename	slope β_{var}	Estimate of $\hat{H}_{var}$
FUNET1	-0.62	0.69
FUNET2	-0.67	0.67
FUNET3	-0.69	0.66
FUNET4	-0.55	0.72
FUNETSTA.T3	-0.60	0.70
FUNETSTA.T4	-0.67	0.67

Table 7.2: Estimated values of H from Variance-Time plots

Note again, that this is a heuristic test in the sense that the statistical properties of these plots are not known, but in case of huge data sets the method gives a rather accurate picture about the self-similar nature of the underlying time series and about the degree of self-similarity.

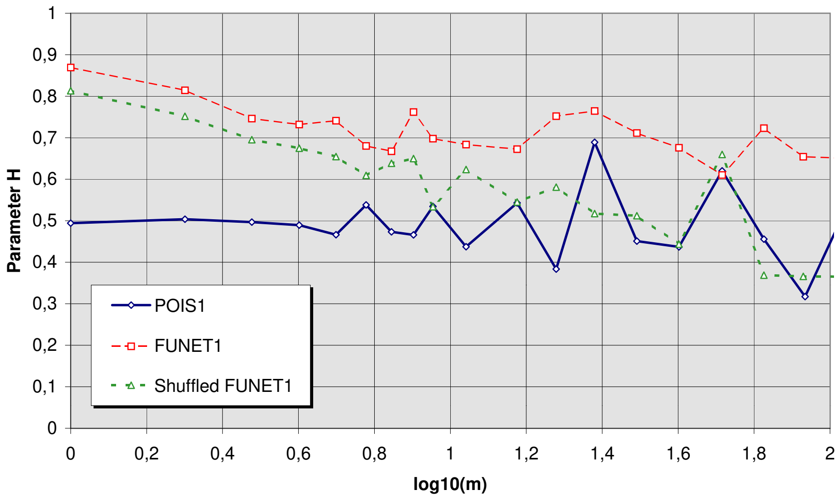
It is easy to derive a new diagram from the variance-time plot, and with the help of this new plot one can follow the evolution of the parameter $\hat{H}$ as the aggregating factor m increases. Seeing that the estimated values of $\hat{H}$ can be derived by estimating the slope of the variance-time plot, plotting the derivative of the same plot using the equation

$$\hat{H}(m_i) = 1 + \frac{1}{2}\frac{\log(var(X^{(m_i)})) - \log(var(X^{(m_{i+1})}))}{\log m_i - \log m_{i+1}} \quad i = 1, 2, \ldots$$

we get an estimate value for H directly as a function of m. Using this method the resulting diagram is shown in Figure (7.17).

Unfortunately, as the aggregation factor m increases, the diagram becomes more and more "noisier" so the exact value of H is not determinable. But the trend is clearly visible. In case of the Poisson process, the value of $\hat{H}(m)$ starts at 0.5 and remains the same as m increases. For the FUNET1 data set, the starting value of $\hat{H}$ is relatively high because of the strong short-range correlations present in the traffic trace, but it drops to about 0.7 and remains the same for larger values of m. As for the shuffled version of the FUNET1 data the value of $\hat{H}$ is high again in the beginning, but it falls constantly to about 0.5 as m increases. (The right side of the diagram is so noisy that the values of $\hat{H}(m)$ become meaningless when m exceeds the value of about 100.)

Figure 7.17: Local estimate of parameter H



7.3.3 R/S Analysis

As the previously described methods, R/S analysis is based on a heuristic graphical approach that tries to estimate the degree of self-similarity of a given data set. In Section 4.1.3 the full description of the method was presented. Here I only summarize the substance of that algorithm.

Given an empirical time series, the whole series is subdivided into K non-overlapping blocks. Then, we compute the rescaled adjusted range $R(t_i, d)/S(t_i, d)$ for a number of values d, where t_i are the starting points of the blocks. For each of d one obtains a number of R/S samples. Plotting $\log R(t_i, d)/S(t_i, d)$ against $\log d$ results in the R/S plot, also known as pox diagram. Next, a least square line is fitted to the points of the R/S plot, where both the smallest and largest values of d are not considered. The slope of that regression line is an estimate for the Hurst parameter H.

Figure (7.18) shows the R/S plot for the FUNET1 measured data set. (For fitting the line only the values d between 35 and 5500 were considered. The values of d greater than 5500 means that the number of R/S samples is less than 7, so those samples are statistically insignificant.) The analyzed process was the sequence of cell counts in every 100μs.

The estimated value of H for this data set is 0.68, which is nearly the same as the values previously calculated by the two another statistical methods.

Again, I calculated the R/S plot for the control Poisson process and also for the shuffled FUNET1 data set. The results can be seen on Figure (7.19) and Figure (7.20).

Figure 7.18: FUNET1 - R/S Plot

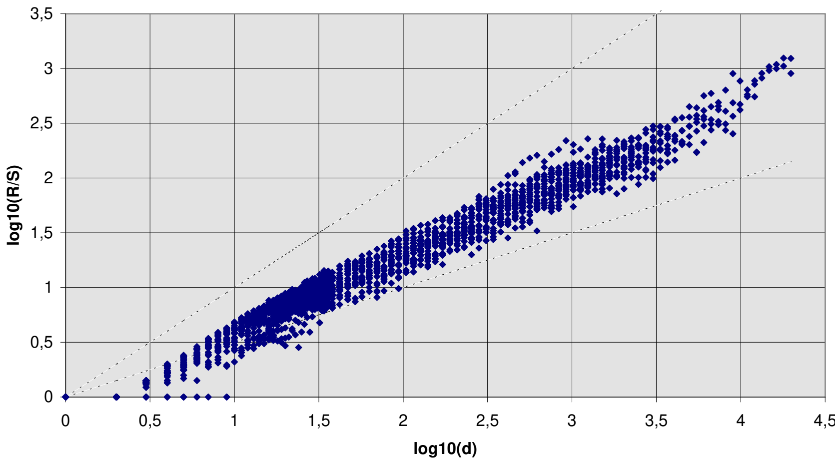
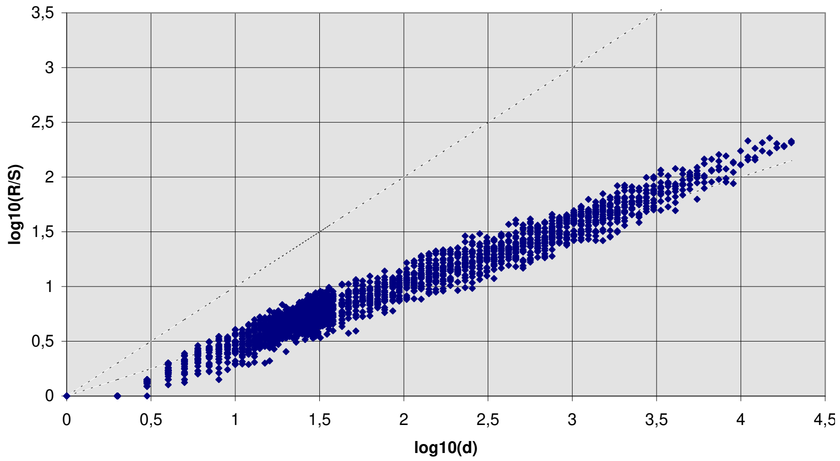


Figure 7.19: POIS1 - R/S Plot

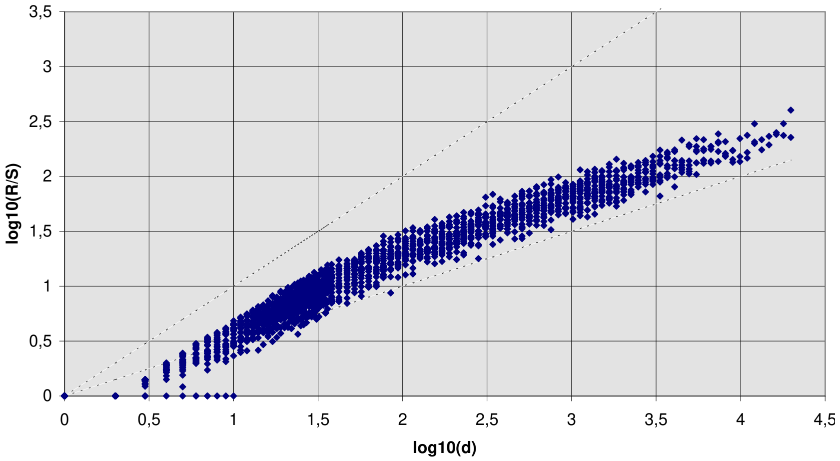


It is remarkable that the estimate for H in case of the Poisson process is 0.53 instead of the theoretical value 0.5. The slope of the diagram for the shuffled FUNET1 data is 0.54.

The same analysis was made for all the FUNET measurement data sets. The resulting values for parameter H are listed in Table (7.3).

For the first four data sets we get what we expected. But in case of the last two data sets, the estimated value of parameter H is higher by about 0.1. The difference between the result and expected value could follow from the

Figure 7.20: Shuffled FUNET1 - R/S Plot



Filename	Estimate of $\hat{H}_{rs}$
FUNET1	0.68
FUNET2	0.67
FUNET3	0.68
FUNET4	0.74
FUNETSTA.T3	0.82
FUNETSTA.T4	0.79

Table 7.3: Estimated values of H from R/S analysis

fact that these data sets contain much less data. Because of the measured data was the number of cells received in a one second interval, the file FUNETSTA.T3 contains only as few as 425 numbers, while FUNETSTA.T4 has 1960 elements. In contrast, FUNET1 contains about 40,000 entries. That is why we should be careful while interpreting the results. From statistical point of view, the values of H calculated for the last two data sets are not so meaningful, but it can be seen that those values are definitely higher than 0.5.

7.3.4 Periodogram plots

Section 4.1.4 gives the full theoretical background for the periodogram-based statistical methods used to determine the self-similarity parameter H. Here, again, only the basics of the simplest method is presented.

The basic idea of the analysis to determine the Hurst-parameter H is to plot the periodogram in a log-log grid, and to compute the slope of the regression line which is fitted to a number of low frequencies. This should be an estimate of $1-2H$. But note again, that in most of the cases this will lead to a wrong estimate of H since the periodogram is not appropriate to estimate the spectral density [3].

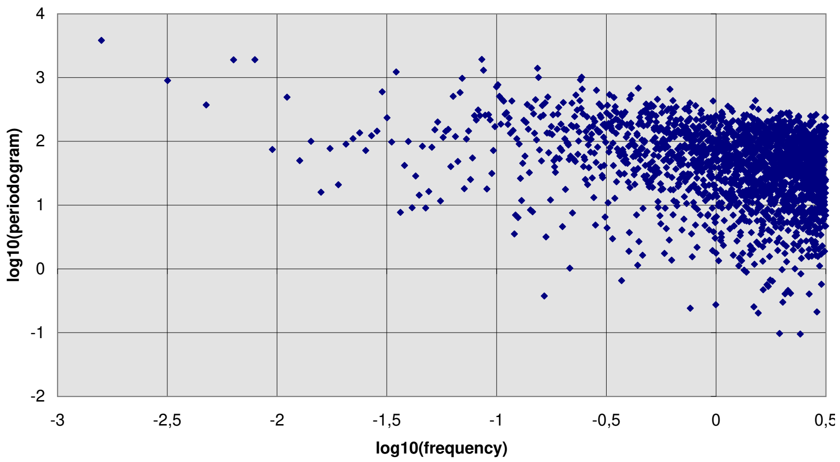
Let $I(\cdot)$ denote the periodogram of X defined by

$$I(\lambda) = \frac{1}{2\pi N} | \sum_{j=0}^{N-1} X_i e^{ij\lambda} |^2, \qquad k \in [0, \pi)$$

Then the periodogram plot is obtained by plotting $\log(I(\lambda))$ against $\log \lambda$. In the present context, the regression line was fitted to the lowest 50% of all frequencies.

Figure (7.21) presents the periodogram plot for the FUNET1 data set, where the analyzed time series was the number of cells in every 1 milliseconds. The slope of the low frequency part is clearly different from 0, the slope estimate is about -0.36 which yields $H = 0.68$. This result corresponds to the previously calculated values of H.

Figure 7.21: FUNET1 - periodogram plot



Following the never-failing procedure by plotting the periodogram for the Poisson process as well as for the shuffled FUNET1 data, we get Figure (7.22) and (7.23). The estimated values for H are 0.51 and 0.52 which was expected.

Figure 7.22: POIS1 - periodogram plot

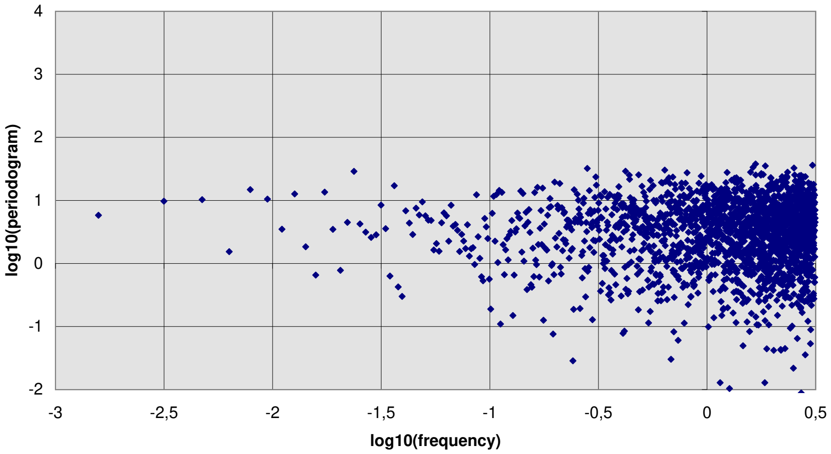
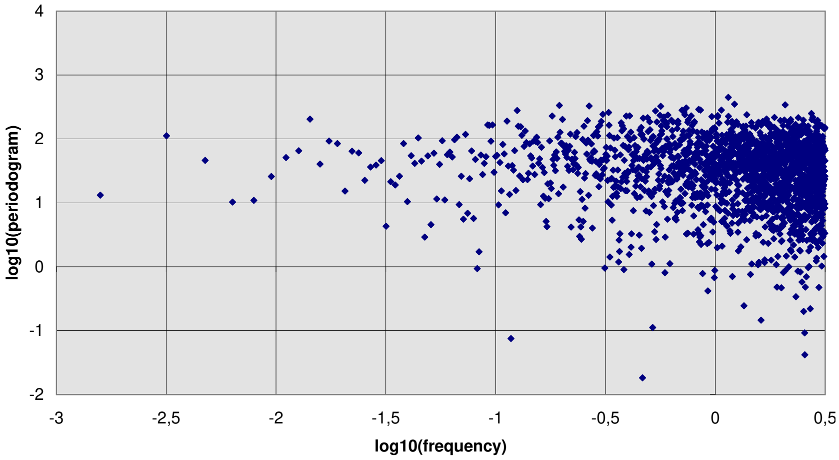


Figure 7.23: Shuffled FUNET1 - periodogram plot



The analysis was made for all the data sets again, and the results are listed in Table (7.4).

For the FUNET1 and FUNET3 data sets the estimated value of H is the same as was before. But for sets FUNET2 and FUNET4 the values of H are a little bit higher than the previous estimates. As for the last two data sets, the same holds as in the case of the R/S statistics. In case of the set FUNETSTA.T3, the regression line was fitted to as few as 98 points, so the inaccuracy could be very high.

Filename	slope γ_{idc}	Estimate of $\hat{H}_{per}$
FUNET1	-0.36	0.68
FUNET2	-0.46	0.73
FUNET3	-0.35	0.68
FUNET4	-0.56	0.78
FUNETSTA.T3	-0.88	0.94
FUNETSTA.T4	-0.79	0.90

Table 7.4: Estimated values of H from periodograms

7.3.5 Summary

Table (7.5) summarizes the estimated values of parameter H for all the data sets and parameter estimation methods (IDC plot, variance-time plot, R/S analysis and periodogram plot) discussed in the previous sections.

Filename	$\hat{H}_{idc}$	$\hat{H}_{var}$	$\hat{H}_{rs}$	$\hat{H}_{per}$
FUNET1	0.69	0.69	0.68	0.68
FUNET2	0.67	0.67	0.67	0.73
FUNET3	0.66	0.66	0.68	0.68
FUNET4	0.72	0.72	0.74	0.78
FUNETSTA.T3	0.70	0.70	0.82	0.94
FUNETSTA.T4	0.67	0.67	0.79	0.90

Table 7.5: Summary of the results

The calculated values for H are pretty much the same, apart from the small differences discussed earlier. But one can conclude, that for the FUNET measurements the estimated value of H is about 0.7 for all the data sets, which is different for the value 0.5 and this yields for the feature of self-similarity present in the traffic traces.

To estimate the self-similarity parameter H more precisely, more refined statistical methods with confidence intervals for H and longer measurements with more data are needed. Here, I only presented an overview of different statistical methods used for testing for the presence of self-similarity in traffic traces. Using these algorithms to analyze data sets from real ATM measurements, I showed that the traffic traces do exhibit special features which are connected to self-similar processes rather than processes used in classical traffic modeling approaches.

Chapter 8

Modeling

"The purpose of models is not to fit the data
but to sharpen the question."

Samuel Karlin (1923–)

The first half of this chapter describes the synthetic traffic trace generation methods used during the modeling. Three self-similar traffic models were chosen to do this. The modeled traffic measurement data set was the FUNET1 data set—all the models were fitted to this file.

The second part of the chapter gives the statistical analysis results for these synthetic traffic traces. The whole methodology and the statistical tests used are the same as were in case of the analysis of the measurement data in the previous chapter. During the evaluation of the results, the goodness of the models are measured by how close the generated data sets are to the measured traffic trace from statistical point of view.

8.1 Synthetic Traffic Traces

In this section I give an overview of the modeling techniques used to generate self-similar synthetic traffic traces. The aim of modeling here is to create data sets by the models and fit the data to real traffic measurements.

The generation method was the following: at first, I analyzed the real traffic measurement file and calculated the statistical parameters needed to set the model parameters. After that, the generation method was done, and finally I made some post-processing on the resulting data set when it was necessary. The original data set for modeling was the FUNET1 data set. Besides the mean rate of the measured traffic and the length of the measurement time, the most important parameter was the self-similarity parameter H.

I generated four different synthetic traffic traces using four different traffic models. The first one is a simple Poisson process. The second one is for generating Fractional Gaussian Noise process based on the Fourier transformation method. The next model generates traffic traces using deterministic chaotic maps, and the last modeling approach is used for generating self-similar traffic sequence by multiplexing many simple ON/OFF processes.

8.1.1 Poisson Process

We can say, that the traffic source model based on a simple Poisson process [1] is the most widespread traffic model in teletraffic theory. The successfulness of the model lies in its simplicity. As regards accuracy, it is a mathematical fact, that the superposition of a large number of independent renewal processes, even periodic ones, converges locally to a Poisson process. Thus, at the cell time scale, a Poisson process, conditioned on the local arrival rate which in fact is a more slowly changing random process, is often a good model for the aggregated traffic [3]. (Note, that the assumption of independence of traffic streams is not always hold.)

As for the self-similarity, the simple Poisson process has a self-similarity parameter $H = 0.5$. In our case, the measured traffic traces possess long-range dependence and self-similarity parameter $H \approx 0.7$ so the Poisson model as a model for aggregated ATM traffic is not appropriate here. But it is useful to compare the results obtained from different traffic models with the results obtained from the Poisson process, which helps to reveal the mean features of self-similarity and the differences between classical traffic models and the self-similar models discussed in this thesis.

To generate a synthetic traffic trace according to a Poisson process is very simple. The task is to generate exponentially distributed cell interarrival times with parameter λ equal to the reciprocal of the expected mean value for that process and rebuild the cell sequence. In our case, the generated Poisson process was fitted to the FUNET1 data set and is referred to as the 'POIS1' traffic trace.

8.1.2 FFT-FGN Method

In this section a Discrete Time Fourier Transformation (DTFT) based method is presented for synthesizing fractional Gaussian noise (FGN). (Because the

[1] Appendix D.1 gives the defining properties of a Poisson process.

DTFT and its inverse can be rapidly computed using the Fast Fourier Transformation (FFT) algorithm, one can refer to this method as the FFT-FGN method.) The strategy behind is taken from [10], and can be summarized as follows [27].

Suppose we know $f(\lambda; H)$, the power spectrum of the FGN process. Then we can construct a sequence of complex numbers z_i corresponding to the power spectrum; z_i is in a sense a frequency-domain sample path. We can then use an inverse-DTFT to obtain x_i, the time-domain counterpart to z_i. Because x_i has the power spectrum of FGN, and because autocorrelation and power spectrum form a Fourier pair, x_i is guaranteed to have the autocorrelational properties of an FGN process, which for many purposes are its most salient characteristic.

The difficulty with this approach lies in accurately computing $f(\lambda; H)$, and in finding z_i truly corresponding to the FGN power spectrum.

Note, that it is not proven that the method results in true FGN, indeed, it does not, due to several approximations made when developing the method. But it *effectively* produces FGN, which means that the sample paths produced by the method are indistinguishable (using current statistical tests) from true FGN. (According to the author of [27], one can refer to this approach as the "quacks like a duck" approach, from the adage that if an object looks like a duck, walks like a duck, and quacks like a duck, one might as well call it a duck.)

The Appendix in [27] gives a program written in the *S* language for implementing the above method. The *C* version of that algorithm was implemented by C. Schuler at the Research Institute for Open Communication Systems in Berlin, Germany. During the modeling phase I used the v1.1 version of that program. (The *S* program and also the *C* version of the same algorithm is accessible from the Internet [39].)

To generate a cell sequence with this program, first I had to calculate the input parameters for the process. The necessary input parameters for the model are the Hurst-parameter H, the output mean value M, the output variance V and the number of sample points n. Besides these input parameters, the command line switch '-m6' was also used to generate integer output.

When the initial parameters were set, the program generates the self-similar sample path. After the generation two further question arise. The first one is that with certain probability the sample path contains negative values also. In case of a traffic process the notion 'negative traffic' is meaningless (or at least in most cases). To solve this problem, I simply replaced all the negative numbers with zeros in the process. But than another question arise: whether or not this conversion affects the statistical properties of the process. (The mean value increases certainly, but we are interested mainly in the self-similar property of the sample path.) In our case, about 8 % of all the values were negative.

The next question is how to convert the sample path to inter-arrival times. The answer comes from the author of the program: interpret the obtained values as cell counts during a constant 'bin' period, say 100μs. Then calculate the inter-arrival time by IAT = (100μs)/(cell count). As a result, the traffic generated by this method changes its datarate in every 100μs. Choosing the appropriate 'bin' period the conversion method does not affect the long-range effects present in the traffic trace so the self-similar structure remains the same.

To fit the generated data to the FUNET1 data set, I calculated the mean rate and the variance for the cell count process for the original data, where the 'bin' period was 100μs. Because of the conversion method discussed earlier affected the mean rate and the variance of the generated traffic trace, the input values of M and V were set to be higher. The calculated input parameters were: $M = 3$ and $V = 30$. As for the Hurst parameter H I decided to use the value 0.75 instead of 0.7 which was the estimated self-similarity parameter for the FUNET1 file. (The reason why I did this was that I used the same input parameter H for all the traffic models used during the modeling, and for some reason the appropriate value turned out to be 0.75 in case of the other two models. I shall explain this later on in the next section.)

The name of the result file is 'G1-75'.

8.1.3 Superposed ON/OFF Sources

The theoretical background of the self-similar traffic model based on the superposition of ON/OFF sources was presented in Section 5.3. A synthetic traffic generator program based on this approach was implemented by O. Matellanes at Dept. of Computer Science, Uppsala University, Sweden in 1995. The name of the program is 'Self-Similar Traffic Generator' and the C code for that product was accessible for me during my work.

The program is able to generate synthetic traffic traces using two self-similar modeling approaches. The first traffic model implemented is the one discussed in this section, and the other one is based on deterministic chaotic maps. (I shall discuss that second model in the next section.)

Again, recall the main idea behind this approach and the way it is implemented in the traffic generator. Consider N traffic sources that are generating cells independently from each other and multiplex the emitted cell streams together. The rules describing the characteristics of each source are the following:

- A source is in an active state in time interval T_{ON} and after this it remains silent for time T_{OFF} (the source is in a passive state).
- During the T_{ON} periods, the source transmits cells according to the Poisson distribution (with parameter λ_i). (This means that the distribution of the cell inter-arrival times is exponential.)
- The distributions for periods T_{ON} and T_{OFF} follow the Pareto-type [2] (or power-law) distribution.

Thus, every traffic source can be viewed as a simple Poisson-source modulated by a two state, Pareto-type process. (Again, the main feature of a Pareto-type distribution is its hyperbolic tail.)

In theory, the superposition of infinite number of traffic sources satisfying the conditions stated above results in a self-similar traffic process. In practice, the number of sources N should be big enough (i.e., $N = 1000$). For the peak rate (when all of the sources are in the active state) to be λ for the aggregated traffic stream, one should set the parameters $\lambda_i = \lambda/N$.

The model parameters for the self-similarity parameter $H = 0.75$ were: $\alpha_{ON} = 1$, $\beta_{ON} = 0.43$, $\alpha_{OFF} = 400$, $\beta_{OFF} = 1.9$ and $\lambda = 1000/N$. (The idea why I set the parameter H to 0.75 instead of 0.7 came from [14] where the traffic generator under discussion was analyzed extensively. The results from [14] imply that the self-similarity parameter of the sample path is usually lower than the input parameter H. For example, for the input value 0.75 the expected value of the Hurst parameter of the output process is about 0.7. That is why I chose that value as an input parameter for all the traffic models. According to my results, it turned out to be over-caution.) The program generated the sequence of cell arrival times for 131,072 cells. To fit the model to the FUNET1 data set, I only had to modify the time stamps by scaling the values to set the last timestamp to be equal to the length of the FUNET1 measurement. (This modification does not disturb the self-similar property of the process because the self-similar structure does not depend on the time scale.) As a result, I got the data set— referred to as the ON1-75 data—which had exactly the same length and the same number of time stamps FUNET1 had.

[2]See Appendix D.2.

8.1.4 Chaotic Maps

The full description of the modeling approach can be found in Section 5.4. To recall the underlying idea, the definition of the traffic source model referred to as *intermittency map* is repeated here briefly.

Consider a nonlinear map $f(\cdot)$ that describes the evolution of a state variable $x \in (0,1)$ over discrete time as $x_{n+1} = f(x_n)$. The packet generation process for an individual source can now be modeled by stipulating that the source generates one or no packet at time n depending on whether x_n is above or below an appropriately chosen threshold. The defining equations for the intermittency map traffic model are the following:

$$x_{n+1} = \begin{cases} \varepsilon + x_n + cx_n^m & \text{if } 0 < x_n \leq d \\ \frac{x_n - d}{1-d} & \text{if } d < x_n < 1 \end{cases} \qquad \text{where} \quad c = \frac{1 - \varepsilon - d}{d^m}$$

As can be seen from the equation the model has only three parameters: ε, c and m.

To generate a synthetic traffic trace using this model, I used the Self-Similar Traffic Generator program again. The input parameters for the generation process were only the Hurst-parameter H and the number of cells generated. For the parameter $H = 0.75$ the implemented algorithm used the parameter values $\varepsilon = 10^{-10}$, $d = 0.8$ and $m = 1.73$ as pre-defined values for the map. The number of cells generated was set to be equal to the number of cells captured during the FUNET1 measurement period. Again, the time stamps were modified according to the length of the original measurement. In the following, I refer to that data set as the 'CH1-75' set.

8.2 Analysis and Verification

As the title implies, the aim of this section is to present the analysis results for the synthetic traffic traces generated by the above mentioned traffic models. The whole methodology and the statistical algorithms used are the same as were in case of the analysis of the FUNET measurement data sets.

The goodness of the models are measured by how close the generated data sets are to the measurement file as far as certain statistical properties concerned. First of all, we expect that the data sets possess the self-similar property and have nearly the same statistical behaviour (i.e., correlation structure, histograms, distributions) as the original traffic trace. The word verification in the title means to check whether these assumptions are hold or not.

Note again, that in the following—just for simplicity—I refer to the data set generated by the superposed ON/OFF traffic model as the 'ON1-75' file, and the set generated by the intermittency map as the 'CH1-75' file. The name of the last traffic trace generated according to the FFT-FGN method is the 'G1-75' file.

8.2.1 Histograms

Because of the huge amount of figures, only the evaluations of the results can be found in this chapter while all the plots and figures are listed in Appendix C. The histograms calculated for the model files can be found in Appendix C.1. As for the FUNET1 measurement, the chosen time interval was 10ms for the histograms. (Note, that in case of the histograms and all the other plots and charts, the full description of the way the figures were obtained can be found in the previous chapter. Here, I only focus on the results.)

Looking at the pictures, the first impression about all the three diagrams is that the traffic is bursty at this time scale. The plots are clearly different from the one obtained from a Poisson process, and in case of the ON1-75 and G1-75 files, the histograms look pretty much the same as the FUNET1 histogram.

But looking at the CH1-75 plot, one strange thing arise which makes the plot different from the others. Sometimes the silent periods between traffic bursts are extremely long, the longest inactive period is about 0.1 seconds. In case of aggregated ATM traffic on a 155Mbps link loaded at a rate as was when the FUNET measurements were done it is almost impossible. The traffic pattern with such a long idle periods can be the characteristic of a single source but not the aggregated network traffic.

8.2.2 Inter-arrival time distributions

Considering the inter-arrival time distribution of the FUNET1 file, it is clear that none of the models will produce such a plot. The IAT distribution for FUNET1 is so special that only more specialized traffic models could generate such a traffic stream. But the diagrams presented in Appendix C.2 reveal some fundamental characteristics of the models.

Analyzing the cell inter-arrival time distribution for the CH1-75 data set, the shape of the distribution function is very peculiar. The tail of the function decays slowly and it has positive values even for $T = 0.1$s. This result is in full harmony with the result obtained from the histogram of the process. Another interesting thing is that the shape of the plot is not continuous, there are some definite peaks in the diagram for small values of T separated by empty bins. This comes from the structure of the model.

The IAT plot for the set G1-75 is missing from the Appendix. The reason is that the generation process of that traffic trace only gives the cell count sequence, not the cell arrival time sequence. The conversion method for generating the IAT process from the sample path (presented in Section 8.1.2) is artificial, so it is not meaningful to plot the result. Instead, the IAT plot for the POIS1 data set is presented in Appendix C.2.

It is not surprising that the IAT plot for the ON1-75 data set looks nearly the same as the plot obtained from the Poisson process. The generation method for that traffic trace is based on the superposition of many independent ON/OFF sources with Poissonian ON periods. There is only a slight (but maybe fundamental!) difference between the two diagrams, the left side of the ON1-75 function decays more slowly than the other one.

8.2.3 Cell distributions

The distributions of the cell counts for all the data sets were plotted and are presented in Appendix C.3.

At first glance, all the three diagrams look like roughly a bell-shaped curve. But looking at the plots more carefully, the CH1-75 data set turns out to be the cuckoo-egg again. The first column of the histogram is about three times higher than the second highest value. This means that the number of cells received in 0.1s is equal to 1 more that 300 times during the process. This also implies that sometimes there is a huge gap between two consecutive cells.

Finally, it is worth to compare all the plots with the plot obtained from the FUNET1 measurement (see Figure (7.8) on page 47). As far as the cell distribution function concerned, the G1-75 data set models the FUNET1 process quite accurately.

8.2.4 Autocorrelation functions

The autocorrelation structure of the process is important in our discussion because the long-range correlations present in the data set are responsible for the self-similar property. Appendix C.4 gives the autocorrelation function plots for the synthetic traffic traces.

Comparing the results with the autocorrelation function of the FUNET1 data series, there is one thing in common. In all four cases the plotted curves possess the long-tailed property. In other words, there are still positive correlations between elements relatively far from each other. (The 'relatively far' here means about $k = 500$ as can be seen from the plots.)

The right side of the plot represents the short range correlations in the process. Again, comparing the diagrams with the FUNET1 autocorrelation function, it can be seen, that the short-range dependence in the real traffic pattern is much stronger than in the synthetic traffic processes. But the short-range dependence does not affect the self-similar feature of the actual process.

Appendix C.5 shows the log-log versions of the diagrams discussed above. For the values $\log k < 2.5$, the trends for $\log r(k)$ are clearly visible from the diagrams. The function $r(k)$ scales as $k^{-\beta_{auto}}$ which is in stark contrast to the conventional short-range dependent processes, that are characterized by an exponential decay of the autocorrelations.

8.2.5 IDC plots and Variance-time plots

After the preliminary basic statistical test described above, our aim is to estimate the parameter standing in the limelight, namely, the self-similarity parameter H of the traffic traces. The first two statistical test discussed here are the variance-time plots and the IDC plots. As I mentioned earlier, the two methods are closely related to each other and give the same information about the analyzed process. Since the results obtained from both methods are the same, there is no use to discuss them separately.

Appendix C.6 presents the IDC plots, and Appendix C.7 shows the variance-time plots calculated from the ON1-75, CH1-75 and G1-75 data sets. The resulting diagrams seem to be nearly ideal. For all the plots, the linear parts are clearly visible and it is easy to fit the regression line to the points and to calculate its slope. Table (8.1) list the calculates values for the slope of the regression line and the estimate of parameter H for all three cases.

Filename	slope γ_{idc}	slope β_{var}	estimate of $\hat{H}$
ON1-75	0.47	-0.53	0.73
CH1-75	0.60	-0.40	0.80
G1-75	0.43	-0.57	0.71

Table 8.1: Estimated values of H from IDC and variance-time plots

The Hurst-parameter for all the models was set to be 0.75. As for the ON/OFF source model, the calculated value for H is 0.73 which is nearly the same as the expected value. But as for the last two traffic models, the difference between the calculated and the expected values are considerable.

Before the concluding remarks, finish our tests and examine the results obtained from the next two statistical tests, namely, the R/S statistics and the periodogram plots.

8.2.6 R/S-analysis

The R/S plots for the synthetic data sets can be found in Appendix C.8.

The diagrams for the ON1-75 and G1-75 data sets are nearly identical. The street of the sample points is straight and its thickness is uniform through the whole interval for d. But the second plot is a little bit different again. The right end of the plot becomes noisy for the values $100 < d$ and the street of the points stretches out. Apart from this, there are some points lying on the horizontal axis, representing the events where the calculated R/S values were infinite for the given d interval. This behaviour affected the line fitting method to the diagram also. Table (8.2) list the calculated values of $\hat{H}$ from the R/S diagrams.

Filename	estimate of $\hat{H}_{rs}$
ON1-75	0.75
CH1-75	0.75
G1-75	0.71

Table 8.2: Estimate of parameter $\hat{H}$ from R/S plots

For the ON1-75 data set the calculated value is 0.75, which was the input self-similarity parameter for the model. The value of $\hat{H}$ of the G1-75 trace is lower than the expected 0.75, but the results is in good harmony with the values estimated previously using the IDC and variance-time plots. In spite of the strange effects present in the CH1-75 plot, the estimate for $\hat{H}$ is 0.75.

8.2.7 Periodogram plots

And finally, the periodogram plots for the ON1-75, CH1-75 and G1-75 data sets are presented in Appendix C.9.

Looking at the diagrams one thing is evident. The shape of the point-clouds clearly indicate the presence of long-range dependence in every cases. The slopes of the fitted lines are different from zero. The calculated values and the estimates of $\hat{H}$ can be found in Table (8.3).

The values of $\hat{H}$ for the ON1-75 and CH1-75 sets are higher than before. The value of 0.83 for CH1-75 is so high that I tried to fit the regression line to different parts of the diagram. The calculated values were considerably different in each case, even values higher than 0.9 were calculated sometimes. But note again, that plotting the periodogram and estimating its slope would lead to a wrong estimate of H since the periodogram is not appropriate to estimate the spectral density.

But in case of the G1-75 data set, the estimate of $\hat{H}$ is 0.72 and that is the same as it was previously.

8.2.8 Summary

At the end of this chapter, the task is to conclude whether or not the traffic models were able to model the real measurement file (FUNET1) in a satisfactory way. The three traffic models were the aggregated ON/OFF source model (ON1-75), the chaotic map model using the intermittency map (CH1-75), and the self-similar traffic model based on the fractional Gaussian noise process (G1-75).

To conclude the results obtained from the four different statistical methods used, Table (8.4) summarizes the calculated values of H for all the data sets.

As for the first traffic model based on the superposition of independent ON/OFF sources, we can say that the generated synthetic traffic trace—as far as the above analyzed statistical properties concerned—is a good approximation of the modeled traffic. The synthetic traffic trace possesses the

Filename	slope δ_{per}	Estimate of $\hat{H}_{per}$
ON1-75	-0.62	0.80
CH1-75	-0.67	0.83
G1-75	-0.45	0.72

Table 8.3: Estimated values of $\hat{H}$ from periodograms

Filename	$\hat{H}_{idc}$	$\hat{H}_{var}$	$\hat{H}_{rs}$	$\hat{H}_{per}$
ON1-75	0.73	0.73	0.75	0.80
CH1-75	0.80	0.80	0.75	0.83
G1-75	0.71	0.71	0.71	0.72

Table 8.4: Summary of the test results for estimating the Hurst-parameter

self-similar feature, and the calculated self-similarity parameter H (apart from the periodogram) is nearly the same as was the model input parameter. Further, the qualitative results derived from the diagrams of the basic statistical part are consistent with the ones obtained from the real traffic data set.

The chaotic map model possesses the self-similar property also, but the other statistical features are different from the expected ones. Informally, the generated traffic trace looks different than the original one. The model could be a good source model for a single traffic source emitting self-similar cell streams, but not appropriate to model aggregated network traffic.

The most promising modeling approach in our case is the FGN traffic model. The synthetic traffic trace generated by the FFT-FGN method has nearly the same basic statistical features than the measured traffic trace, and the self-similar property of the generated traffic is clear. The calculated value of the Hurst-parameter is 0.71 for the first three statistical methods, and even the periodogram plot gives nearly the same result. The value 0.71 is smaller than the input value 0.75 for the model. But remember, that the conversion method to turn a fractional Gaussian noise process to a traffic process required two approximations. At first, the real values of the sequence were rounded to integers, and next, the negative values were eliminated in the process. These steps could cause the difference between the calculated and expected values of H.

As a concluding remark, I would like to emphasize once more that we cannot say that a particular traffic model is good or not on its own. We can only say that a 'good' traffic model should possess certain features that are important for us. For example, the goodness of a traffic model can be measured by generating synthetic traffic traces and putting these cell streams into a running network or network element (i.e., queues, multiplexers, etc.). If everything behaves like in case of real traffic processes (i.e., the queue length distributions and waiting times remain unchanged), the model is said to be appropriate.

In the present context, the main feature was the self-similarity present in the process. The conclusions stated above should be interpreted according to this.

Chapter 9

Conclusions

"A theory has only the alternative
of being right or wrong.
A model has a third possibility:
it may be right but irrelevant."

Manfred Eigen (1927–)

Modeling of ATM traffic is a significant research topic in the design and performance evaluation of future telecommunication networks. Careful statistical analysis of a large number of traffic traces from a wide variety of working packet-based networks indicate traffic variations over a wide range of time scales. These variations can be described using the concept of self-similarity. Due to the significance of the topic of self-similar traffic modeling, there is an extensive related literature, while important research is still in progress. At first, an overview about the recent findings based on the literature was presented.

The first part of the thesis described the modeling and analysis techniques of self-similar processes. As for the analysis methods, the question was, how could we detect the self-similarity property present in actual network traffic. To answer this, the statistical methods used for testing for self-similarity were discussed. Besides, after a brief mathematical overview, I have summarized the self-similar modeling techniques for ATM network traffic.

The second part of the thesis presented my investigations on the FUNET ATM network. Measurements were made on this network, and the analysis of the measurements were done. For the self-similarity tests, four different graphical statistical test out of the analysis methods discussed earlier were used during the study, the so-called IDC plot, the variance-time plot, the R/S statistic, and an analysis method based on periodograms. The results show, that the measured tarffic traces do possess the self-similarity property characterized with self-similarity parameter H about 0.7.

The next task was to model the measured traffic processes. To do this, I have dealt with three different self-similar traffic models in more details. The examined traffic models were:

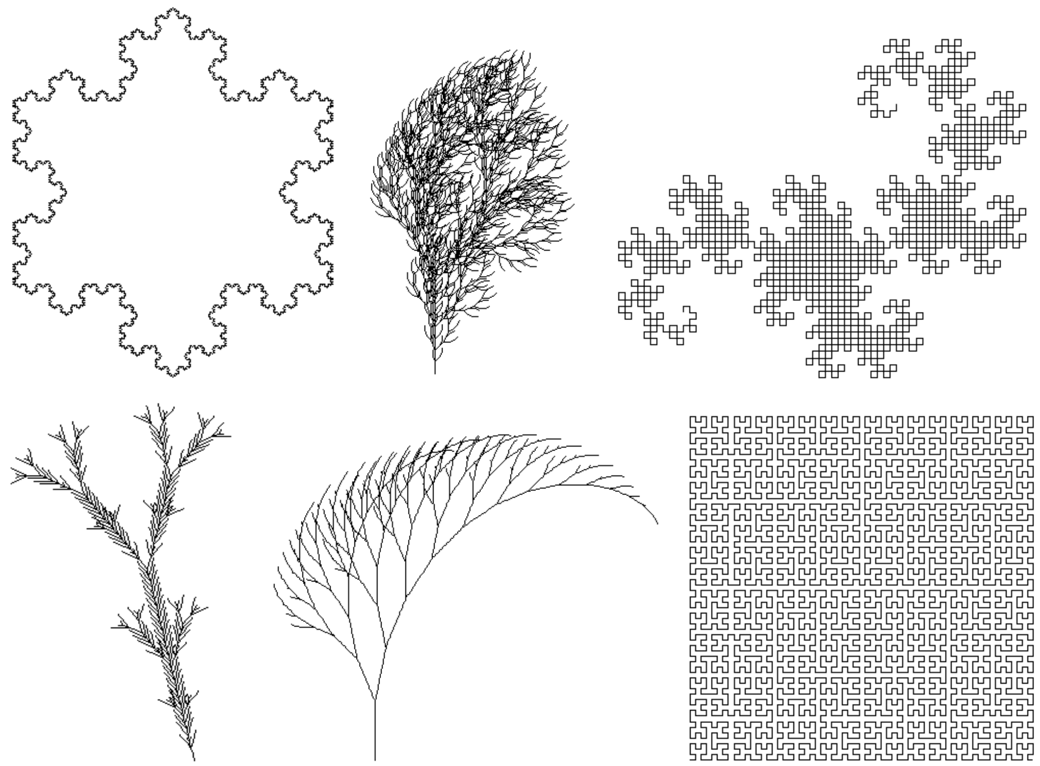
- an exactly self-similar stochastic model based on fractional Gaussian noise processes,
- a deterministic traffic model based on chaotic maps,
- a traffic model based on the superposition of independent ON/OFF sources, where all the sources possess the so-called long-tailedness property for the ON periods.

The appropriateness of a model depends heavily upon the matching of the self-similarity parameter to that measured experimentally. As already mentioned, all the three models were able to generate self-similar traffic. But as for the basic statistical properties of the synthetic traffic traces, there were considerable differences between the different traffic models. After the statistical analysis made for each of the synthetic traffic traces, the following conclusions can be drawn:

- The most promising modeling approach in our case is the FGN traffic model. The traffic trace generated by the FFT-FGN method has nearly the same basic statistical features than the measured traffic trace, and the self-similar property of the generated traffic is clear. The output of a single source is usually very far from a Gaussian process. Fractional Gaussian traffic is a rather generally applicable model for traffic aggregated from a large number of independent sources with peak rates are sufficiently low in comparison to the aggregated mean rate.
- As for the traffic model based on the superposition of ON/OFF sources, the generated synthetic traffic trace is a good approximation of the modeled traffic also.
- The chaotic map model can produce self-similar traffic, but the other statistical features of the generated process are different from the expected ones. The intermittency map model could be a good model for a single traffic source, but it is not appropriate to model aggregated network traffic.

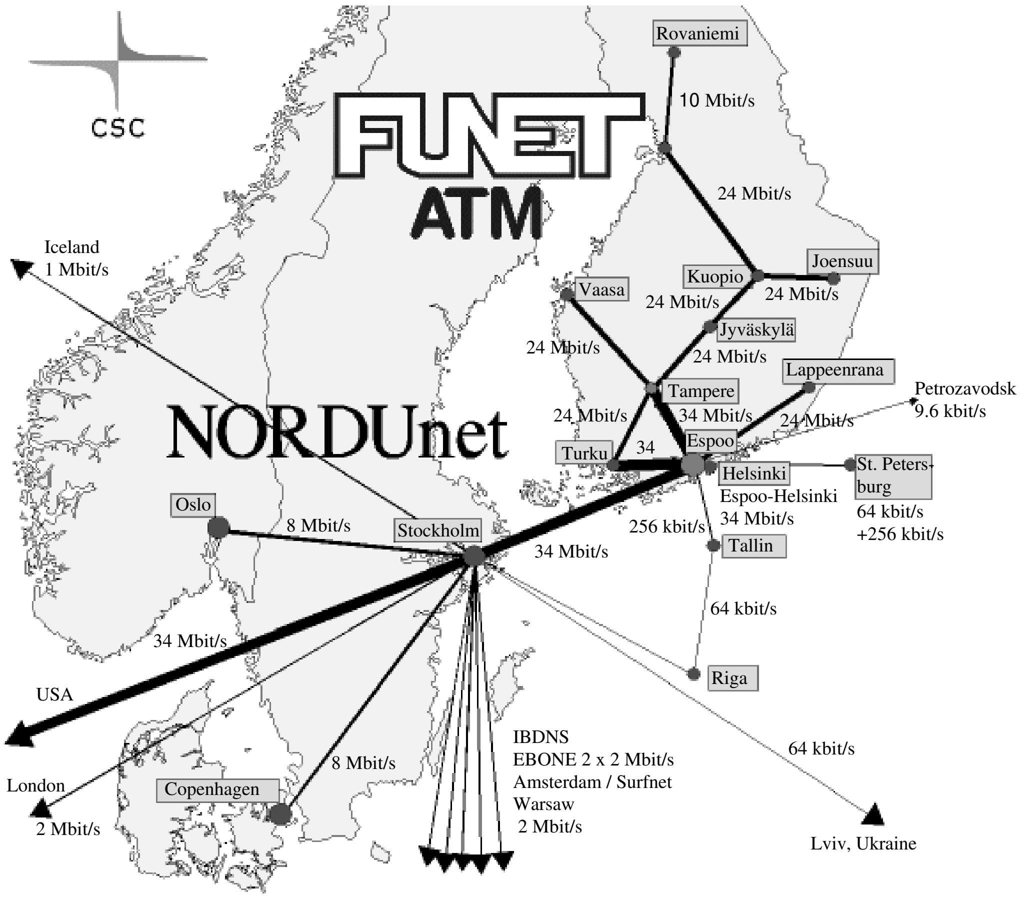
Appendix A

Self-similar point sets



Appendix B

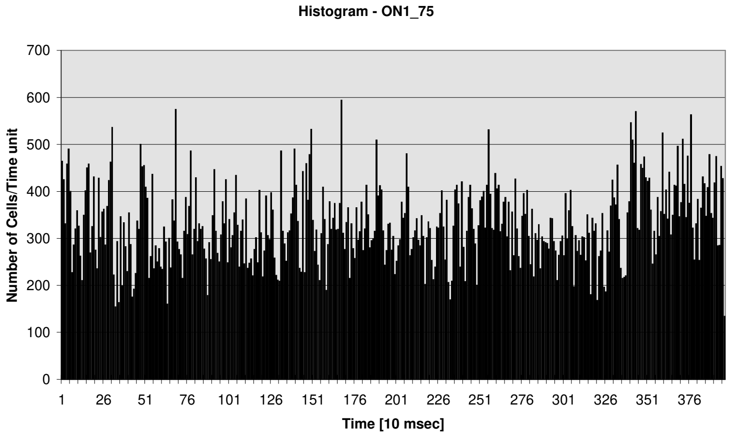
The FUNET Network

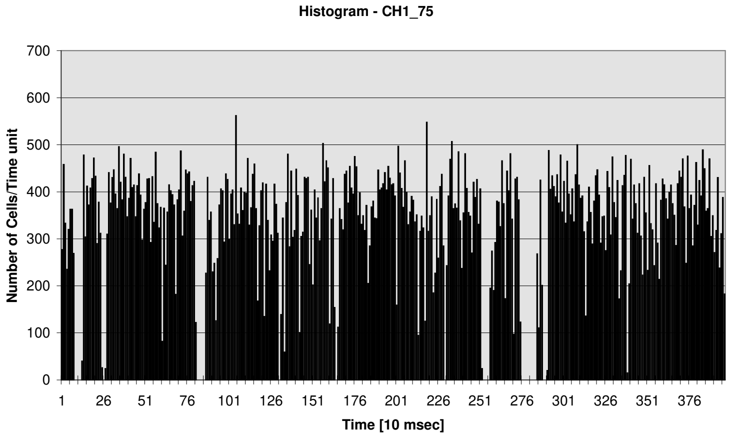


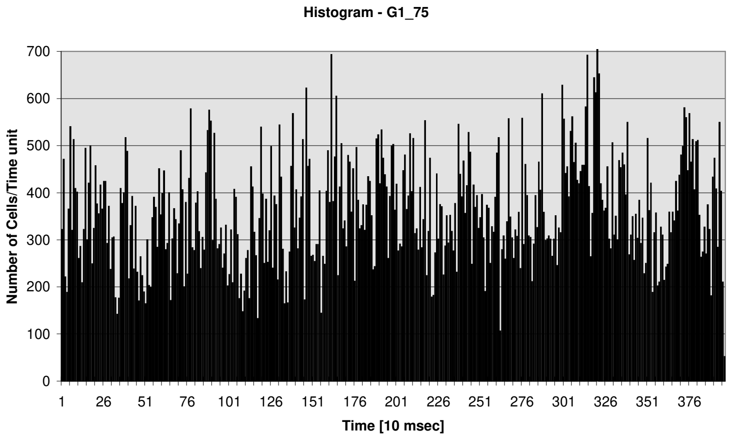
Appendix C

Test Results

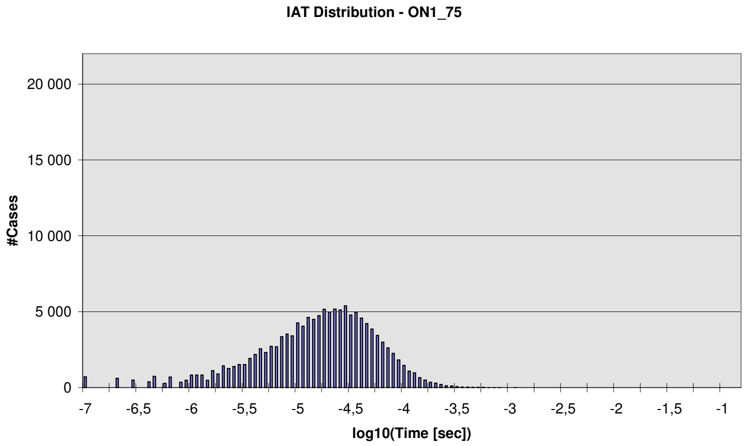
C.1 Histograms

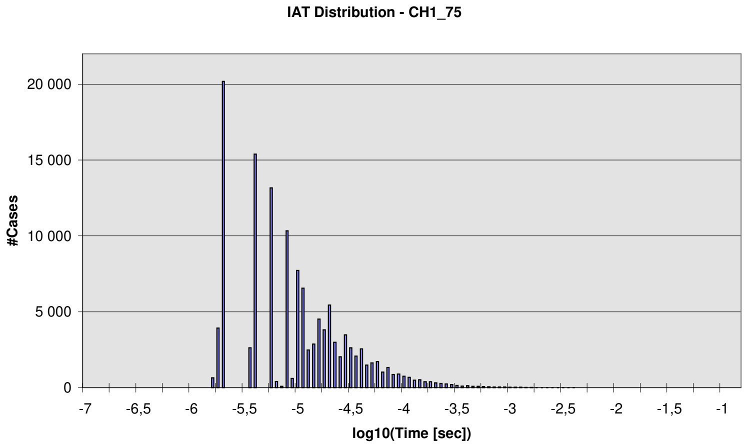


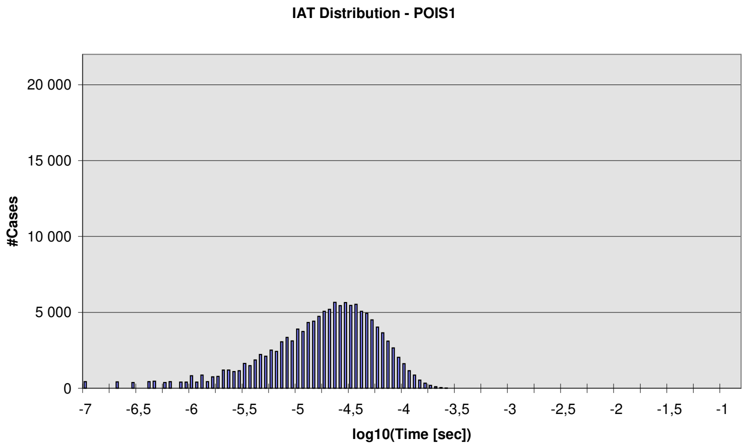




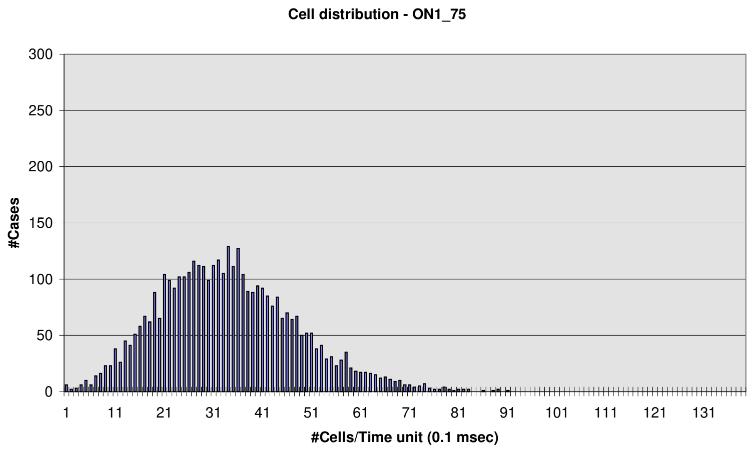
C.2 Inter-Arrival Time Distributions

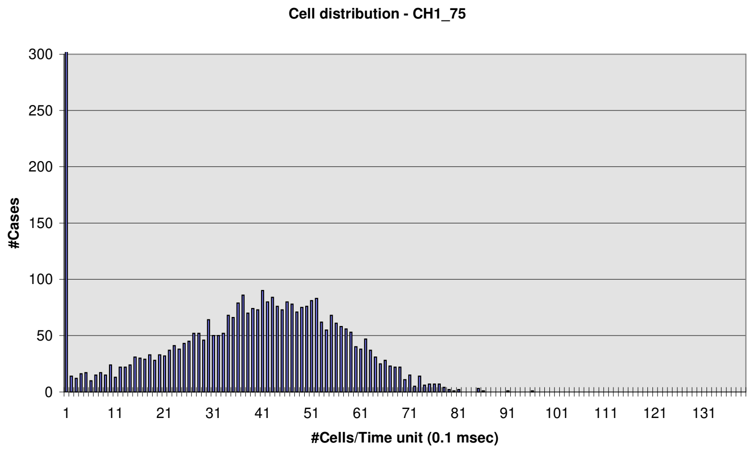


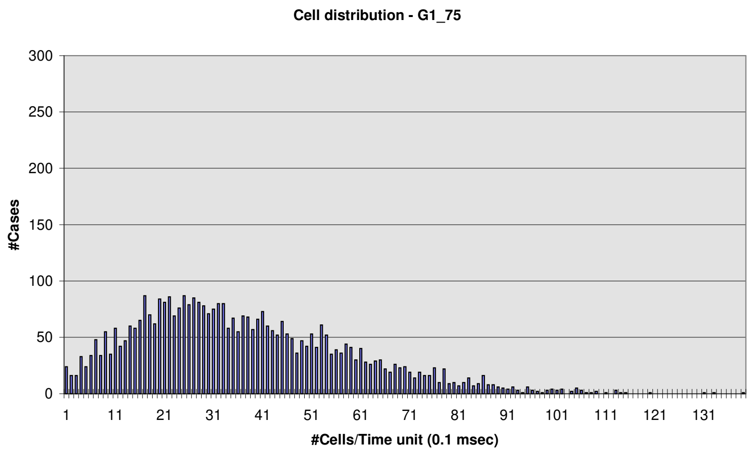




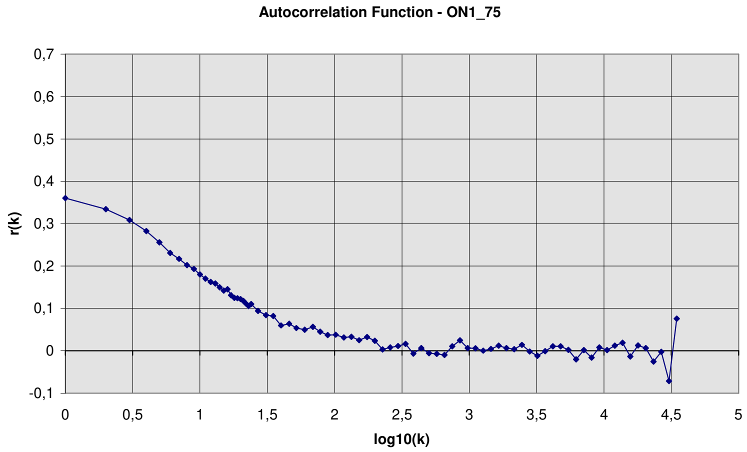
C.3 Cell Distributions

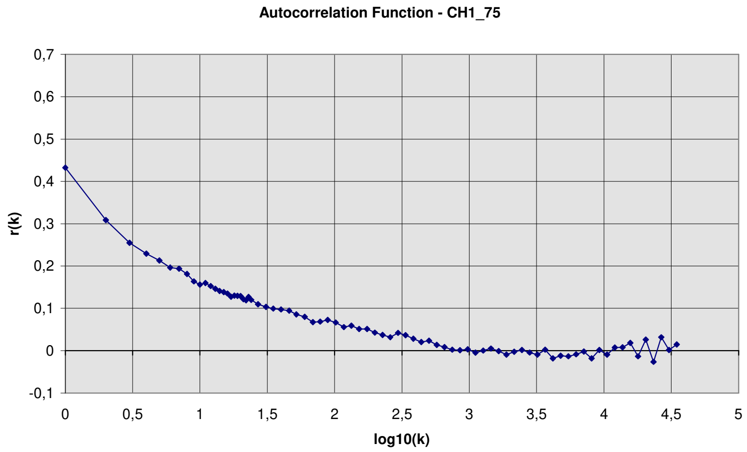


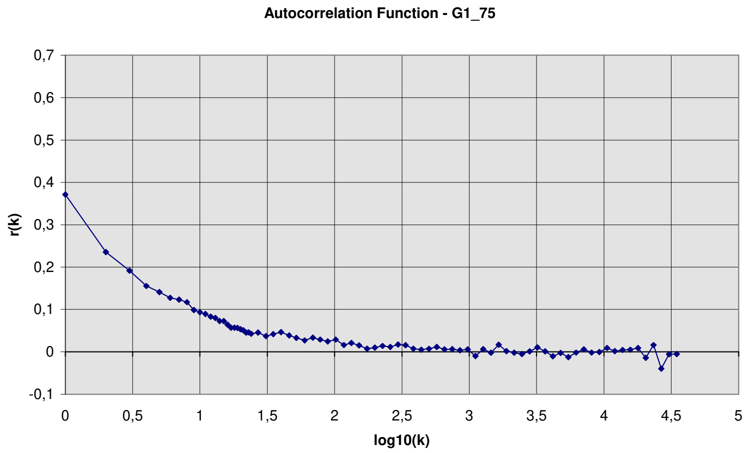




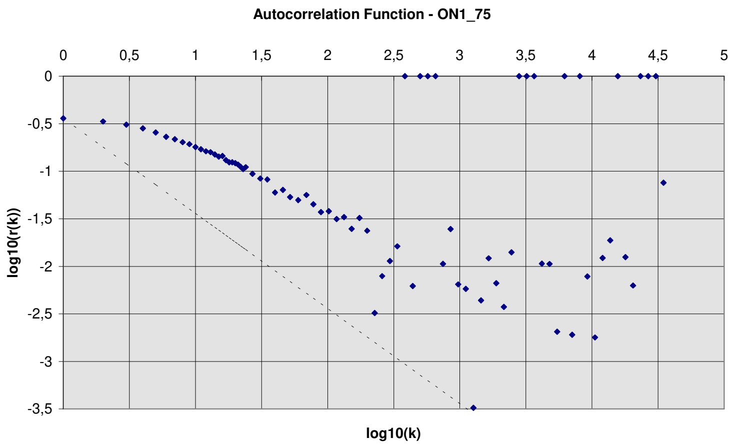
C.4 Autocorrelation Functions

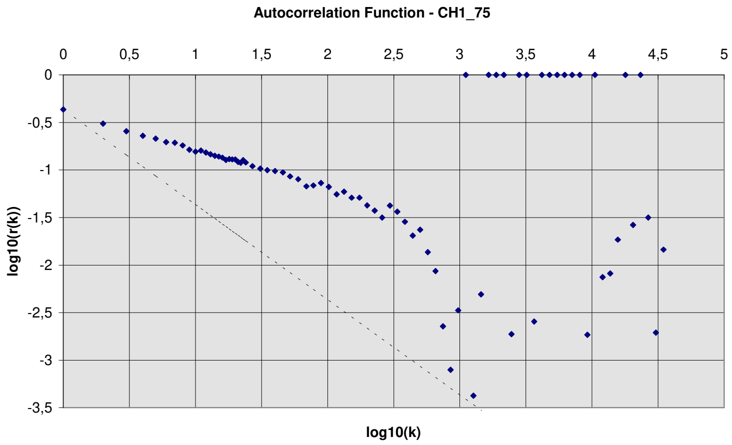


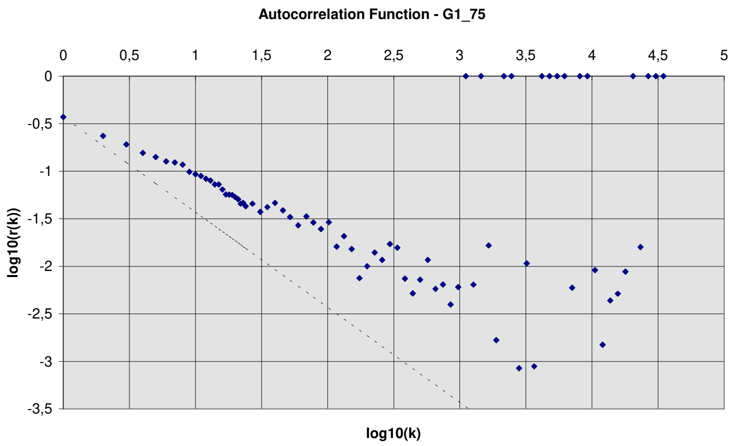




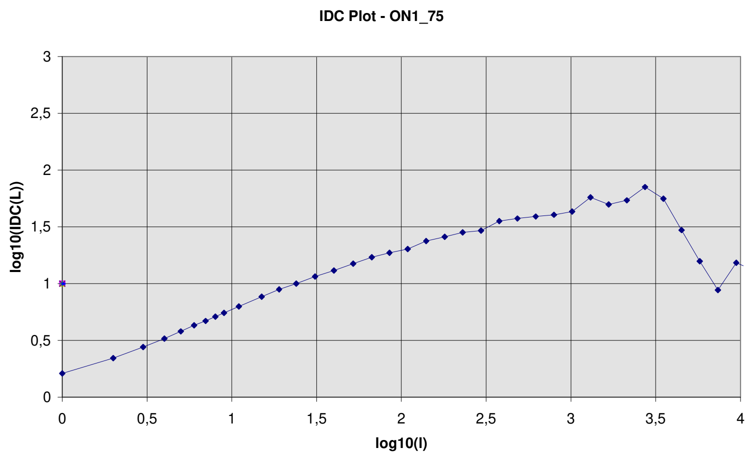
C.5 Autocorrelation Functions (log-log Plot)

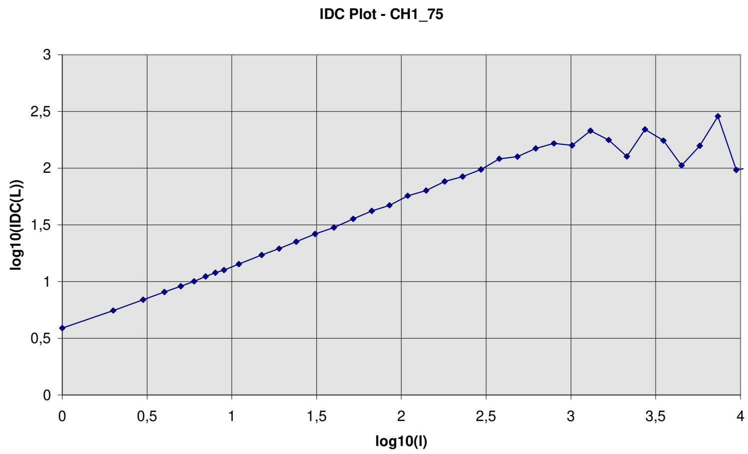


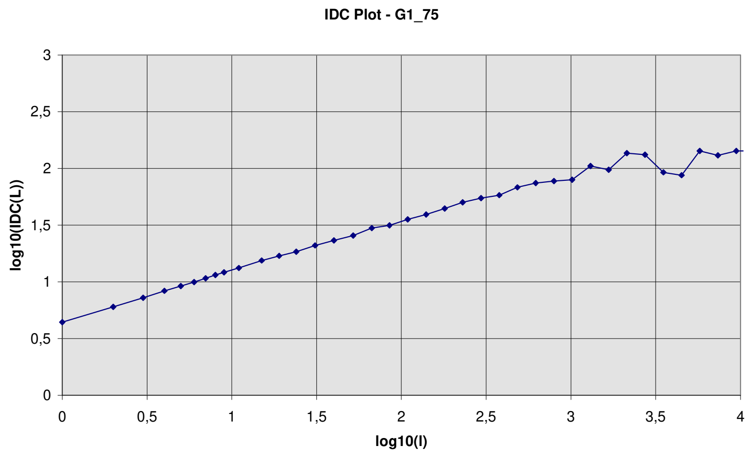




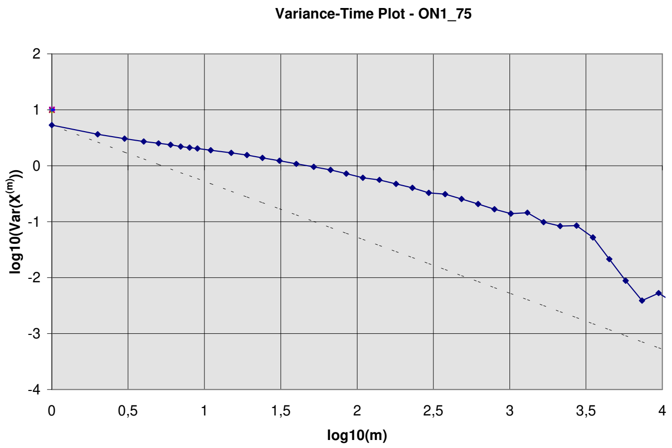
C.6 IDC Plots

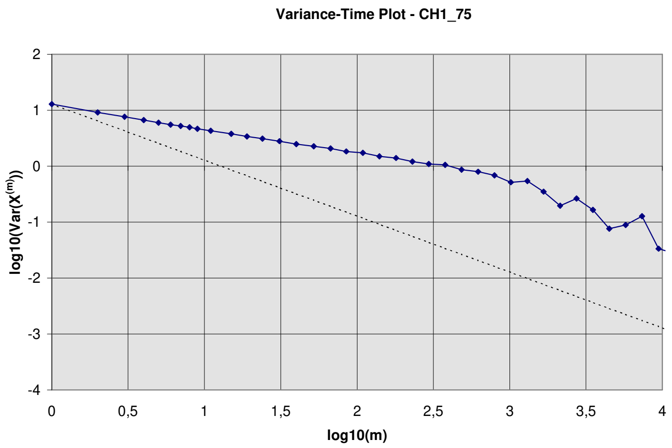

IDC Plot - ON1_75
log10(IDC(L))
log10(l)
0
0,5
1
1,5
2
2,5
3
3,5
4

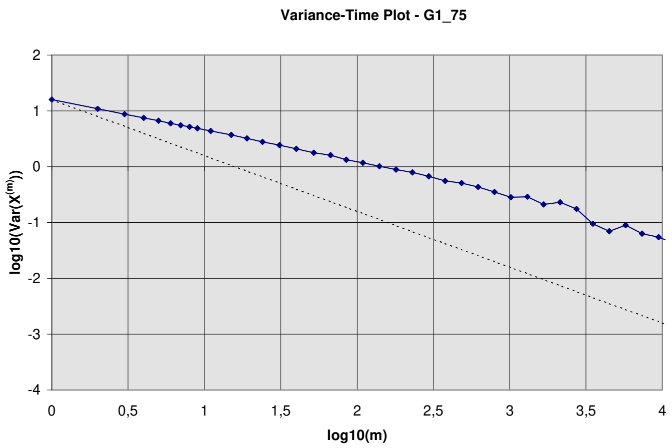

IDC Plot - CH1_75
log10(IDC(L))
log10(l)
0
0,5
1
1,5
2
2,5
3
3,5
4


IDC Plot - G1_75
log10(IDC(L))
log10(l)
0
0,5
1
1,5
2
2,5
3
3,5
4

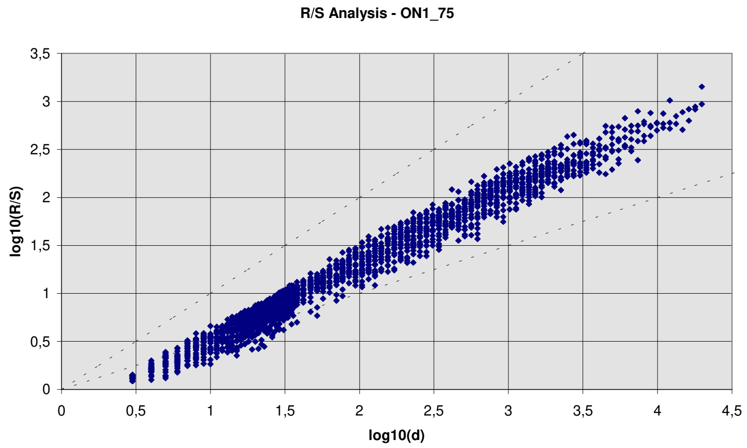
C.7 Variance-Time Plots

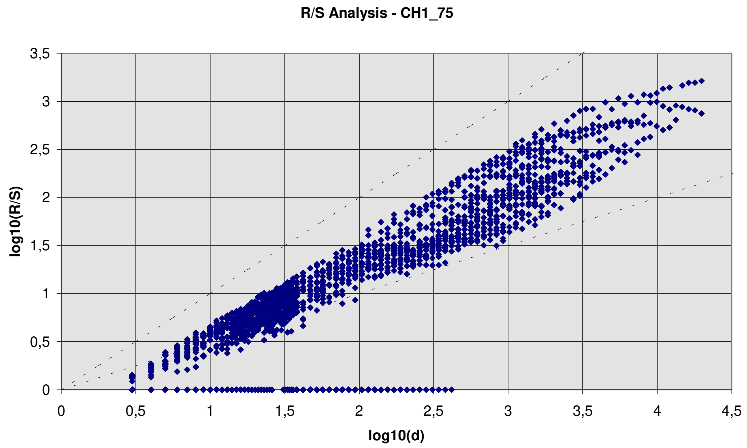


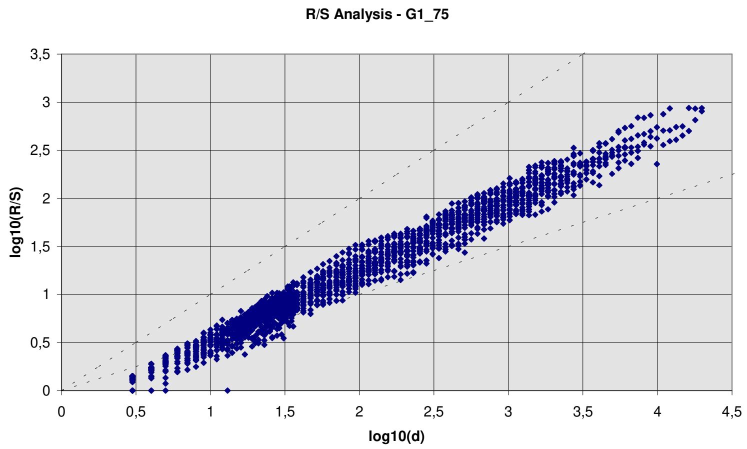




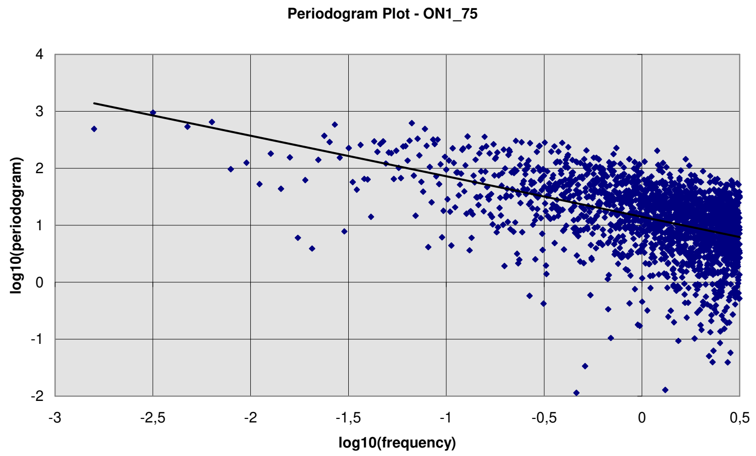
C.8 R/S Analysis

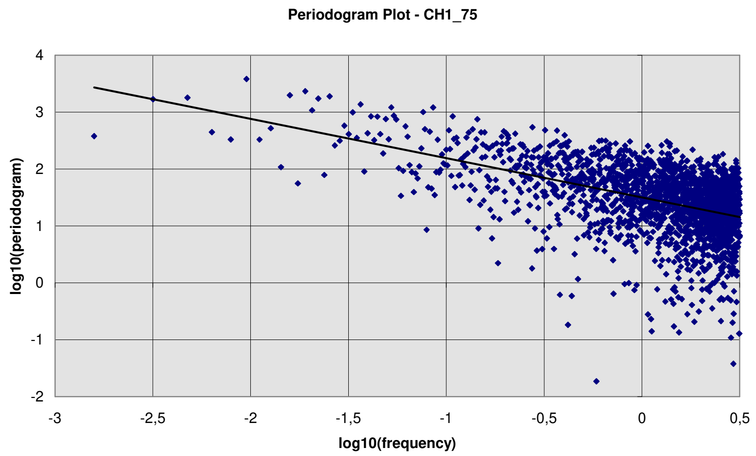


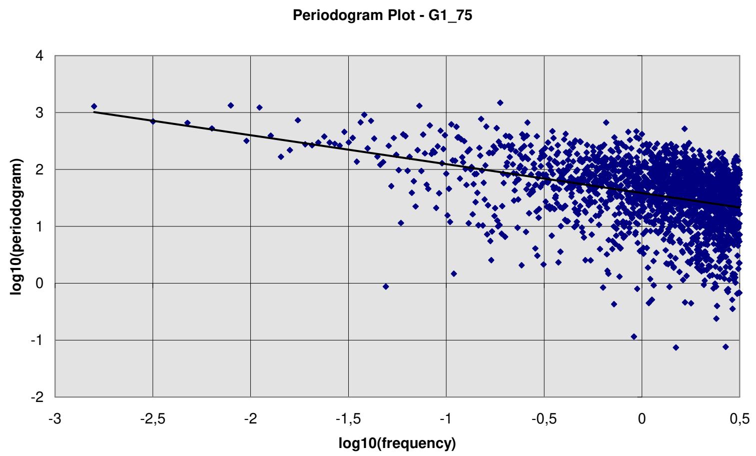




C.9 Periodogram Plots







Appendix D

Distributions

D.1 Poisson distributions

A Poisson process with parameter λ is a renewal process with exponentially distributed interarrival times. The defining formulae

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

gives the probability that the number of arrivals (or events) in time interval t is equal to n. The cumulative distribution function and the corresponding probability density function of the process are:

$$\begin{aligned} F(x) &= P(T \leq t) = 1 - e^{-\lambda t}, \\ f(x) &= \lambda e^{-\lambda t}. \end{aligned}$$

D.2 Pareto distributions

The classical Pareto distribution with shape parameter β and location parameter α has the cumulative distribution function:

$$F(x) = P(X \leq x) = 1 - (\alpha/x)^{\beta}, \quad \alpha, \beta \geq 0, \quad x \geq \alpha,$$

with corresponding probability density function:

$$f(x) = \beta \alpha^{\beta} x^{-\beta-1}.$$

If $\beta \leq 2$, then the distribution has infinite variance, and if $\beta \leq 1$, then it has infinite mean. (The Pareto distribution is also referred to as the power-law distribution, the double-exponential distribution, and the hyperbolic distribution.)

Appendix E

Software Tools

The following short statistical routines were used during the analysis. For the more detailed description of the programs listed below, see [14].

POISSON.C generates a simple Poisson process with parameter *lambda*.

MEAN.C calculates the mean value and the variance of the process.

HISTO.C generates the histograms for the cell count process.

DIST.C plots the cell distribution function for the given data set.

COV.C creates the autocorrelation function for the traffic trace.

IAT.C calculates the inter-arrival distribution.

IDC.C generates the IDC plot for a given data set.

VAR.C plots the variance-time diagram.

RS.C generates the R/S statistics for the input process.

PER.CPP calculates the periodogram for the data set.

SHUFFLE.C shuffles the data set by mixing the inter-arrival times.

TXT2PDA.C, IDX2PDA.C, CONVERT.C, HST2HST converters.

Bibliography

[1] J. Beran "Statistical Methods for Data with Long-Range Dependence", *PhD Thesis*, ETH Zurich, Switzerland, 1986.

[2] J. Beran "Statistical methods for data with long-range dependence", *Statistical Science*, 7(4):404–427, 1992.

[3] COST242 "The New COSTbook", *draft* Mar. 1996.

[4] D. R. Cox "Long-Range Dependence: A Review", *Statistics: An Appraisal*, H. A. David and F. T. David (Eds.), The Iowa State University Press, Ames, Iowa, pp. 55–74, 1984.

[5] N. G. Duffield, J. T. Lewis, N. O. Connel, R. Russel, F. Toomey "Predicting Quality of Service for Traffic with Long-Range Fluctuations", *Proc. ICC'95*, Seattle, WA, pp. 473–477, 1995.

[6] A. Erramilli, R. P. Singh "Application of Deterministic Chaotic Maps to Model Packet Traffic in Broadband Networks", *Proc. 7th ITC Specialists Seminar*, Morristown, NJ, 8.1.1–8.1.3, 1990.

[7] A. Erramilli, D. Gosby, W. Willinger "Engineering for Realistic Traffic: A Fractal Analysis of Burstiness", *Proc. of ITC Special Congress*, Bangalore, India, 1993.

[8] A. Erramilli, W. Willinger "Fractal Properties in Packet Traffic Measurements", ITC Regional Seminar, St. Petersburg, 1993. 1994.

[9] A. Erramilli, P. Pruthi, W. Willinger "Self-Similarity in High-Speed Network Traffic Measurements: Fact or Artifact?", *Proc. of the 12th Nordic Teletraffic Seminar*,Espoo, Finland, Jul. 1995.

[10] P. Flandrin "Wavelet Analysis and Synthesis of Fractional Brownian Motion", *IEEE Transactions on Information Theory*, 38(2), pp. 910–917, Mar. 1992.

[11] C. W. J. Granger, R. Joyeux "An Introduction to Long-Memory Time Series Models and Fractional Differencing", *J. Time Series Anal. 1*, pp. 15–29, 1980.

[12] F. R. Hampel "Data Analysis and Self-Similar Processes", *Proc. of the 46th Session of ISI*, Tokyo, pp. 235–254, 1987.

[13] A. Hegedűs, A. Vidács "Examination of the Self-Similar Features of Packet Traffic in High-Speed Networks", *Students' Technical Conference*, Technical University of Budapest, Hungary, 1995.

[14] A. Hegedűs "Fitting self-similar traffic models to real traffic traces", *Master's Thesis*, Technical University of Budapest, Hungary, May 1996.

[15] B. E. Helvik "Synthetic load generation for ATM traffic measurements", *Telektronikk*, Vol.91, No.2/3, pp. 171–194, Oct. 1995.

[16] J. R. M. Hosking "Modeling persistence in hydrological time series using fractional differencing", *Water Resources Research*, 20(12) pp. 1898–1908, 1984.

[17] Zs. Kenesi, A. Vidács "Measurement and Analysis of Ethernet LAN Traffic; Examination of Fractal Properties", *Students' Technical Conference*, Technical University of Budapest, Hungary, Oct. 1994.

[18] Zs. Kenesi, S. Molnár, A. Vidács "Fractals in Telecommunications", *Hungarian Telecommunications*, Vol.6, No.7, pp. 3–9, 1995.

[19] K. Kilkki, "Traffic Characterisation and Connection Admission Control in ATM Networks", *PhD Thesis*, Helsinki University of Technology, Finland, Jun. 1994.

[20] W. E. Leland, M. S. Taqqu, W. Willinger, D. W . Wilson "On the Self-Similar Nature of Ethernet Traffic (extended version)", *draft*, Mar. 1993.

[21] K. Lindberg "The FUNET Network Connection to Foreign Countries", *CSC News*, Vol.7, No.4, pp. 19, Dec. 1995.

[22] B. B. Mandelbrot "The Fractal Geometry of Nature", *Freeman*, New York, 1983.

[23] B. B. Mandelbrot "Long-run linearity, locally Gaussian processes, H-spectra and infinite variances", *International Economic Review*, Vol.10, pp. 54–60, 1994.

[24] F. Mosteller, J. W. Tukey "Data Analysis and Regression: A Second Course in Statistics", *Addison-Wesley Publ. Comp.*, Reading, MA, 1977.

[25] I. Norros "The Management of Large Flows of Connectionless Traffic on the Basis of Self-Similar Modeling", *IEEE International Conference on Communications*, Seattle, pp. 451–455, 1995.

[26] I. Norros "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks", *IEEE JSAC*, Vol.13, No.6, pp. 953–962, Aug. 1995.

[27] V. Paxson "Fast Approximation of Self-Similar Network Traffic", *technical report LBL-36750/UC-405*, University of California, Berkeley, Apr. 1995.

[28] M. de Prycker "Asynchronous Transfer Mode—Solution for Broadband ISDN" (Second. Ed.), *Ellis Horwood*, Great Britain, 1993.

[29] P. Pruthi "An Application of Chaotic Maps to Packet Traffic Modeling", *PhD Thesis*, Royal Institute of Technology, Stockholm, Sweden, Oct. 1995.

[30] K. Rahko, S. Hertzberg "Traffic Measurements (A TETRAPRO Specialized Course)", *ITU CODEVTEL Course/10.03:01 851*, Otapublishing, Espoo, Finland, 1988.

[31] V. Ramaswami "Traffic Performance Modeling for Packet Communication: Whence, Where and Whither?", *Proc. Australian Teletraffic Seminar*, Nov. 1988.

[32] O. Rose, "Estimation of the Hurst Parameter of Long-Range Dependent Time Series", *Report No. 137*, University of Würzburg, Institute of Computer Science, Feb. 1996.

[33] G. D. Stamoulis, M. E. Anagnostou, A. D. Georgantas "Traffic Source Models for ATM Networks: a Survey", *Computer Communications*, vol.17, No.6, pp. 428–438, Jun. 1994.

[34] M. S. Taqqu, J. Levy "Using renewal processes to generate long-range dependence and high variability", *Dependence in Probability and Statistics*, pp. 73–89, Boston, Birkhäuser, 1986.

[35] V. Teverovsky, M. S. Taqqu "Testing for Long-Range Dependence in the Presence of Shifting Means or a Slowly Declining Trend Using a Variance-Type Estimator", *preprint*, 1995.

[36] A. Vidács, Zs. Kenesi, Á. Rétfalvi, P. Pozsgai, S. Molnár "Traffic Description for Telecom Networks", *Hungarian telecommunications—Selected Papers*, pp. 18-25, 1995.

[37] P. Whittle "Estimation and information in stationary time series", *Arkiv för Matematik*, 2:423–434, 1953.

[38] W. Willinger, M. S. Taqqu, R. Sherman, D. V. Wilson "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level", *to appear in: Proc. of the ACM/Sigcomm'95*, 1995.

[39] The International Traffic Archive, `http://town.hall.org/Archives/pub/ITA/`

[40] "HP Broadband Series Test System—System Overview", *Hewlett-Packard Company*, 5091–7870E 12/93, 1993.