Discrete Self-Similarity

A. Gefferth, S. Molnár
Budapest University of Technology and Economics
High Speed Networks Laboratory
Magyar Tudósok körútja 2, H-1117 Budapest, Hungary
{gefferth,molnar}@ttt-atm.ttt.bme.hu

D. Veitch
ARC Special Research Center for Ultra Broadband Information Networks
Dept. of Electrical and Electronic Engineering, The University of Melbourne.
Victoria 3010, Australia
dveitch@unimelb.edu.au

Abstract

There is now a substantial literature demonstrating that traffic arrival proceses in packet switched networks can be characterised as long-range dependent and asymptotically self-similar. The goal of this paper is to explain these terms for traffic engineers, who are not interested in detailed mathematical proofs, but have to face datasets with long-range dependence during their work. The different terms occurring in this field are defined, and their connections are explored in the light of the latest results. The importance of the new results are highlighted using examples of practical importance.

1. Introduction

For our purposes a dataset or *time series* is formed by calculating the number of bytes or packets traversing a link in a telecommunication network in neighbouring intervals of the same duration. This dataset can be viewed as a sample path of a stochastic process. Analyzing the statistical properties of this process is necessary for efficient traffic management and dimensioning.

It was found in earlier works ([16, 13, 6]) that in contrast to traditional telephone networks, where the traffic can be described by traditional stochastic processes, the traffic of packet switched networks is highly self-correlated, and so has a so-called long-range dependent, or asymptotic self-similar nature.

This long-range dependence influences among other things queuing and multiplexing performance ([12, 14, 10]) and makes the dimensioning of networks difficult.

Although many papers deal with the mathematical and the practical aspects of long-range dependence it was found that in view of our latest results ([8]) the mathematics becomes simpler, easier to understand, and these results also have some practical consequences worth knowing.

The rest of this paper is organised as follows: Section 2 introduces the stochastic processes and their basic descriptors, like the autocovariance and autocorrelation functions. In section 3 the exact definitions of self-similarity, long-range dependence, etc. will be given and their connection explored. In section 4 some practical consequences of the new results will be highlighted, while section 5 summarises and concludes the paper.

2. Preliminaries

In this paper no measured traffic traces will be analysed, rather those mathematical models will be investigated that can serve as a model for them.

Real life datasets are in general not stationary for their whole length, but stationary intervals can usually be identified. In fact all of the mathematical analysis tools require one or other type of stationarity so this assumption ought to be verified in each case.

It is however important to understand the stationary models before turning our attention to the more sophisticated non-stationary ones, therefore in this paper all mathematical models will describe stationary processes.

Definition 2.1 (Second-order stationarity)
A discrete time stochastic process $\{X(t),\ t \in \mathbb{Z}\}$, where $\mathbb{Z}$ denotes the set of integers, will be called *second-order stationary* if the mean and variance of $X(t)$ are each independent of t, and also $\mathrm{E}[X(t+k)X(t)]$ is independent of t for any $k \in \mathbb{Z}$, where $\mathrm{E}[Q]$ means the expected value of the random variable Q.

For simplicity, without loss of generality it will be assumed that the process has zero mean. Apart from the existence of a finite variance no assumption is made on the marginal distributions of the process. In the following the term stationarity will mean second-order stationarity.

Stationarity enables to define the *autocovariance function* as

Definition 2.2 (Autocovariance function)
The autocovariance function (ACVF) of a stationary process $X(t)$ is defined as

$$\gamma(k) := \mathrm{E}[X(0)X(k)].$$

Normalising γ by $\gamma(0)$, which by definition is equal to $\mathcal{V}$, the variance of X, yields the *autocorrelation function*, $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$. Normalisation hides the "size" of the variation it gives a qualitative description of the covariances.

The autocovariance function plays an important role in our analysis because in this paper we only concentrate on second-order properties, and it is the autocovariance function that unambiguously characterises the process for our analysis.

Self-similarity (see definition 3.1) corresponds to the invariance of the qualitative behaviour of the process when viewed at different time scales. Going to a coarser time scale from a finer time scale can be achieved by recalculating the sum, or average of the values in neighbouring non-overlapping intervals. This operation will be called aggregation and the aggregated process ($X^{(m)}$) is defined as

$$X^{(m)}(t) := \frac{1}{m} \sum_{j=m(t-1)+1}^{mt} X(j).$$

The variance, γ, and ρ functions of $X^{(m)}$ will be denoted by $\mathcal{V}^{(m)}$, $\gamma^{(m)}$ and $\rho^{(m)}$ respectively.

$\mathcal{V}^{(m)}$, also called as *aggregated variance* or, *variance-time function*, can be expressed in terms of $\gamma(k)$ as follows:

$$\mathcal{V}^{(m)} = E\left[\left(\frac{X_1 + X_2 + \cdots + X_n}{m}\right)^2\right] \qquad (1)$$

$$= \frac{1}{m^2} \sum_{j=0}^{m-1} \sum_{i=-j}^{j} \gamma(i) \qquad (2)$$

For the traditional so called short-range dependent processes the dataset gets asymptotically uncorrelated as the level of aggregation increases, that is for each fixed $k \neq 0$ $\rho^{(m)}(k) \overset{m}{\to} 0$. The variance of $X(m)$ has a fast decay $\mathcal{V}^{(m)} \sim C\frac{1}{m}$.

The traffic traces measured in packet switched networks behaves differently. Even for large values of m $\rho^{(m)}(k)$ does not vanish, and the decay rate of variance can be modelled as $\mathcal{V}^{(m)} \sim Cm^{\alpha}$, where $\alpha \in (0, -1)$.

This influences among others queuing behaviour, makes an estimation of the mean of the traffic volume more unreliable, and so makes the dimensioning and management of the network difficult.

3. Results

We start with the definition of self-similarity.

Definition 3.1 (Second-order self-similarity (*SS*))
A process is (exactly) second-order self-similar if $\rho^{(m)} \equiv \rho$ for all $m = 1, 2, 3, \ldots$.

As it can easily be justified by some simple algebra and is well-known [5] the fractional Gaussian noise (fGn) satisfies definition 3.1 of self-similarity.

Fractional Gaussian noise is a Gaussian process (all finite distributions are Gaussian) with an autocorrelation function of $\rho(k) = \frac{1}{2}\delta_m^2\{m^{2H}\}(k)$, where δ^2 is the double-differencing operator defined as:

$$\delta_i^2\{f(i)\}(n) = \begin{cases} 2f(1) & : n = 0 \\ f(2) - 2f(1) & : n = 1 \\ f(n+1) - 2f(n) + f(n-1) & : n > 1. \end{cases}$$

and the parameter $H \in [0, 1]$ is the famous Hurst parameter.

Because the exact distributions of the random variables play no role in this paper, we will use fractional noise (FN) instead of fGn, which has the same autocorrelation function, but the requirement of Gaussianity has been dropped.

It hasn't been investigated before whether there are any other SS processes, and if yes whether they can also appear in communication networks and what influence they may have. Such an investigation is also useful for a better understanding of the nature of the already known fractional noise.

Contrary to the common belief that there are no other SS processes besides FN [5], our investigations revealed a new kind of self-similar process, which was named almost periodic [9] because of the visual periodicity of its autocorrelation function, as depicted in figure 1.

Definition 3.2 (The almost periodic fixed point family $\mathrm{AP}_{q,c}$)
The autocorrelation function of a $\mathrm{AP}_{q,c}$ process is $\rho(k) = \frac{1}{2}\delta_s^2\{f_{q,c}(s)\}(k)$, where $f_{q,c}(rq^n) = c^n$, where n is non-negative integer, q is a prime number that is not an integer divisor of r and $c \in (0, 1)$.

Theorem 3.1
Members of the $\mathrm{AP}_{q,c}$ family are self-similar.

In [8] it was also shown that besides the well known fractional noise family and the newly discovered AP family no other self-similar processes exist.

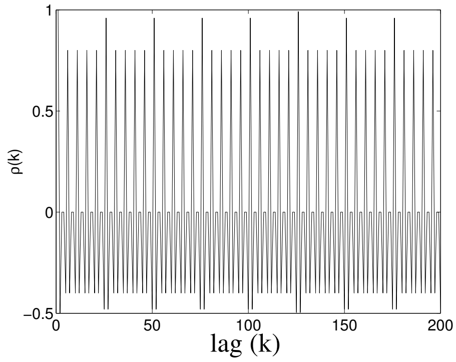

Figure 1. Autocorrelation function of AP$_{5,0.2}$

Having a quick look at the very complicated definition of the autocorrelation function of AP processes it can immediately be seen that they are not likely to appear in telecommunication networks. We can thus conclude that associating fractional noise with self-similarity was wrong from the mathematicians but acceptable from the traffic engineers point of view, and also that having a clear view in this topic is beneficial for both of the above.

Because of their greater importance in the following we will concentrate on fractional noise and related processes.

Besides the exactly self-similar processes, where $\rho^{(m)}$ is independent of m, there exists the class of *asymptotically self-similar (ASS)* processes, where $\rho^{(m)}$ changes with m, but $\lim_{m\to\infty} \rho^{(m)} =: \rho^*$, is the autocorrelation function of an exactly self-similar process.

Definition 3.3 (Asymptotic second-order self-similarity (*ASS*))
A process is asymptotically second-order self-similar (to ρ^*) if $\lim_{m\to\infty} \rho^{(m)}(k) =: \rho^*(k)$, for all $k \in \mathbb{Z}$.

It can be shown that in this case ρ^* is SS.

The importance of asymptotically self-similar processes lies in the fact that the traffic volume measured on telecommunication networks belongs to a subset of them, namely the class of so called *long-range dependent (LRD)* processes. The difference between LRD and ASS is not always clear, and in some papers they are even treated as equal. Furthermore several different definitions of long-range dependence can be found in the literature.

Definition 3.4 (LRD1)
LRD1 processes are those whose ACVFs obey $\gamma(k) \sim c_\gamma k^{2H-2}$, $H \in (0.5, 1)$, and c_γ is a positive constant.

This definition is the simplest and most frequently encountered. For example it is used in [5, 2, 13, 1].

This definition can be generalised by allowing arbitrary slowly varying function (see definition 3.7) in place of the constant c_γ.

Definition 3.5 (LRD2)
LRD2 processes are those whose ACVFs obey $\gamma(k) = c_\gamma(k)k^{2H-2}$ with $H \in (0.5, 1)$ and c_γ is a *slowly varying function*.

Slow and regular variation was originally introduced for continuous time functions [3].

Definition 3.6 (Regular variation in continuous time)
A function $S(x)$ is slowly varying if for all $t \in \mathbb{R}^+$

$$\frac{L(xt)}{L(x)} \xrightarrow{x} 1, \tag{3}$$

where $\mathbb{R}^+$ is the set of positive real numbers. A function $R(x)$ is regularly varying with index α if for all $t \in \mathbb{R}^+$

$$\frac{R(xt)}{R(x)} \xrightarrow{x} t^\alpha. \tag{4}$$

It is clear that regular variation with index 0 is slow variation. It is not complicated to show that

Theorem 3.2
The continuous time function $R(x)$ is regularly varying with index α if and only if $R(x) = L(x)x^\alpha$, where $L(x)$ is slowly varying.

Examples for slowly varying functions include any function that converges to a positive constant, or the logarithm function.

It makes sense to define slow and regular variation in discrete time. Some authors define them analogously to equations (3) and (4) where $t \in \mathbb{R}^+$ is replaced by $n \in \mathbb{Z}^+$, and R is said to be regularly varying if

$$\frac{R(nm)}{R(m)} \xrightarrow{m} n^\alpha. \tag{5}$$

This definition although simple and straightforward has a drawback. Equation (5) imposes much less constraint on the function than its continuous counterpart equation (4). As a result of this the class of functions defined by (5) is to big, including functions with some undesired behaviour. Therefore nice properties of the continuous time regular variation like the cumulative sum of regularly varying functions remains regularly varying, the ratio of neighbouring values converges to 1, etc. do not carry over. To save most of the convenient properties discrete time regular variation will be defined in a more restrictive way as follows [7]:

Definition 3.7 (Discrete regular variation (DRV))
A function f defined on $\mathbb{Z}^+$ is *regularly varying* with index α if there exists a continuous time regular varying function $\widetilde{f}$ such that $f(n) = \widetilde{f}(n)$ for all $n \in \mathbb{Z}^+$.

Unfortunately none of the above definitions for long-range dependence were found to be flexible enough to capture all processes that have such a slowly decaying autocorrelation function that $\rho^{(m)}$ does not vanish even as $m \to \infty$. Therefore we propose using the definition of [8], which says:

Definition 3.8 (LRD)
All asymptotically self-similar processes that converge to FN_H with $H \in (0.5, 1)$ are long-range dependent (LRD).

Later on it will turn out that LRD is not equivalent to LRD2.

The first part of the following theorem appeared in several places, including in [2] and [15].

Theorem 3.3
For LRD1 processes

$$\mathcal{V}^{(m)} \sim c_\gamma \frac{m^{2H-2}}{H(2H-1)}. \tag{6}$$

The result can be extended to LRD2 processes as

$$\mathcal{V}^{(m)} \sim c_\gamma(m) \frac{m^{2H-2}}{H(2H-1)}. \tag{7}$$

Although equation (6) appears in many places, its proof is either omitted or is only valid for a special case of LRD1 processes, or uses some additional unproved lemma. Equation (7) and its proof can be found in [8].

Definition 3.9
A stochastic process that satisfies $\mathcal{V}^{(m)} \sim c_\gamma(m) \frac{m^{2H-2}}{H(2H-1)}$ with $H \geq 0.5$ and $c_\gamma(m)$ is slowly varying will be called a process of *slowly decaying variance (SDV).*

SDV processes are in fact regularly varying. The word "slowly" in this definition should not be confused with slow variation. It only highlights, that the variance has a slower decay than in the case of traditional short-range dependent models.

The connection between slowly decaying variance and asymptotic self-similarity is given by the following

Theorem 3.4
All processes having a slowly decaying variance are asymptotically self-similar, converging to FN_H.

Theorem 3.4 together with (7) means that LRD2 is a subset of LRD. The question naturally arises whether the above statements can be reversed. The quick answer is that there exist processes with slowly decaying variance, which are not LRD2, and it is not known yet whether there exist any processes which converge to FN_H, with $H \geq 0.5$ but do not

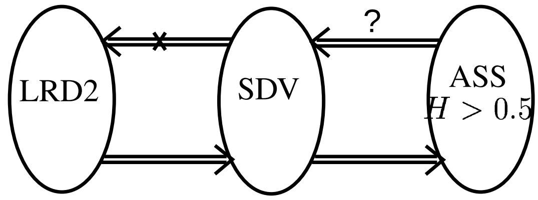

Figure 2. Connection of long-range dependence, slow-decay of variance (SDV) and asymptotic self-similarity (ASS). All processes that satisfy $\gamma(k) \sim c_\gamma(k) k^{2H-2}$ (LRD2) also satisfiy $\mathcal{V}^{(m)} \sim c_\gamma(m) \frac{m^{2H-2}}{H(2H-1)}$ (SDV), and these processes are ASS converging to FN_H. In the first case the converse is not true, while in the second case it is not known yet.

have slowly decaying variance. This is depicted in figure 2.

In detail:

To show that LRD contains LRD2 as a subclass a process will be constructed that is a member of the set LRD$\setminus$LRD2. Let X_1 and X_2 be independent copies of an FN_H process with ACVF γ^*, $H \in (0.5, 1]$. We define $Y(t)$, $t \in \mathbb{Z}$ by deterministically alternating between the two copies:

$$Y(t) := \begin{cases} X_1(t/2), & t \text{ even} \\ X_2(\frac{t-1}{2}), & t \text{ odd} \end{cases}$$

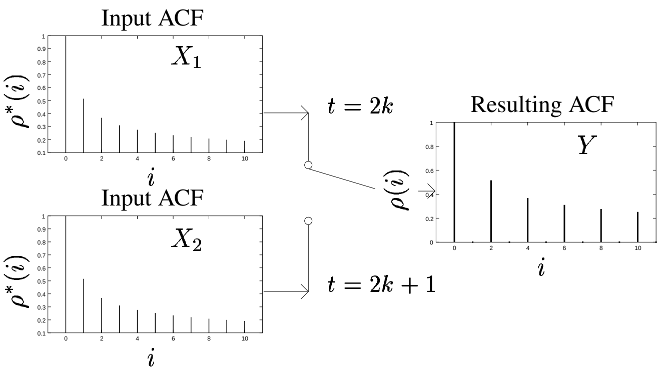

Figure 3. Example of a process in LRD$\setminus$LRD2. Odd lags compare different processes and are uncorrelated. (ACF: Autocorrelation Function)

as illustrated in figure 3. It can be shown that Y is a second-order stationary process, with ACVF

$$\gamma_Y(k) = \begin{cases} \gamma^*(k/2), & k \text{ even} \\ 0, & k \text{ odd} \end{cases},$$

and after some algebraic transformations using (2) we get:

$$\mathcal{V}_Y^{(m)} = \begin{cases} \frac{1}{2}\mathcal{V}^{*(m/2)}, & m \text{ even} \\ \frac{1}{2}\left(\frac{\mathcal{V}^{*((m-1)/2)}+\mathcal{V}^{*((m+1)/2)}}{2}\right), & m \text{ odd} \end{cases}, \quad (8)$$

where $\mathcal{V}^*$ is the aggregated variance function of FN.

Equation (8) shows that the aggregated variances of Y ($\mathcal{V}_Y^{(m)}$) behave asymptotically as $\frac{1}{2}\mathcal{V}^{*(m/2)}$ the aggregated variances of the fractional noise, and so according to theorem 3.4 Y is LRD but obviously not LRD2.

Although this example might first seem awkward its existence shows that LRD2 and LRD are not equivalent, and there is no reason (yet) why other examples could not be constructed. In section 4 an example will be given of how the process Y can appear in a teletraffic network.

To explain the difference between processes that converge to FN_H , with $H \geq 0.5$ (LRD processes) and processes of slowly decaying variance a sufficient and necessary condition for being ASS to FN_H is presented.

Theorem 3.5
A process is ASS to FN_H if and only if

$$\lim_{m\to\infty} \frac{\mathcal{V}^{(nm)}}{\mathcal{V}^{(m)}} = n^{2H-2},$$

for each $n \in \mathbb{Z}^+$.

This looks very similar to the definition of the continuous time regular variation and in fact it can easily be shown that also in discrete time if R is regularly varying with index α then it satisfies (5). It is also clear that there are functions satisfying (5) that are not regularly varying [7]. But it is not trivial whether such a function can appear as an aggregated variance time function. The autocorrelation function has to satisfy the constraint of *positive semi-definiteness* [4], which is a simple consequence of the positivity of the variance of any finite linear combination of some values of the process.

Definition 3.10 (Positive semi-definiteness)
the function γ is *positive semi-definite* if

$$\sum_{1\leq i,j\leq n} a_i\gamma(i-j)a_j \geq 0. \quad (9)$$

where $\underline{\mathbf{a}} = (a_1, a_2, \ldots, a_n)^T$ is any constant real vector of length n.

Positive semi-definiteness excludes many of the discrete time functions to be valid as an autocorrelation function and so also many functions cannot appear as aggregated variance functions. So the question is whether there are any functions that satisfy (5), are not regularly varying, and the corresponding autocorrelation function is positive semi-definite. Although this has not been investigated before, we believe that even if such functions exist they are highly pathological and not likely to appear in practical situations.

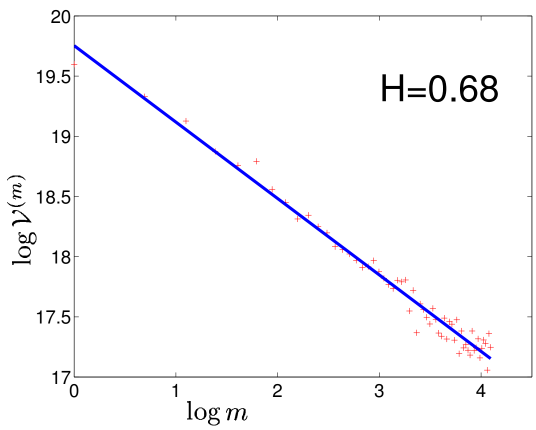

Figure 4. Hurst parameter estimation using the Variance Time Plot method

4. Consequences

The effect of long-range dependence on various QoS parameters have already been studied and reported in [12, 14, 10]. In this section we only concentrate on the effects of the new results, using practical examples.

Several statistical tests and estimators have already been developed to detect the presence and estimate the parameters (H, c_γ) of long-range dependence and asymptotic self-similarity. In view of our new results, it is possible to highlight some possible hazards during the interpretation of the output of these estimators.

In our example one of the simplest estimators, the variance time-plot will be investigated, although similar examples could also be constructed for some other estimators too. Variance time-plot is based on equation (6). Using numerical methods $\mathcal{V}^{(m)}$ is estimated and is plotted against m in a log-log scale, as visualised in figure 4.

If the process in question was LRD1 with Hurst parameter H, then the tail of the plot should be a straight line with slope $2H - 2$. So first the straightness of the tail should be judged. If it is straight its slope has to be measured and the Hurst parameter has to be estimated. In fact this particular estimator is statistically not very reliable, meaning roughly that very long traces are needed to get a reasonable estimate, and also that non-stationarities migh easily mislead the estimator [1] [11].

[1] Other, more sophistcated estimators perform much better in these aspects.

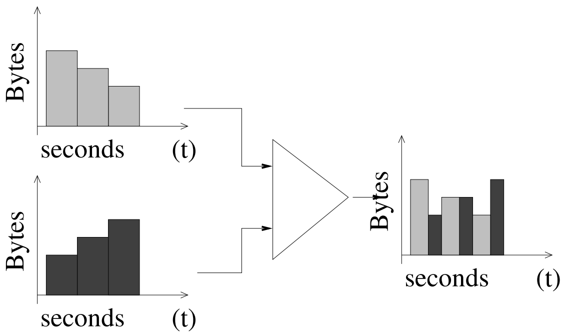

Figure 5. Deterministical multiplexing of two independent traffic streams.

But in view of our results we can identify two other shortcomings of this estimator.

Example 1a:

Assume that the process in question is LRD2 but not LRD1. In this case, even if a perfect estimation of $\mathcal{V}^{(m)}$ is assumed, the tail of the slope will never be a straight line. This estimator is not suitable to detect the presence of LRD2.

This drawback is common to all estimators that restrict attention to LRD1, such as the R/S plot, or the periodogram plot [2].

Example 1b:

Variance time-plot is able to detect if a process is of slowly decaying variance or not. It is a common mistake to assume that SDV implies LRD1 or LRD2 and so conclude LRD1 from the straightness of the tail of the plot.

It has to be noted however that the presented drawbacks do not mean that the estimators are useless or wrong. They only show that the interpretaion of the output of the estimator might include some hidden hazards, which can be eliminated by careful analysis.

Example 2:

This example shows a possible scenario, mentioned in Example 1b, where the process is SDV but not LRD.

Let $I_1(k)$ and $I_2(k)$ be two independent traffic streams, carrying e.g. real-time variable bitrate multimedia information. Here $I_i(k)$ denotes the amount of bytes transmitted in timeframe $[kt_0, (k+1)t_0)$ of process i. These two streams are fed into a network that lacks any type of congestion control, and are deterministically multiplexed to form a single output stream, $O(n)$, with double bitrate where

$$O(n) = \begin{cases} I_1(n/2), & n \text{ even} \\ I_2((n-1)/2), & n \text{ odd} \end{cases}$$

is the amount of bytes transmitted in timeframe $[n\frac{t_0}{2}, (n+1)\frac{t_0}{2})$ as illustrated in figure 5.

Now if both I_1 and I_2 were LRD1, with H_1 and H_2, where $H_1 > H_2$ then similarly to the example presented after theorem 3.4 it can be shown that $O(n)$ will be LRD and so SDV with $H = H1$, but it will definitely not be LRD2, and so it will show different queueing and statistical behaviour. Therefore if such sort of deterministical multiplexing is present in the network it is advisable to analyse the traffic traces before entering the multiplexer. If the traffic can only be measured after multiplexing the input streams have to be separated to gain an accurate view of the traffic statistics.

5. Conclusions

In this paper the recent complete definitions of exact and asymptotic self-similarity, discrete regular variation, and several forms of long-range dependence were reviewed and summarised with practical implications discussed. All possible self-similar processes were presented. It was shown that a regular variation of the autocovariance function results in a regular variation of the variance-time function, and a regular variation of the variance-time function results in being asymptotically self-similar. It was also shown that the converses are not necessarily true. Practical examples were presented to highlight the importance of the results from the traffic engineers point of view. It was shown that care should be taken when interpreting the results of traffic analysis tools.

Acknowledgment

András Gefferth gratefully acknowledges the support by the Australian-European Awards Program, an Australian Government funded Award administered by the Department of Employment, Education, Training and Youth Affairs. The authors gratefully acknowledge the support of Ericsson Australia.

References

[1] P. Abry and D. Veitch. Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, January 1998.

[2] J. Beran. *Statistics for Long-Memory Processes*. Chapman and Hall, New York, 1994.

[3] N. Bingham, C. Goldie, and J. Teugels. *Regular Variation*. Cambridge University Press, Cambridge England, 1987.

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

[5] D. R. Cox. *in Long-range dependence: a review,* H.A. David and H.T. David editors, chapter Statistics : an Appraisal, pages 55–74. Iowa State University Press, Ames (IA), 1984.

[6] I. Csabai. 1/f Noise in Computer Network Traffic. *Journal of Physics*, 1994.

[7] J. Galambos and E. Seneta. Regularly Varying Sequences. *Proceedings of the American Mathematical Society*, 41(1):110–116, November 1973.

[8] A. Gefferth, D. Veitch, I. Maricza, S. Molnár, and I. Ruzsa. The Nature of Discrete Second-Order Self-Similarity. *submitted for publication*, 2002.

[9] A. Gefferth, D. Veitch, I. Ruzsa, I. Maricza, and S. Molnár. A New Class of Second Order Self-Similar Processes. *submitted for publication*, 2002.

[10] T. G. Kurtz. Limit theorems for workload input models. In *Stochastic Networks: Theory and Applications*, pages 339–366. Clarendon Press, Oxford,UK, 1996.

[11] S. Molnár and T. D. Dang. Pitfalls in Long Range Dependence Testing and Estimation. In *GLOBECOM 2000*, San Francisco, CA, USA, Nov. 2000.

[12] S. Molnár and I. M. (eds.). Source Characterization in Broadband Networks. Technical report, COST 257, Jan. 1999. Interim Report.

[13] S. Molnár and A. Gefferth. On the Scaling and Burst Structure of Data Traffic. In *8th International Conference on Telecommunication Systems, Modelling and Analysis*, Nashville, Tennessee, USA, Mar. 2000.

[14] S. Molnár and A. Vidács. On modeling and shaping self-similar ATM traffic. In *15th International Teletraffic Congress*, Washington, D.C.,USA, June 1997.

[15] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall, 1994.

[16] W. Willinger, M. Taqqu, and A. Erramilli. *in Stochastic Networks: Theory and Applications,* F. P. Kelly and S. Zachary and I. Ziedins, editors, chapter A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High-Speed Networks, pages 339–366. Clarendon Press (Oxford University Press), Oxford, 1996.