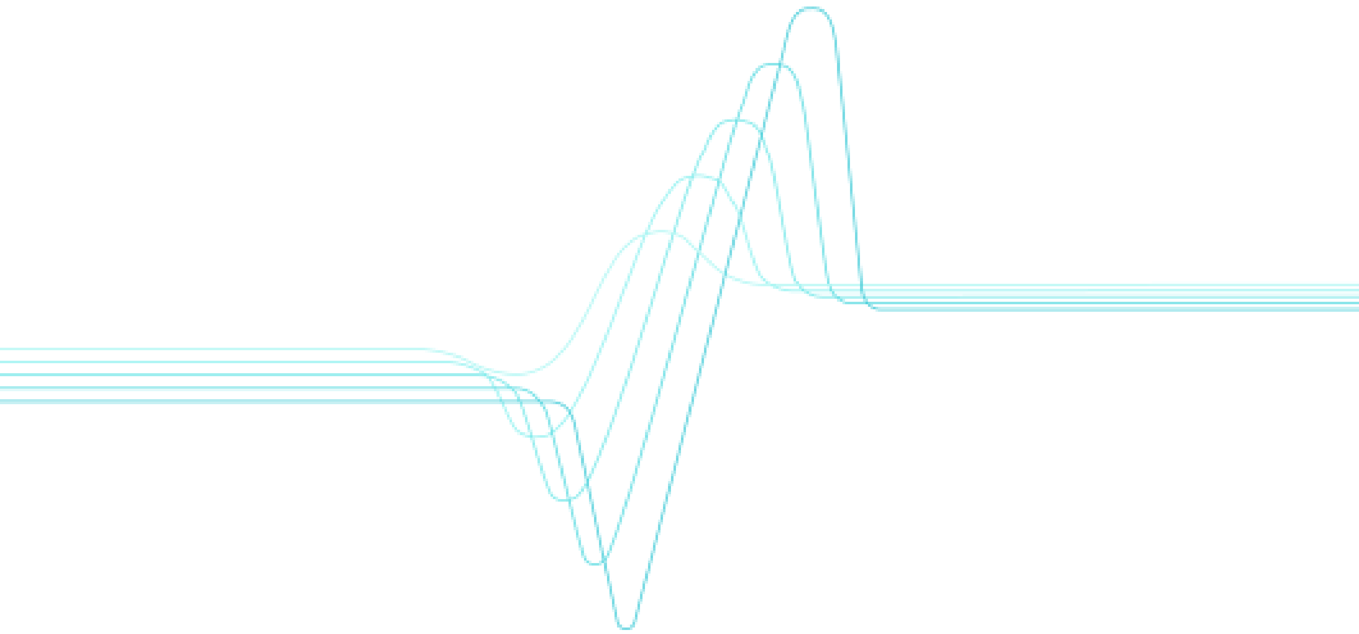
ESPOO 2003 VTT PUBLICATIONS 491

Petteri Mannersalo

Gaussian and multifractal processes in teletraffic theory



VTT PUBLICATIONS 491

Gaussian and multifractal processes in teletraffic theory

Petteri Mannersalo

VTT Information Technology

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Electrical and Communications Engineering for public examination and debate in Auditorium S4 at Helsinki University of Technology (Espoo, Finland) on the 25th of April, 2003, at 12 o'clock noon.

ISBN 951–38–6036–1 (soft back ed.)
ISSN 1235–0621 (soft back ed.)

ISBN 951–38–6037–X (URL: http://www.inf.vtt.fi/pdf/)
ISSN 1455–0849 (URL: http://www.inf.vtt.fi/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 4374

VTT, Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (09) 4561, fax (09) 456 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 4374

VTT Tietotekniikka, Otakaari 7 B, PL 1202, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 7013

VTT Informationsteknik, Otsvängen 7 B, PB 1202, 02044 VTT
tel. växel (09) 4561, fax (09) 456 7013

VTT Information Technology, Otakaari 7 B, P.O.Box 1202, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 7013

Technical editing Maini Manninen

Otamedia Oy, Espoo 2003

Mannersalo, Petteri. Gaussian and multifractal processes in teletraffic theory. Espoo 2003. VTT Publications 491. 44 p. + app. 109 p.

Keywords Gaussian processes, multifractals, queueing systems, performance analysis, traffic modeling

Abstract

In this thesis, we consider two classes of stochastic models which both capture some of the essential properties of teletraffic. Teletraffic has two time regimes where profoundly different behavior and characteristics are seen. When traffic traces are observed at coarse resolutions, properties like self-similarity and long-range dependence are visible. In small time-scales, traffic exhibits complex scaling laws with much more spiky bursts than in coarser resolutions. The main part of the thesis is devoted to a large time-scale analysis by considering Gaussian processes and queueing systems with Gaussian input. In order to understand the small time-scale dynamics, first steps are taken towards general multifractal models offering a suitable basis for short time-scale teletraffic modeling.

The family of Gaussian processes with stationary increments serves as the traffic model for large time-scales. First, we introduce a fast and accurate simulation algorithm, which can be used to generate long approximate Gaussian traces. Moreover, the algorithm is also modified to run on-the-fly. Then approximate queue length distributions for ordinary, priority and generalized processor sharing queues are derived using a most probable path approach. Simulation studies show that the performance formulae appear to be quite accurate over the full range of buffer levels. Finally, we construct a semi-stationary predictor, which uses a constant variance function and mean rate estimation based on a moving average method. Moreover, we show that measuring the past of a process by geometrically increasing intervals is a good engineering solution and a much better way than equally spaced measurements.

We introduce a family of multifractal processes which belongs to the framework of *T*-martingales and multiplicative chaos introduced by Kahane. The family has many desirable properties like stationarity of increments, concave multifractal spectra and simple construction. We derive, for example, conditions for non-degeneracy, establish a power law for the moments and obtain a formula for the multifractal spectrum.

Preface

The major part of the research reported in this thesis was done at VTT Information Technology while I worked on the international COST 257 and COST 279 projects during 1997–2002. In addition, I was employed by the Finnish Academy project "Fractal processes in telecommunications" in 2000. This period also included a 6 month visit to AT&T Labs-Research. The completion of this thesis has been funded by VTT and a grant from the Sonera research foundation.

I am deeply grateful to Research Professor Ilkka Norros. In addition to his valuable instructions and support, he has been an excellent role model as a broad-minded researcher having interests from horticulture to brass bands. I would also like to thank my supervisor Professor Jorma Virtamo who, even offering his own free time, ensured that the process advanced in time. Professor Chris Blondia and Professor Bo Friis Nielsen deserve acknowledgments for the pre-examination of the manuscript.

Additionally, I wish to express my gratitude to Ron Addie, PhD, and Rolf Riedi, PhD, for rewarding research collaboration, and to my room mate at VTT, Jorma Kilpi, Lic. Phil., for having the patience to discuss even the simplest mathematical questions. Finally, I would like to thank the personnel of VTT Sähkötalo, and especially our coffee club, for creating a good working environment.

Espoo, February 1, 2003

Petteri Mannersalo

Contents

Appendices of this publication are not included in the PDF version.
Please order the printed version to get the complete publication
(http://www.vtt.fi/inf/pdf/)

List of publications

[1] Norros, I., Mannersalo, P. and Wang, J. Simulation of fractional Brownian motion with conditionalized random midpoint displacement. *Advances in Performance Analysis*, 1999. Vol. 2(1), pp. 77–101.

[2] Addie, R., Mannersalo, P. and Norros, I. Performance formulae for queues with Gaussian input. In *Proceedings of ITC 16*. Edinburgh, UK, 1999. Pp. 1169–1178.

[3] Addie, R., Mannersalo, P. and Norros, I. Most probable paths and performance formulae for buffers with Gaussian input traffic. *European Transactions in Telecommunications*, 2002. Vol. 13(3), pp. 183–196.

[4] Mannersalo, P. and Norros, I. Approximate formulae for Gaussian priority queues. In *Proceedings of ITC 17*. Salvador, Brazil, 2001. Pp. 991–1002.

[5] Mannersalo, P. and Norros, I. GPS schedulers and Gaussian traffic. In *Proceedings of IEEE Infocom 2002*. New York, USA, 2002. Pp. 1660–1667.

[6] Mannersalo, P. and Norros, I. A most probable path approach to queueing systems with general Gaussian input. *Computer Networks*, 2002. Vol. 40(3), pp. 399–412.

[7] Mannersalo, P. Some notes on prediction of teletraffic. In *Proceedings of 15th ITC Specialist Seminar*. Würzburg, Germany, 2002. Pp. 220–229.

[8] Mannersalo, P., Norros, I. and Riedi, R. Multifractal products of stochastic processes: construction and some basic properties. *Advances in Applied Probability*, 2002. Vol. 34(4), pp. 888–903.

List of abbreviations

ADSL	Asymmetric Digital Subscriber Line
ARPA	Advanced Research Projects Agency, USA
ATM	Asynchronous Transfer Mode
CLT	Central Limit Theorem
CPU	Central Processing Unit
DARPA	Defense Advanced Research Projects Agency, USA
FBM	Fractional Brownian Motion
FCLT	Functional Central Limit Theorem
FFT	Fast Fourier Transformation
FTP	File Transfer Protocol
FUNET	Finnish University and Research Network
GPS	Generalized Processor Sharing
IP	Internet Protocol
LAN	Local Area Network
QoS	Quality of Service
RAND	Research and Development, U.S. government agency
RFLA	Rough Full Link Approximation
RKHS	Reproducing Kernel Hilbert Space
RMD	Random Midpoint Displacement
TCP	Transmission Control Protocol
WAN	Wide Area Network
WWW	World Wide Web

1. Introduction

1.1 Teletraffic theory

Teletraffic theory, in its widest definition, is a term linking together mathematics applicable to performance analysis, design, control and management of telecommunication systems. This includes topics like mathematical modeling, queueing theory and optimization. Since almost all the problems in this area have some random components, stochastic analysis and modeling play a major role.

The history of teletraffic theory is a successful combination of mathematics and engineering profiting both disciplines. It all began 100 years ago when the mathematical analysis of telephone networks was started by Erlang [Erl09]. Erlang's studies are often considered to be the birth of queueing theory. On the other hand, by using Erlang's formula and related mathematics, engineers have been able to design and operate traditional telephone networks for over fifty years so that the utilization rate is very high without impairing the quality of service seen by customers.

The next big step was moving from circuit switching (telephone networks) to packet switching (data networks). In early 1960s, there were three independent teams working with the packet network concept: RAND Corporation, British National Physical Laboratory and ARPA[1]. There is some disagreement about who originally invented the notion of packet switching (see e.g. [Dav01, Kle02, Bar02]). However, Baran, Davies and Kleinrock are the recognized pioneers who all worked on this idea. The mathematical analysis of data networks progressed jointly with technological advances. The principles of packet switching and queueing theory were part of the innovations leading to ARPANET and, later on, to the Internet.

Though the early data networks had a strong basis in analytical studies, the driving force of the Internet has been experimental engineering curiosity, the role of teletraffic theory being very small. Only recently have the Internet operators started to show interest in the performance analysis of their networks. Best effort service is not always enough since there are applications requiring a guaranteed quality of service (QoS). This has forced the Internet operators to change their view. Unfortunately, the current Internet protocols have very limited traffic engineering capabilities.

The ATM technology was proposed as an intelligent compromise between telephone and data networks. It was designed from the start to carry both voice and data traffic over the same transmission channel. The traffic theory community participated very actively in the design process (see e.g. [RMV97]) and the definition of ATM includes, for example, differentiation of traffic (five traffic classes), QoS features and congestion algorithms. Although, ATM was originally proposed to

[1] currently DARPA

replace IP entirely, its present role is quite small. ATM is usually met in backbone networks and as a part of access networks (e.g. ADSL home connections); end-to-end ATM connections are almost completely nonexistent.

1.2 Traffic modeling

One of the principal questions in teletraffic theory is the traffic-performance relation linking network capacity, offered traffic and realized performance. The proper characterization of traffic is an essential requirement. Only after this does the performance analysis of a system make any sense. The traffic characterization should not only include the (statistical) properties seen in measurements, but also an understanding of their origin.

In traditional telephone networks, the standard approach was to model call arrivals by a Poisson process and call durations by an exponentially distributed random variable — these assumptions were validated by several measurements and they remained unchanged for decades. In the early days of data-networks, similar types of models were also used in data network analysis. There were virtually no attempts to validate modeling assumption against the actual data traffic measurements in the 1970s and 1980s. It took almost 20 years before the qualitative difference between voice and data traffic was fully recognized.

The first evidences of the new era were noticed in the local area network (LAN) measurements reported by Wilson and Leland [LW91]. A careful statistical analysis of these measurement [LTWW93, LTWW94] showed that the Ethernet LAN traffic exhibits properties which are inconsistent with the old assumptions of independence, Poisson arrivals and exponentially distributed job sizes. Instead, LAN traffic appeared to be statistically self-similar and have long-range correlations. Process A is called self-similar with self-similarity parameter H if the processes $\alpha^H A_t$ and the time-scaled version of it $A_{\alpha t}$ have the same finite-dimensional distributions for all $\alpha > 0$. If the autocorrelation function decays slowly, usually according to a power-law, then the process is long-range dependent. In practice, these findings mean that there is no characteristic time-scale, and bursts appear in all time resolutions. This, the so-called Bellcore study, triggered a series of follow-ups which all confirmed that the traditional Poissonian models were inadequate for data networks (video streams [BSTW95], Wide Area Network traffic [PF95], World Wide Web downloads [CB97]).

After measurement equipment improved and accurate packet level measurements became possible (the original Bellcore data was accurate to within 20-100μs) even stranger behavior has been observed: traffic exhibits complex scaling laws with much spikier bursts in small time-scales than in the coarser resolutions. Riedi and Lévy Véhel were the first to describe this phenomena using the language of multifractals [RLV97, LVR97]. Subsequently, several papers have reported similar observations (see e.g. [MN97, FGW98, RCRB99]).

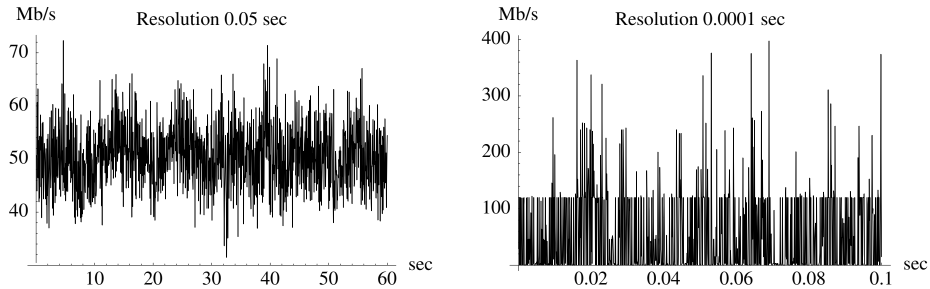


Figure 1. Traffic in an international link of the Finnish University and Research Network (FUNET), measured in Feb, 2001.

The difference between large and short time-scale behavior is clearly seen in Figure 1, where an IP traffic trace measured in an international link of the Finnish University and Research Network (FUNET) is shown in two resolutions. The time series have distinctly different characteristics: the larger time-scale plot has a Gaussian character whereas the high resolution one has more and bigger bursts. (Note that the terms "small" and "large" time-scale depend strongly on the network studied.) Moreover, the latter trace seems to have some preferable deterministic values, which are probably due to the maximum transfer rates of the individual input ports or links. Broadly speaking, aggregation both in time and space, that is, using coarser measurements and more contributing sources, results in self-similarity and long-range dependency. In contrast, multifractality is best observed in a single TCP flow and at small time-scales. This suggests that these phenomena probably do not have the same physical origin.

The observations indicate that new models and explanations are needed. However, there is still a relative large group of teletraffic engineers who are reluctant to accept the change. They often reason that the family of Markov processes is enough for their purposes. Although it is true that the Markovian models are very flexible and, in principle, any process can be approximated by a Markovian one, the number of model parameters easily becomes very large. An even more crucial shortcoming is that these models often fail to provide an insight into the underlying system.

The large-time-scale dynamics of teletraffic is understood reasonable well. A common explanation for the long-range dependence and self-similarity is heavy-tailed transmission times originating mainly from the heavy-tailed file sizes (files in Unix systems [BHK$^+$91], FTP bursts [PF95], WWW-files [CB97]). There are several simple and parsimonious traffic models which capture this phenomenon, like superposition of on-off sources with heavy-tailed sojourn times, or the infinite source Poisson model, also known as the M/G/∞ input model (see e.g. [WTSW97, MRRS02]). When considering the cumulative traffic at large time-scales, these models have the same properties as the observed data traffic: self-similarity and long-range dependence [TWS97, MRRS02].

Understanding the short time-scale dynamics is a more difficult task. Currently, no mathematically rigorous and intuitively appealing construction exists which could fully explain the observations. The networking mechanisms, like protocols and queues, determine how the flows of packets are actually moving in a network. Thus, it is likely that they are responsible for the local irregularities and high variations seen in real traffic (see e.g. [FGHW99, RW00]). Unfortunately, we lack (close-loop) models able to describe the joint dynamics of heterogeneous sources, protocols and queues in the right way.

1.3 Stochastic-process limits

A stochastic-process limit is usually constructed from initial processes by properly scaling time and space. Analyzing the limiting system gives us relevant information about the original system, since the continuous-mapping theorem [Bil68, Bil99] guarantees that most of the performance functionals also converge (see e.g. [Add99, Whi02]). *"Stochastic-process limits often give a macroscopic view of the system and they strip away unessential details and reveal key features determining performance*" [Whi02].

The main reasons to consider Gaussian traffic are the Central Limit Theorem (CLT) and the Functional Central Limit Theorem (FCLT). Intuitively, they express the fact that a sum of many small independent or weakly dependent random variables is approximately normally distributed, given that the variances are finite. For cases with infinite variances, there are results of the same type with a stable random variable or a stable random process as the limit (see e.g. [ST94, Whi02]). For example, in [MRRS02], Mikosch et al. study whether the network traffic can be approximated by fractional Brownian motion or stable Lévy process by determining the limit process for a rescaled infinite Poisson model.

While Gaussian processes are due to addition/aggregation, multifractal processes are usually limits of multiplicative constructions. Unfortunately, such universal limit processes as Gaussian processes, or more generally stable processes, are not yet available in the toolbox of multifractal modeling.

The stochastic limit processes are often pure mathematical objects that may have conflicting or undefined properties compared to the real system studied. For example, Gaussian traffic has positive probability for negative increments and the scaling laws of a multifractal are defined on infinitely small time-intervals. This is not necessarily a problem since one can still determine the functionals of the traffic process, like queue length, even though the physical meaning may be questionable. However, this means that performance results derived from the limit models cannot be more than approximations for the real system.

It is clear that only continuous time processes are feasible models for the small time-scale behavior. Otherwise, the notion of the local scaling laws would be meaningless. Both discrete and continuous time approaches can be used for processes describing the large time-scale properties. However, a time discretization brings an extra artificial parameter, the sampling interval. In order to keep models as simple as possible, we prefer continuous time models.

1.4 Contents of the thesis

The motivation of this thesis comes from traffic modeling and performance analysis of telecommunication systems. However, it does not consider structural modeling, but its emphasis is on analysis of the related stochastic limit processes. The main part of thesis is devoted to large time-scale analysis by considering Gaussian processes and queueing systems with Gaussian input. The family of Gaussian processes with stationary increments serves as our traffic model. We develop a toolbox that makes Gaussian models applicable for real teletraffic applications. This includes a simulation method, a prediction algorithm and performance formulae for Gaussian queueing systems. In order to understand the small time-scale dynamics, the first steps are taken towards general multifractal models offering a suitable basis for teletraffic modeling. We introduce and analyze a novel family of multifractal processes with stationary increments.

In Chapter 2, generation (i.e. simulation) methods for Gaussian processes are studied. We introduce a fast and accurate simulation algorithm that can be used to generate long approximate Gaussian traces. Moreover, the algorithm is also modified to run on-the-fly. We consider only the simulation of fractional Brownian motion, but the same approach can be used for arbitrary Gaussian processes with stationary increments (which is done in the queueing studies summarized in Chapter 3). Comparison with other simulation methods shows that our approach is performing well and has many advantages, such as simplicity and only linearly increasing computational complexity.

Chapter 3 considers queueing systems with Gaussian input traffic. Approximate queue length distributions for ordinary, priority and Generalized Processor Sharing (GPS) queues are derived using a most probable path approach. In multiclass cases, our approach leads either to the empty buffer approximation (a notion first introduced by Berger and Whitt [BW98a, BW98b]) or to new heuristics called Rough Full Link Approximation. Moreover, some results concerning the buffer emptiness probabilities and lower bounds of the queue length distribution are given. Simulation studies show that the performance formulae appear to be quite accurate over the full range of buffer levels. Although most of the performance formulae are just heuristic approximations, they are accurate enough, for example, for network dimensioning purposes. One of the advantages is that the approach is easily implemented as a tool that can be used without a profound knowledge of Gaussian processes.

Linear minimum-least-square-error predictors are studied in Chapter 4. First, the optimal way to measure the past of a fractional Brownian motion (FBM) is determined. We find that conditioning with respect to geometrically increasing intervals is a good engineering solution and a much better way than equally spaced measurements. This result holds for any process having an FBM-type variance function. Then we introduce a semi-stationary predictor, which uses a constant variance function and a moving average based mean rate estimation. The performance of the predictor is tested with real traffic traces. The algorithm works according to the analytical expectation even in a non-stationary case. However, a combination of a resource reservation algorithms and the predictor may result in unexpected behavior.

In Chapter 5, we introduce a family of multifractal processes which belongs to the framework of T-martingales and multiplicative chaos introduced by Kahane [Kah85, Kah87, Kah89]. The family has many desired properties like stationarity of increments, concave multifractal spectra, and simple construction. In its simplest form, our model is based on the multiplication of independent rescaled processes $\Lambda^{(i)}(\cdot) \stackrel{dist}{=} \Lambda(b^i \cdot)$. We study properties of the limit process $\lim_{n\to\infty} A_n$, where $A_n(t) = \int_0^t \prod_{i=0}^n \Lambda^{(i)}(s)\mathrm{d}s$. For example, we derive conditions for non-degeneracy, establish a power law for the moments and obtain a formula for the multifractal spectrum.

2. Generating Gaussian traces

2.1 Introduction

Recent measurement studies have shown that burstiness of packet traffic is associated with long-range correlations that can be efficiently modeled in terms of fractal or self-similar processes, like fractional Brownian motion (FBM). To gain a better understanding of queueing and network-related performance issues based on simulations, as well as to determine network element performance and capacity characteristics based on load testing, it is essential to be able to accurately and quickly generate long traces from FBM processes. In addition to that, almost all performance formulae for a queueing system with Gaussian traffic are either asymptotic or heuristic results. For that reason, these simulation traces also have an important role in accuracy checking purposes.

FBM has a special role in telecommunications and numerous approximate algorithms for simulating FBM have been proposed: a short memory approximation by Mandelbrot [Man71], a queueing based method using the M/G/∞ queue length with Poisson arrivals and heavy-tailed service times by Cox [Cox84], a fast Fourier transform based method by Paxson [Pax97], a wavelet transform based method by Flandrin, Abry and Sellan [Fla92, AS96], aggregation of a large number ON-OFF sources with infinite-variance sojourn times by Willinger et al. [WTSW97], to mention just some of the known approaches.

The simulation algorithm of Publication [1] belongs to the bisection methods which are based on generating points of the process path "top-down" by properly interpolating from the existing points. The simplest version of the bisection methods is the random midpoint displacement (RMD) whose performance was analyzed by Lau et al. [LEWW95].

In a recent thesis by Dieker [Die02], different simulation methods, including the method of Publication [1], are studied. His findings are in accordance with those presented in [1].

2.2 Basics of Gaussian processes

In this thesis, we consider only Gaussian processes with stationary increments. A process is called Gaussian if all its finite-dimensional distributions are multivariate Gaussian. Let $A = (A_t)_{t\in\mathbb{R}}$ be a Gaussian process and set $A_0 = 0$. The stationarity of the increments means that for any $t_0 \in \mathbb{R}$ the processes A and $(A_{t+t_0} - A_{t_0})_{t\in\mathbb{R}}$ have the same finite-dimensional distributions.

We can represent process A in the form

$$A_t = mt + Z_t,$$

where $m = \mathrm{E} A_1$ and Z is a centered Gaussian process, i.e., $\mathrm{E} Z_t = 0$, with $\mathrm{Var} Z_t = v(t)$. A Gaussian process with stationary increments is uniquely defined by its mean and variance function. For example, the covariance of A is given simply by

$$\mathrm{Cov}(A_t, A_s) = \mathrm{Cov}(Z_t, Z_s) = \frac{1}{2}(v(t) + v(s) - v(t-s)).$$

It is clear that it is enough to simulate the centered process Z_t.

An interesting special case is the fractional Brownian motion. The normalized FBM has variance $v(t) = t^{2H}$, where the self-similarity parameter $H \in (0,1)$. Its single most important property is self-similarity: the processes $Z_{\alpha t}$ and $\alpha^H Z_t$ have the same path space distributions for any $\alpha > 0$. If $H > \frac{1}{2}$, then FBM is long range-dependent.

Since all the finite dimensional distributions of Z are Gaussian (by the definition), one could generate theoretically exact traces at any finite resolution just by using the conditional multinormal distributions or a direct Cholesky decomposition of the correlation matrix. Unfortunately, the computational complexities of the direct algorithms are $\mathcal{O}(N^2)$, where N denotes the number of steps. Using the discrete Fourier transformation, the number of operations can be reduced to $\mathcal{O}(N \log N)$ (see e.g. [DH87]). However, the exact generation of very long traces is infeasible in practice due to the amount of storage and CPU time required.

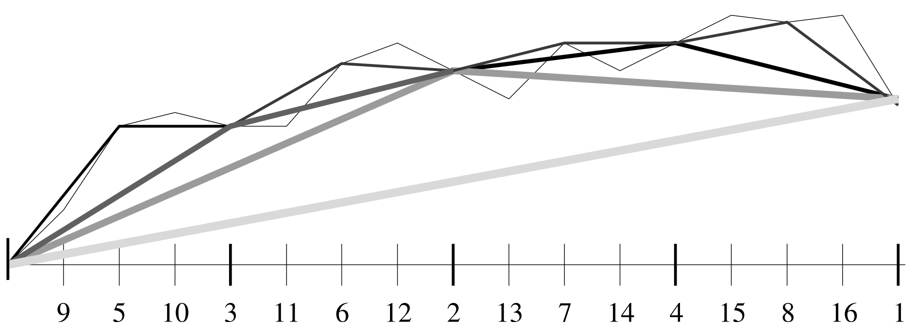


Figure 2. Bisection algorithm: the numbers denote the order of the point generation.

The bisection approach works by progressively subdividing the interval over which the sample path is to be generated. Let us consider a Gaussian process on the unit interval $[0,1]$. The trace is generated in the following order: Z_1, $Z_{0.5}$, $Z_{0.25}$, $Z_{0.75}$, $Z_{0.125}$ and so on (see Figure 2). In the basic RMD algorithm, we only condition with respect to the values of the endpoints of the subdivided interval and the value of the midpoint is drawn from the corresponding conditional Gaussian distribution. For example, $Z_{0.75}$ is drawn conditioning with respect to the earlier generated values $Z_{0.5}$ and Z_1. The generator proposed in Publication [1] is a natural generalization of RMD; conditioning is taken with respect to m nearest points to the left and n nearest points to the right. Although [1] considers only FBM, the conditionalized RMD can be used to generate arbitrary Gaussian processes with stationary increments.

A common problem with the approximate methods is that it is very difficult or even impossible to state any analytical results on the accuracy of a specific method. Usually, only empirical accuracy tests can be used in order to show that a method is functioning properly. This also holds true for our RMD_{mn} algorithm. The recent paper by Dzhaparidze and van Zanten [DZ02] introduces a new series presentation for FBM

$$Z_t = \sum_{n=1}^{\infty} \frac{\sin x_n t}{x_n} X_n + \sum_{n=1}^{\infty} \frac{1-\cos y_n t}{y_n} Y_n,$$

where the X_n and the Y_n are independent Gaussian random variables with mean zero and variances decreasing as a function of n, and the x_n and the y_n are zeros of certain Bessel functions. This result might offer a novel basis for FBM generation with a proper control over the accuracy of the traces.

2.3 Summary of [1]

This paper proposes a new method for generating approximate fractional Brownian motion traffic: a generalization of random midpoint displacement called RMD_{mn}. The algorithm is a bisection method where the process path is generated “top-down” by properly interpolating (i.e. conditioning) the existing points. Moreover, it is shown that the basic algorithm can be modified to produce long accurate FBM traces on-the-fly with memory requirements increasing only at the logarithmic speed. The role of parameters m and n (numbers of interpolated points from the left and from the right) is studied in several ways: calculating the relative ℓ_1 error, estimating how close the Hurst parameter is to the desired one and studying the computational complexity. The analysis suggests that quality of the generated traces is reasonable good even with as small values as $m = 2, n = 1$. Finally, RMD_{mn} is compared with two popular generation methods: a fast Fourier transformation (FFT) method and an aggregation method. We find that the FFT method is comparable to the RMD_{mn} algorithm in terms of quality and time complexity, however, the FFT method is strictly a top-down algorithm, that is, the whole trace has to be generated before it can be used in any application. The aggregation method requires a couple of orders of magnitude of time periods to converge to the asymptotic results, and is therefore much slower.

2.4 Author's contribution to [1]

The present author is responsible for the implementation of the RMD_{mn} algorithms (both the pure top-down and on-the-fly), design and complexity analysis of the on-the-fly algorithm, and the pathwise accuracy analysis of the traces. The idea of the RMD generalization and the on-the-fly algorithm is by Norros, comparison with other algorithms is done by Wang. The paper is jointly written, Norros being the corresponding author.

3. Queueing systems with Gaussian input

3.1 Introduction

Gaussian processes play a special role in performance analysis since their use can be motivated by the central limit theorem type of arguments. For example, Addie [Add99] argues that as networks grow and the number of traffic streams multiplexed on a single link increases, the shape of traffic must become closer and closer to Gaussian. In order to apply Gaussian models in performance analysis of a real-life network, one must first validate the model assumptions: adequate aggregation, and (approximate) stationarity, independence and homogeneity of the sources. After that, parameters can be estimated and the goodness of the Gaussian approximation can be checked. How to perform all that properly is not yet fully clear. In the case of 1-dimensional marginal distributions, Kilpi and Norros [KN02] have determined the minimal source and time aggregation levels needed to make a Gaussian approximation at least theoretically reasonable. Moreover, they also consider tests which can be used to measure the goodness of a Gaussian fit. Kilpi [Kil03] is currently trying to extend methods to multivariate Gaussian distributions.

Gaussian processes are not the only feasible models for traffic aggregates. There are cases where the limit process should have jumps and much heavier marginal distributions than Gaussian. Then the proper limit process could be a stable Lévy process (see e.g. [KSJ98, MRRS02]) or something in between Gaussian and Lévy processes [GK02]. If we relax the independence and/or homogeneity assumptions, the possibilities become virtually endless.

From the point of view of queueing theory, the assumption of Gaussian input is never fully acceptable, because there is always a positive probability of negative input, which is nonsense from practical point of view and destroys many classical arguments on the theoretical side. In a Gaussian framework, the rigorous constructions of queueing theory must be replaced by analogously defined functionals of a Gaussian process. Moreover, it seems that general results on the distributions of these functionals cannot be better than inequalities and limit theorems. In this thesis, we must often be satisfied with heuristic approximations only.

In early 1990s, Norros introduced his famous FBM storage model [Nor93, Nor94]. Following this, the FBM queue, and especially its tail behavior, have been extensively studied by several authors (see e.g. [Nar98, Nor99, MS99, Pit01]). Currently, the behavior of the ordinary, that is, single class, queue with FBM input is known reasonably well.

In this thesis, the analysis is extended to general Gaussian input processes and multi-class queues. In addition to ordinary queues, we study priority and generalized processor sharing (GPS) systems. Berger and Whitt [BW98a, BW98b] are probably the first who studied Gaussian priority systems. In the discrete case, the large deviations of Gaussian priority queues are considered by Wischick [Wis01]. Recently, Mandjes and Uitert [MU02] have proved some exact large deviation results for the tandem and priority queues with continuous Gaussian processes. The results of Mandjes and Uitert are in accordance with our heuristics.

3.2 Queueing systems

The basic traffic modeling idea is to approximate the cumulative traffic process in a telecommunication system by a continuous Gaussian process $A = (A_t)_{t\in\mathbb{R}}$ with stationary increments and $A_0 \equiv 0$ (see Section 2.2). For $s < t$, $A(s,t) \doteq A_t - A_s$ presents the amount of traffic in time interval $(s,t]$. We denote

$$A_t = mt + Z_t,$$

where Z is a centered Gaussian process with variance $v(t) = \mathrm{Var}(Z_t)$ and covariance $\Gamma(s,t) \doteq \mathrm{E}\, Z_s Z_t = \frac{1}{2}(v(t) + v(s) - v(t-s))$. The ordinary queue with a constant rate server and an infinite buffer is defined by

$$Q_t = \sup_{s\le t}(A(s,t) - c(t-s)),$$

where c is the server rate.

The extension to the multiclass system is simple. Let the input traffic consist of k classes, and denote the cumulative arrival process of class $j \in \{1,\dots,k\}$ by $(A_t^{\{j\}})_{t\in\mathbb{R}}$. For the superposition of a set of traffic classes $J \subseteq \{1,\dots,k\}$ we write

$$A_t^J \doteq \sum_{j\in J} A_t^{\{j\}}$$

and use the similar superscript notation also for other quantities defined later. Moreover, the processes $A^{\{j\}}$ are assumed to be independent, continuous Gaussian processes with stationary increments. As above, denote $A_t^{\{j\}} = m_j t + Z_t^{\{j\}}$, $A^{\{j\}}(s,t) \doteq A_t^{\{j\}} - A_s^{\{j\}}$, $v_j(t) = \mathrm{Var}\, Z_t^{\{j\}}$ and $\Gamma_j(s,t) = \mathrm{E}\, Z_s^{\{j\}} Z_t^{\{j\}}$.

Assume that the classes are numbered with descending priority, with class 1 having the highest priority. Since our model is continuous, there is no distinction between preemptive and non-preemptive priority. Lower class traffic does not disturb upper class traffic[1] and a simple approach is the following: Define $Q^{\{1\}}$,

[1] This is not necessarily true in Gaussian queues; depending on how the negative traffic is interpreted there may be interaction also from lower class traffic to an upper class queueing process, see [6].

$Q^{\{1,2\}}$, $Q^{\{1,2,3\}}$ etc. as ordinary queues served at rate c and with inputs $A^{\{1\}}$, $A^{\{1,2\}}$, $A^{\{1,2,3\}}$, etc., respectively. Then the class-wise queues are given by

$$\begin{aligned} Q^{\{2\}} &= Q^{\{1,2\}} - Q^{\{1\}}, \\ &\cdots \\ Q^{\{k\}} &= Q^{\{1,\ldots,k\}} - Q^{\{1,\ldots,k-1\}}. \end{aligned} \tag{3.1}$$

In a GPS system, all classes are given a guaranteed server rate $\mu_j c$, with $\sum_j \mu_j = 1$. In the case of unlimited buffers, the queue of class i, $Q_t^{\{i\}}$, and the total queue $Q_t = \sum_{i=1}^k Q_t^{\{i\}}$ satisfy [Mas99]

$$\begin{aligned} Q_t^{\{i\}} &= \sup_{s\le t}(A^{\{i\}}(s,t) - \mu_i c T(s,t)), \\ Q_t &= \sup_{s\le t}(\sum_{i=1}^k A^{\{i\}}(s,t) - c(t-s)), \end{aligned} \tag{3.2}$$

where $T(s,t) = T_t - T_s$ and T_t is a non-decreasing stochastic process with $T_0 \equiv 0$. Thus, $\mu_i c T(s,t)$ presents in a sense the amount of potential service for each class i in time interval $(s,t]$. If all the classes are queueing on interval $[s,t]$, then $T(s,t) = t-s$, i.e., everyone gets exactly its guaranteed service. Otherwise, the role of $T(s,t)$ is to redistribute the excess capacity.

Note that since the input processes have stationary increments, the queue processes are stationary in all three cases: in ordinary, priority and GPS systems.

3.3 Most probable paths and queue length distributions

The main goal of this part of the thesis is to find approximations for the queue length distributions

$$\mathrm{P}\left(Q^{\{i\}} > x\right), \quad i = 1,\ldots,k.$$

Moreover, the approximations should be valid over the full range of buffer levels. In the case of one class, it is often possible to derive exact asymptotics or asymptotical bounds for the equivalent distribution

$$\mathrm{P}\left(\sup_{t>0}(A_t - ct) > x\right)$$

(see e.g. [Nar98, DMR98, CS99, HP99, Pit01, DĘ02]). Unfortunately, methods like the Fourier expansion and the double sum method (see [Pit96]) are quite complicated and their applicability to the multi-class case is questionable. In contrast, large deviation based analysis is less accurate, but usually the outcome is also much simpler.

The reproducing kernel Hilbert space (RKHS) of a Gaussian process (see e.g. [Adl90]) plays an important role in the path space large deviations. Basically, the RKHS is defined by the covariance function $\Gamma(\cdot,\cdot)$. Let us consider a centered Gaussian process Z on T. Start with the set

$$S = \left\{ f \ : \ f(\cdot) = \sum_{i=1}^{n} a_i \Gamma(s_i,\cdot),\ a_i \in \mathbb{R},\ s_i \in T,\ n \geq 1 \right\}$$

and define an inner product on S by

$$\langle f, g \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \Gamma(t_i, s_j),$$

where $f(\cdot) = \sum_{i=1}^{n} a_i \Gamma(t_i,\cdot)$ and $g(\cdot) = \sum_{i=1}^{m} b_i \Gamma(s_i,\cdot)$. It easy to see that the inner product has the "reproducing property": for all f in S,

$$f(t) = \langle f, \Gamma(t,\cdot) \rangle. \tag{3.3}$$

The closure of S under the norm $\|f\|_R = \langle f, f \rangle^{\frac{1}{2}}$ is called the RKHS of Z. In this thesis, we denote this space by R.

A large deviation principle for Gaussian measures in Banach space is given by the generalized Schilder's theorem (Bahadur and Zabell [BZ79], see also [Aze80, DS89]). It gives us the following logarithmic asymptotics: If Z is a centered Gaussian process, then

$$\text{for } F \text{ closed in } \Omega: \qquad \limsup_{n\to\infty} \ \frac{1}{n} \log \mathrm{P}\left(\frac{Z}{\sqrt{n}} \in F \right) \ \leq - \inf_{\omega \in F} I(\omega),$$

$$\text{for } G \text{ open in } \Omega: \qquad \liminf_{n\to\infty} \ \frac{1}{n} \log \mathrm{P}\left(\frac{Z}{\sqrt{n}} \in G \right) \ \geq - \inf_{\omega \in G} I(\omega),$$

where the function $I : \Omega \to \mathbb{R} \cup \{\infty\}$,

$$I(\omega) = \begin{cases} \frac{1}{2}\|\omega\|_R^2, & \text{if } \omega \in R, \\ \infty, & \text{otherwise.} \end{cases}$$

In the general case, it is almost impossible to decide whether an arbitrary function belongs to the RKHS. Only some special cases have been studied. Let us consider centered Gaussian distributions on $\mathbb{R}^N$, i.e., multivariate Gaussian distributions, in details. Denote $\mathbf{f} = (f_1, \ldots, f_N)$ and $\mathbf{g} = (g_1, \ldots, g_N)$. Assuming a non-singular covariance matrix Γ, arbitrary $\mathbf{f}$ and $\mathbf{g}$ can be represented as $\mathbf{f} = \sum_{i=1}^{N} a_i \Gamma_i$ and $\mathbf{g} = \sum_{i=1}^{N} b_i \Gamma_i$, where Γ_i denotes the ith column of Γ. Thus,

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i \Gamma_{i,j} b_j = \mathbf{a}^{\mathrm{T}} \Gamma \, \mathbf{b} = \mathbf{f}^{\mathrm{T}} \Gamma^{-1} \mathbf{g}$$

and the space R is $\mathbb{R}^N$ itself, but equipped with an inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \mathbf{f}^{\mathrm{T}} \Gamma^{-1} \mathbf{g}$. Since the density function of a multivariate Gaussian vector $\mathbf{Z}$ is given by

$$\text{const } e^{-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\Gamma^{-1}\mathbf{z}} = \text{const } e^{-\frac{1}{2}\|\mathbf{z}\|_R^2},$$

minimizing the R-norm means maximizing the density.

The other simple case is Brownian motion on $[0,1]$ (see e.g. [Adl90]); then the RKHS consists of functions which are absolutely continuous, vanishing at 0, and whose derivative belongs to L_2. The inner product is given by

$$\langle f, g \rangle = \int_0^1 f'(s) g'(s) ds.$$

The RKHSs of fractional Brownian motion have been determined by Decreusefond [Dec00]. The RKHS of a FBM with self-similarity parameter H is a space of functions which are $(H + \frac{1}{2})$ differentiable, vanishing at 0 and whose $(H + \frac{1}{2})$-th derivative belongs to L_2.

Although the full characterization of a RKHS is usually difficult, finding the minimizing element in a given subset — the most probable path of that set — is simple if the set can represented in the form of $F = \{ f \in R \ : \ f(t) = y \}$. By the reproducing property (3.3)

$$F = \{ f \in R \ : \ f(t) = y \} = \{ f \in R \ : \ \langle f, \Gamma(t, \cdot) \rangle = y \}.$$

On the other hand,

$$\operatorname{argmin} \{ \| f \|^2 \ : \ \langle f, g \rangle = y \} = \frac{y}{\|g\|^2} g$$

holds in an arbitrary Hilbert space. Thus the most probable path in F is

$$f_{t,y}(\cdot) = \frac{y}{\Gamma(t,t)} \Gamma(t, \cdot) = \frac{y}{v(t)} \Gamma(t, \cdot). \tag{3.4}$$

In a single class queueing system, the most probable path leading to buffer occupancy x or more can be determined using (3.4). Since

$$\begin{aligned} \{Q_0 \geq x\} &= \left\{ \sup_{t<0} (A(t,0) - c(0-t)) \geq x \right\} = \left\{ \sup_{t<0} (-Z_t + ct) \geq x \right\} \\ &= \bigcup_{t<0} \{ -Z_t \geq c|t| + x \} = \bigcup_{t<0} \bigcup_{y \geq x + c|t|} \{ Z_t = -y \}, \end{aligned}$$

the minimization can be done in three steps. First determine the most probable path for $Z_t = -y$ which is by (3.4) $f_{t,y} = \frac{-y}{v(t)} \Gamma(t, \cdot)$. Then minimize the corresponding R norm $\| f_{t,y} \|_R^2 = \frac{y^2}{v(t)}$ with respect to $y \geq x + c|t|$ and $t < 0$. The solution is shown in Section 3.4.

Motivated by the sample path large deviations, we seek the most probable paths for the class i buffer exceedance events, that is, we try to find path

$$f_x^{(i)} = \operatorname{arginf}\left(\|f\|_R^2 \ : \ f \in R,\ Q_0^{\{i\}}(f) > x\right).$$

But instead of taking limits, we approximate

$$\mathrm{P}\left(Q^{\{i\}} > x\right) = \mathrm{P}\left(Q_0^{\{i\}} > x\right) \approx e^{-I(f_x^{(i)})}. \tag{3.5}$$

In certain multi-class cases, we are unable to determine the most probable paths. Then the exact buffer exceedance sets are replaced by approximate sets in which we can solve the minimization problem.

Identifying most probable paths is also interesting with its own rights — it is like "seeing what really happens" when the rare event occurs. For ordinary queues, this has mainly heuristic value, but we show that identifying these paths has an essential role in choosing a good approximation in the case of priority and GPS queues.

For reasons of clarity, a consistent notation is used in the following summarizing sections differing from the notations used in the publications.

3.4 Summary of [2]

Publication [2] considers ordinary single class queues with general Gaussian input. First the mathematical setup, i.e., the path space of the Gaussian processes, the reproducing kernel Hilbert space and generalized Schilder's theory are reviewed. Then the basic approximation (3.5) is introduced. We show that the most probable paths of Z in $\left\{Q_0 \geq x\right\}$ have the form

$$f_{x,t_x}(\cdot) = -\frac{x + (c-m)(-t_x)}{v(t_x)}\Gamma(t_x, \cdot),$$

where $t_x < 0$ is the value of t which minimizes the expression

$$\frac{(x + (c-m)|t|)^2}{v(t)}.$$

The corresponding basic approximation is given by

$$\mathrm{P}(Q > x) \approx \exp\left(-\frac{(x + (c-m)|t_x|)^2}{2v(t_x)}\right).$$

A differentiable Gaussian process (with stationary increments) can be presented in the form $Z_t = \int_0^t X_s ds$, where X_s is a continuous stationary zero-mean Gaussian process with standard deviation σ. For these cases, we derive upper and lower bounds for the probability of buffer emptiness. It appears that those bounds are close to $2\mathrm{P}(X_t > c-m)$, i.e.,

$$\mathrm{P}(Q>0) \approx 2\overline{\Phi}((c-m)/\sigma),$$

where $\overline{\Phi}$ denotes the complementary standard normal distribution function.

The basic approximation, its rescaled version (based on the estimate of the buffer emptiness probability), and an exact lower bound

$$\mathrm{P}(Q>x) \geq \overline{\Phi}\left(\frac{x-(c-m)t_x}{\sqrt{v(t_x)}}\right)$$

are compared with simulations. Both short-range and long-range dependent Gaussian processes are considered. The accuracy of the approximations is quite good over the full range of buffer lengths. Moreover, the buffer emptiness probabilities behave according to our formulae.

3.5 Summary of [3]

Publication [3] also considers the ordinary single class queue with Gaussian input. It is an extended version of [2]. In addition to rigorous proofs of the statements of [2] and repeated simulation studies, there are several new mathematical results.

We show that the condition

$$\lim_{t\to\infty} \frac{v(t)}{t^{\alpha}} = 0$$

for some $\alpha < 2$ guarantees that the input process satisfies

$$\lim_{t\to\pm\infty} \frac{Z_t}{1+|t|} = 0 \text{ almost surely.}$$

This means that the queueing process is finite almost surely if $m < c$.

The general properties of the most probable paths are studied carefully. The following properties are proved: antisymmetry of the input process around the time point $t_x/2$, emptiness of the queue at t_x, non-emptiness of the queue on $(t_x, 0]$ and draining of the queue just after 0.

Using a counter-example, it is shown that the most probable paths are not necessarily unique. Moreover, there are several illustrating examples. Those include periodic Gaussian processes and multiplexing heterogeneous Gaussian sources.

3.6 Summary of [4]

Publication [4] considers priority queues with Gaussian input. The basic approximation is extended to the multiclass case by introducing the corresponding reproducing kernel Hilbert space for k independent Gaussian processes. Assuming that a higher class queue does not see lower class traffic, it is enough to consider only 2-class priority systems. Moreover, the highest class queue behaves as an ordinary queue served at rate c.

The simplest problem is to find the most probable paths of both classes such that the combined (i.e. total) queue reaches a value x. The most probable paths in $\{Q_0^{\{1,2\}} \geq x\}$ have the form

$$g_{t_x,x}(\cdot) \quad \doteq \quad -\frac{x+(c-m_1-m_2)(-t_x)}{v_1(t_x)+v_2(t_x)}(\Gamma_1(t_x,\cdot),\Gamma_2(t_x,\cdot)),$$

where $t_x < 0$ is the value of t which minimizes the expression

$$\frac{(x+(c-m_1-m_2)|t|)^2}{v_1(t)+v_2(t)}$$

and the basic approximation is given by

$$\mathrm{P}\left(Q^{\{1,2\}} > x\right) \approx \exp\left(-\frac{(x+(c-m_1-m_2)|t_x|)^2}{2(v_1(t_x)+v_2(t_x))}\right). \tag{3.6}$$

The analysis of the lower class queue must be done in two steps. First, we determine the most probable paths for the total queue problem as above. If the paths are such that at $t=0$ only class 2 traffic is queueing, then it follows that these paths are also the most probable ones to achieve level x in the class 2 queue alone and $\mathrm{P}\left(Q^{\{2\}} > x\right)$ can be approximated by (3.6). This happens in most priority systems. However, if this is not the case, the situation is more difficult and the most probable paths remain generally unknown (Mandjes and Uitert [MU02] have some new results on this topic). As a heuristic solution, we suggest the Rough Full Link Approximation (RFLA). During $[t,0]$, $t<0$, class 1 offers in total the amount $-ct$ of traffic, class 2 offers in total the amount x of traffic, and $[\tilde{t}_x,0]$ is the most probable interval under the above conditions. The most probable paths satisfying the RFLA are given by

$$\phi^{\mathrm{RFLA}}_{\tilde{t}_x,x}(\cdot) = \left(\frac{(c-m_1)\tilde{t}_x}{v_1(\tilde{t}_x)}\Gamma_1(\tilde{t}_x,\cdot), -\frac{x+m_2\tilde{t}_x}{v_2(\tilde{t}_x)}\Gamma_2(\tilde{t}_x,\cdot)\right),$$

where $\tilde{t}_x < 0$ is the value of t which minimizes the norm

$$\|\phi^{\mathrm{RFLA}}_{t,x}\|^2_R = \frac{(c-m_1)^2t^2}{v_1(t)} + \frac{(x+m_2t)^2}{v_2(t)},$$

and the basic estimate is given by $P(Q^{\{2\}} > x) \approx \exp\big(-\frac{1}{2}\|\phi^{\mathrm{RFLA}}_{\tilde{t}_x,x}\|_R^2\big)$. For the events satisfying the RFLA conditions we have an exact lower bound

$$\mathrm{P}(\mathrm{RFLA}_x) \geq \overline{\Phi}\left(-\frac{(c-m_1)\tilde{t}_x}{\sqrt{v_1(\tilde{t}_x)}}\right)\overline{\Phi}\left(\frac{x+m_2\tilde{t}_x}{\sqrt{v_2(\tilde{t}_x)}}\right).$$

In the simulation studies we find the following: If the most probable paths are such that the total queue consists only of the lower class queue, then class 2 and total queue (empirical) distributions are indistinguishable and the basic approximations work as well as in the single class case. If RFLA is needed then the accuracy of the approximations decreases being, however, still qualitatively rather good.

3.7 Summary of [5]

In publication [5], the most probable path approach is extended to GPS queues with Gaussian input. In addition to demonstrating the applicability of the performance formulae, we show that a GPS system with Gaussian input is very sensitive; in some regions small changes of parameters may change the performance a lot. Moreover, mean rate based weight assignments do not usually give desired results. It is important to take into account the variance structure carefully, since one cannot assume similar queueing behavior — not even qualitatively — for processes which have different types of variance functions.

The basic setup is as in [4]. The combined queue and the corresponding input processes are the same as in the priority systems, only the class-wise queueing paths differ (due to the service guarantees). Also similarly, when considering a specific class, say class 1, there are two possibilities: either the most probable total queue of size x consists only of class 1 traffic, or other classes are also contributing. In the previous case, the distributions of the combined queue and the class 1 queue will be almost indistinguishable. The former case is dealt with the Rough Full Link Approximation.

We define RFLA for a GPS queue with two input classes as follows. During $[t,0]$, $t<0$, class 1 offers in total the amount $x+\mu_1 c|t|$ of traffic, class 2 offers in total the amount $\mu_2 c|t|$ of traffic, and $[t_x,0]$ is the most probable interval under the above conditions. The most probable paths satisfying the RFLA conditions are of the form

$$\phi^{\mathrm{RFLA}}_{\tilde{t}_x,x}(\cdot) = \begin{pmatrix} \frac{-x+(\mu_1 c-m_1)\tilde{t}_x}{v_1(\tilde{t}_x)}\,\Gamma_1(\tilde{t}_x,\cdot) \\ \frac{(\mu_2 c-m_2)\tilde{t}_x}{v_2(\tilde{t}_x)}\,\Gamma_2(\tilde{t}_x,\cdot) \end{pmatrix},$$

where $\tilde{t}_x < 0$ is the value of t which minimizes the expression

$$\|\phi^{\mathrm{RFLA}}_{t,x}\|_R^2 = \frac{(x-(\mu_1 c-m_1)t)^2}{v_2(t)} + \frac{(\mu_2 c-m_2)^2 t^2}{v_2(t)},$$

and the basic approximation is given by $P(Q^{\{1\}} > x) \approx \exp\big(-\frac{1}{2}\|\phi^{\text{RFLA}}_{\tilde{t}_x,x}\|^2_R\big)$. For the events satisfying the RFLA conditions, denoted by RFLA_x, we have an exact lower bound

$$\text{P}(\text{RFLA}_x) \geq \overline{\Phi}\left(-\frac{x-(\mu_1 c - m_1)\tilde{t}_x}{\sqrt{v_1(\tilde{t}_x)}}\right)\overline{\Phi}_1\left(-\frac{(\mu_2 c - m)\tilde{t}_x}{\sqrt{v_2(\tilde{t}_x)}}\right).$$

The rescaling of the basic formulae can be done so that the buffer emptiness is approximated by a worst case scenario. The class i queue cannot be larger than the total queue. On the other hand, class i queue is always smaller than the queue of a single class system with input $A^{\{i\}}$ and server rate $\mu_i c$. Applying the non-emptiness approximation to these two non-priority queues gives

$$\text{P}\left(Q^{\{i\}} > 0\right) \lesssim 2\min\left\{\overline{\Phi}\left(\frac{\left(c-\sum m_j\right)\Delta t}{\sqrt{\sum v_j(\Delta t)}}\right), \overline{\Phi}\left(\frac{\left(\mu_i c - m_j\right)\Delta t}{\sqrt{v_i(\Delta t)}}\right)\right\},$$

where Δt denotes the smallest relevant time resolution of a Gaussian model.

In addition to testing the accuracy of the approximate formulae (which is again rather good), the simulation studies are used to study how the GPS parameters μ_i affect the queueing performance.

3.8 Summary of [6]

Publication [6] is a unified presentation of the most probable path approach for priority and GPS queues. In addition to the main results of [4] and [5], there is also some new material.

The problem of negative traffic is discussed. In Equations (3.2) the negative traffic is considered as an extra service capacity to be shared among other classes. Instead of Equations (3.1), one could define also a priority system where negative traffic plays a role.

The lower bound for the events satisfying the RFLA condition is improved using a half-space approximation: our second RFLA approximation, which the simulations indeed indicate to be a lower bound, is

$$\text{P}(\text{RFLA}_x) \geq \overline{\Phi}\left(\sqrt{\frac{(x-(\mu_1 c - m_1)\tilde{t}_x)^2}{v_1(\tilde{t}_x)} + \frac{(\mu_2 c - m_2)^2\,\tilde{t}_x^2}{v_2(\tilde{t}_x)}}\right).$$

New examples and simulations are added, including a case of two FBM input classes, where everything can be calculated analytically.

3.9 Author's contribution to [2]–[6]

In [2], the present author modified the conditionalized RMD algorithm for general Gaussian processes, performed all the simulations and calculated the corresponding performance estimates for the examples. The paper is jointly written. In [3], most of the new material (compared to [2]) is by the present author. This includes especially the rigorous mathematical analysis of the most probable paths and several illustrating examples. The paper is jointly written.

In [4]–[6], the basic ideas of how to apply the most probable path approach to priority and GPS queues, as well as the heuristic rough full link approximation, are by Norros. The actual analysis and accuracy testing are mainly by the present author. [4] and [5] are written by the present author under the guidance of Norros. In [6], Norros was the corresponding author, the present author also contributing to the content; especially the examples and simulations are mostly by the present author.

4. Prediction of teletraffic

4.1 Introduction

Although the capacities of networks are increasing continuously, there are always some applications and sub-networks where resources are scarce. In these situations, even a slight improvement by predicting the future traffic load can make a difference.

After the discovery of long-range dependence of data traffic in various network environments there have been great — even unrealistic — expectations concerning how one could utilize this property. The theoretical basis for predictors of long-range dependent traffic can be found in the paper on prediction of fractional Brownian motion by Gripenberg and Norros [GN96] (see also [Nor95]). Their study shows that the past of a process is relevant only as far as we are going to predict the future, for example, the next second is mainly predicted by the previous one. Lately, there have been several studies which, using time series approaches, have considered traffic prediction and related problems (see e.g. [ÖS01, ÖSH01, PT01, SL02]).

4.2 Basics of the linear predictions

The basic setup for a prediction problem is the following: we have observed random variable $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}} \in \mathbb{R}^n$ and using this knowledge we would like to predict another random variable $Y \in \mathbb{R}$. The natural predictor would be $\mathrm{E}\,(Y|\mathbf{X})$. Unfortunately, this is impractical in many cases due to the need for exact knowledge of the (joint) distributions.

If we assume that $Y, X_i \in \mathscr{L}_2$, $i = 1, \ldots, n$, we can consider the classical regression problem: given any closed subset M of $\mathscr{L}_2$, find the element of this subset which minimizes the mean-square distance from Y. In time series analysis, this element is often called the best predictor in M (see e.g. [BD91]). In this thesis, we consider only linear predictions: if M is the closed linear space spanned by $(X_1, \ldots, X_n)$, then the best prediction of Y in M is found by determining $\mathbf{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$ and μ such that

$$\mathrm{E}\,\left(Y - (\mathbf{a}^{\mathrm{T}}\mathbf{X} + \mu)\right)^2$$

is minimized. The well known solution (see e.g. [BD91]) is

$$\mathbf{a} = \Gamma(\mathbf{X},\mathbf{X})^{-1}\Gamma(\mathbf{X},Y), \qquad \mu = \mathrm{E}\,Y - \mathbf{a}^{\mathrm{T}}\,\mathrm{E}\,\mathbf{X},$$

where $(\Gamma(\mathbf{X},\mathbf{X}))_{i,j} = \mathrm{Cov}\left(X_i, X_j\right)$ and $(\Gamma(\mathbf{X},Y))_i = \mathrm{Cov}(X_i, Y)$.

If $(\mathbf{X}, Y)$ is a multivariate Gaussian random variable, then the minimum-least-square-error predictor of Y is also the conditional expectation of Y with respect to $\mathbf{X}$. Thus, the linear $\mathscr{L}_2$ predictors suit especially well for Gaussian processes.

Let us consider the cumulative traffic model

$$A_t = mt + \sigma Z_t, \tag{4.1}$$

where m is the mean rate, σ^2 is the variance in a unit time and Z_t is a centered (i.e. $\mathrm{E}\,Z_t = 0$) stochastic process with stationary increments and $\mathrm{Var}\,Z_1 = 1$. Moreover, it is assumed that $Z \in \mathscr{L}_2$. The prediction problem of A can be represented as follows: the past of A is given by a finite number of measurements

$$X_i = A_{t-s_i} - A_{t-s_{i-1}}, \qquad i = 1, \ldots, n,$$

where $s_0 = 0$, and the task is to predict future traffic $Y = A_{t+h_2} - A_{t+h_1}$ based on this knowledge (see Figure 3). By the stationarity of the increments, the covariance is given by

$$\mathrm{Cov}(A_t, A_s) = \frac{\sigma^2}{2}(v(t) + v(s) - v(t-s)),$$

where $v(t) = \mathrm{Var}(Z_t)$. Thus, the prediction formula depends only on m and $v(t)$, not on σ.

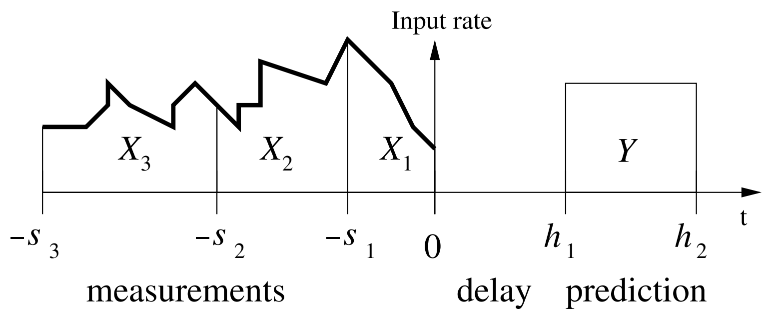


Figure 3. Prediction at time $t = 0$ and conditioning with respect to three intervals in past.

4.3 Summary of [7]

In the case of fractional Brownian motion, we answer the questions: how should one optimally choose the measurement resolution and how long time should each measurement data be stored. We first determine the reference value, predictions based on the full history. The predictor for the incoming traffic on interval $[h_1, h_2]$, $\hat{Z}(h_1, h_2) = \mathrm{E}\left(Z_{h_2} - Z_{h_1} \mid Z_s, s < 0\right)$, has a relative error

$$\frac{\mathrm{E}\left(Z(h_1,h_2) - \hat{Z}(h_1,h_2)\right)^2}{\mathrm{E}Z(h_1,h_2)^2} = 1 - \frac{\sin(\pi(H-\frac{1}{2}))}{2\pi(h_2-h_1)^{2H}} \int_0^\infty \frac{(h_1^{2H}+h_2^{2H})(1+t)^{2H} - (h_1+h_2 t)^{2H} - (h_2+h_1 t)^{2H}}{(1+t)t^{H+\frac{1}{2}}} dt.$$

Comparing the above integral with the (numerically) optimized discrete conditions, we find that conditioning with respect to two or three optimally chosen measurement intervals gives almost as accurate prediction as it is theoretically possible. We are especially interested in predictions in which the delay h_1 satisfies $h_1 = h_2 - h_1$. In this case, an engineering solution is to condition on the geometrically increasing intervals with the smallest resolution being half of the delay. It has a satisfactory accuracy if five or more intervals are used.

Using real traces from a modem pool measurement, we study the accuracy of the predictions when a "semi-stationary" traffic model is applied. We adapt the traffic model (4.1), but every time a prediction is made, we use mean rate which is estimated by a moving average method. Since the magnitude of the variance function, that is σ^2, plays no role, we can use the same variance function throughout each trace. Here a FBM type variance function $v(t) = t^{2H}$ is chosen. Moreover, instead of equally spaced measurements of the past, we condition with respect to geometrically increasing intervals.

When the goodness of a prediction is considered in the $\mathscr{L}_2$-sense, our prediction algorithm is performing well, also even if the number of traffic sources is very small. In addition, the algorithm is robust in the sense that one need not care too much about how accurately the parameters are estimated.

On other hand, a straightforward application of a traffic predictor in resource reservation may lead to problems. As a simple example we consider ε-overallocation algorithm: let g_k denote the predicted amount of traffic on interval I_k and reserve for this interval the bandwidth

$$c_k = (1+\varepsilon)g_k,$$

where $\varepsilon > 0$. We show that the parameter values for a good predictor with respect to an ε-overallocator are often different to those chosen with respect to the $\mathscr{L}_2$-criteria.

4.4 Author's contribution to [7]

The present author is the sole author of Publication [7].

5. Multifractal modeling

5.1 Introduction

Multifractals by themselves are a relatively old topic first introduced by Mandelbrot in the context of turbulence [Man72, Man74] in the early 1970s. In telecommunications, multifractals have only recently appeared: Appleby combined multifractal analysis of population distributions with network planning [App94, App95], Riedi and Lévy-Véhel applied multifractal analysis to data traces [RLV97], Taqqu, Teverovsky and Willinger considered whether network traffic is self-similar or multifractal [TTW97], Feldmann, Gilbert and Willinger modeled the multifractal nature of data traffic using a cascade based construction [FGW98], and Riedi et al. developed a multiscale modeling framework suitable for network traffic characterization [RCRB99].

There are many ways to construct random multifractal measures varying from the simple binomial measures to measures generated by branching processes (see e.g. [Man72, Man74, Fal94, AP96, Pat97, RCRB99]). In teletraffic modeling, we would also like to have, in addition to a simple and causal construction, the stationarity of the increments. Unfortunately, most of the 'classical' multifractal models, in particular tree-based cascades, lack these properties.

There have been some thoughts concerning how to use multifractal processes in teletraffic modeling. Unfortunately, neither our constructions and reasonings, nor the ones that have been suggested in different publications (e.g. random cascade measures), are fully satisfactory. Currently, the connection between multifractal processes and the reality of telecommunication systems is not completely clear. However, the family introduced in [8] is quite general and has many desired properties, thus, being a promising basis for the modeling purposes. For example, Kulkarni et al. [KMS01] have suggested a closely related approach of cascaded on-off models for TCP connection traces.

5.2 Multifractal processes

Mathematically, the local scaling behavior is measured by some singularity exponent, for example, the local Hölder exponent or the pointwise dimensions. We consider here only non-decreasing processes; a comprehensive presentation of the general case is found in [Rie02]. Let A be a (stochastic) process on $[0,1]$. The lower and upper pointwise dimensions at t are defined by

$$\begin{aligned}\underline{h}(t) &\doteq \liminf_{\delta\downarrow 0} \frac{\log(A_{t+\delta}-A_{t-\delta})}{\log\delta},\\ \overline{h}(t) &\doteq \limsup_{\delta\downarrow 0} \frac{\log(A_{t+\delta}-A_{t-\delta})}{\log\delta},\end{aligned}$$

correspondingly. When a singularity exponent is itself a highly irregular (random) function of t, i.e., sets like $E_\alpha = \{t : \underline{h}(t) = \overline{h}(t) = \alpha\}$ and $K_\alpha = \{t : \underline{h}(t) \le \alpha\}$ have fractal dimensions, the process is said to be multifractal. The irregularity of the paths can be described via the Hausdorff dimension of the sets. Function $f(\alpha) = \dim E_\alpha$ is often called the multifractal spectrum of A.

In applications, the fractal dimensions of the scaling sets are unattainable and one must consider global scaling properties. The so-called partition function presents a global summary of the scaling behavior. The (pathwise) partition function

$$\tau(q) = \liminf_{\delta\downarrow 0} \frac{\log S_\delta(q)}{\log\delta}$$

and the corresponding deterministic partition function

$$T(q) = \liminf_{\delta\downarrow 0} \frac{\log \mathrm{E}\, S_\delta(q)}{\log\delta}$$

are defined through the partition sum

$$S_\delta(q) = \sum_k (A_{k\delta} - A_{(k-1)\delta})^q,$$

where the sum is over k for which $A_{k\delta} - A_{(k-1)\delta} > 0$. The Legendre transformations of τ and T, i.e.,

$$\begin{aligned}\tau^*(\alpha) &= \inf_q(\alpha q - \tau(q)),\\ T^*(\alpha) &= \inf_q(\alpha q - T(q)),\end{aligned}$$

give approximations for the fractal dimensions, since $\dim(E_\alpha) \le \tau^*(\alpha) \le T^*(\alpha)$ with probability one. Similar type of results also hold for other singularity exponents and more general processes (see [Rie02]).

The binomial measure is the standard example of multifractal processes. The simplest construction is the deterministic one: Start from the uniform measure on unit interval, split it into two parts and weight the intervals by m and $p = 1-m$, respectively. Continue analogously by splitting intervals and redistributing the mass (see Figure 4). There are several ways to generalize and randomize this construction, for example, Mandelbrot's martingale [Man72] has i.i.d. mean one random weights and splitting into n intervals in each step. This family of measures was first analyzed by Kahane and Peyriére almost 30 years ago [KP76].

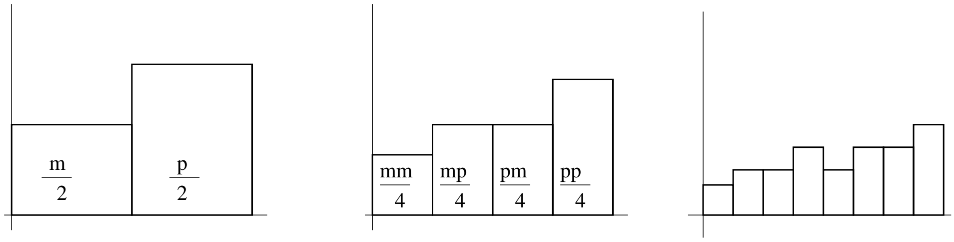


Figure 4. Construction of a binomial measure.

5.3 Summary of [8]

The construction of a binomial measure can be represented by a product of piecewise constant processes. A natural extension is to consider random measures where the redistribution is determined by a product of arbitrary stationary processes. Publication [8] studies when such constructions result in non-trivial limit measures and what kind of properties these limits have.

In a very general setting of T-martingales, this question was originally studied by Kahane [Kah87, Kah89]. Here we restrict us to a more specific family. Let us consider independent positive stationary processes $\{\Lambda^{(i)}(t)\}_{t\in[0,1]}$ with

$$\mathrm{E}\,\Lambda^{(i)}(t) = 1, \quad \forall t \in [0,1],\ i = 0,1,2\ldots.$$

Define the finite product processes

$$\Lambda_n(t) \doteq \prod_{i=0}^{n} \Lambda^{(i)}(t)$$

and the corresponding cumulative processes

$$A_n(t) \doteq \int_0^t \Lambda_n(s)\mathrm{d}s = \int_0^t \prod_{i=0}^{n} \Lambda^{(i)}(s)\,\mathrm{d}s, \qquad n = 0,1,\ldots.$$

Denote $A(t) = \lim_{n\to\infty} A_n(t)$.

The main result for processes in $\mathscr{L}_2$ is a condition for convergence to a non-trivial limit process. Assume that $\mathrm{Var}\,\Lambda^{(i)}(t) = \sigma^2 < \infty$ for all i and denote

$$\mathrm{Cov}\Big(\Lambda^{(i)}(t_1), \Lambda^{(i)}(t_2)\Big) = \sigma^2 \rho_i\,(t_1 - t_2)\,,$$

where $\rho_i(0) = 1$. Then $A_n(t) \to A(t)$ in $\mathscr{L}_2$ for all $t \in [0,1]$ if

$$\sum_{n=0}^{\infty} a_n(1) < \infty, \tag{5.1}$$

where $a_n(t) = \int_0^t (t-s)\rho_n(s) \prod_{i=0}^{n-1}(1+\sigma^2\rho_i(s))\mathrm{d}s$.

Most of the results in [8] are for self-similar products. We assume that the processes $\Lambda^{(i)}$ are independent rescaled copies of a stationary mother process Λ, i.e.,

$$\Lambda^{(i)}(\cdot) \stackrel{dist}{=} \Lambda(b^i \cdot),$$

where $b > 1$ and $\mathrm{E}\Lambda = 1$. For a large set of covariance functions the $\mathscr{L}_2$-convergence condition (5.1) simplifies a lot: $b > \mathrm{E}\Lambda^2$ is enough. After showing that $\mathrm{E}A(t)^q \sim t^{q-\log_b \mathrm{E}\Lambda^q}$, we calculate the deterministic partition function which happens to be very similar to the binomial case: $T(q) = q - 1 - \log_b \mathrm{E}\Lambda^q$. Moreover, it is shown that the long-range dependence is preserved in the construction.

Finally, we introduce an application-friendly family which is constructed from piece-wise constant Markov jump processes.

The model has natural generalizations to $\mathbb{R}^n$ and they could be used, for example, as a model for fractal population distributions. Moreover, analysis of processes which are built from products of piecewise constant stationary processes has inspired to a modification of the standard multifractal spectra: the scaling properties are easiest observed through the natural random partitionings defined by the constant periods of $\prod_{i=0}^n \Lambda^{(i)}(t)$, $n = 0, 1, \ldots$, (see [MRN03]).

5.4 Author's contribution to [8]

Riedi instructed the other two authors how this kind of problems should be handled and gave them several valuable references and hints. However, most of the new mathematical results in [8] are by the present author and Norros; the contributions being about half and half. The paper is mainly written by the present author.

References

[Add99] Addie, R. On the weak convergence of long range dependent traffic processes. *Journal of Statistical Planning and Inference*, 1999. Vol. 80, pp. 155–171.

[Adl90] Adler, R. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Hayward, California, Institute of Mathematical Statistics, 1990.

[AP96] Arbeiter, M. and Patzchke, N. Random self-similar multifractals. *Mathematiche Nachrichten*, 1996. Vol. 181, pp. 5–42.

[App94] Appleby, S. Fractal telecommunications networks. *British Telecom Technology Journal*, 1994. Vol. 12(2), pp. 19–29.

[App95] Appleby, S. Estimating the cost of a telecommunications network using the fractal structure of the human population distribution. *IEE proceedings: Communications*, 1995. Vol. 142(3), pp. 172–178.

[AS96] Abry, P. and Sellan, D. The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation. *Applied and Computational Harmonic Analysis*, 1996. Vol. 3, pp. 377–383.

[Aze80] Azencott, R. *Ecole d'Eté de Probabilités de Saint-Flour VIII-1978*, Vol. 774 of *Lecture Notes in Mathematics*, Berlin. Springer, 1980. Pp. 1–176.

[Bar02] Baran, P. The beginnings of packet switching: some underlying concepts. *IEEE Communications Magazine*, 2002. Vol. 40(7), pp. 42–48.

[BD91] Brockwell, P. and Davis, R. *Time series: theory and methods*, 2nd edition. Berlin, Springer, 1991.

[BHK+91] Baker, M., Hartman, J., Kupfer, M., Shirriff, K. and Ousterhout, J. Measurements of a distributed file system. In *Proceedings of the 13th ACM Symposium on Operating System Principles*. Pasific Grove, CA, USA, 1991. Pp. 977–984.

[Bil68] Billingsley, P. *Convergence of Probability Measures*, 1st edition. New York, Wiley, 1968.

[Bil99] Billingsley, P. *Convergence of Probability Measures*, 2nd edition. New York, Wiley, 1999.

[BSTW95] Beran, J., Sherma, R., Taqqu, M. and Willinger, W. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, 1995. Vol. 43, pp. 1566–1579.

[BW98a] Berger, A. and Whitt, W. Effective bandwidths with priorities. *IEEE/ACM Transaction on Networking*, 1998. Vol. 6(4), pp. 447–460.

[BW98b] Berger, A. and Whitt, W. Extending the effective bandwidth concept to networks with priority classes. *IEEE Communications Magazine*, August 1998, pp. 78–83.

[BZ79] Bahadur, R. and Zabell, S. Large deviations of the sample mean in general vector spaces. *Annals of Probability*, 1979. Vol. 7(4), pp. 587–621.

[CB97] Crovella, M. and Bestavros, A. Self-similarity in World Wide Web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 1997. Vol. 5, pp. 835–847.

[Cox84] Cox, D. Long-range dependence: a review. In David, H. and David, H., eds., *Statistics: An Appraisal*, Ames, Iowa. The Iowa State University Press, 1984. Pp. 55–74.

[CS99] Choe, J. and Shroff, N. On the supremum distribution of integrated stationary Gaussian processes with negative linear drift. *Advances in Applied Probability*, 1999. Vol. 31, pp. 135–157.

[Dav01] Davies, D. An historical study of the beginnings of packet switching. *The Computer Journal*, 2001. Vol. 44(3), pp. 152–162.

[D02] Dębicki, K. Ruin probability for Gaussian integrated processes. *Stochastic Processes and Applications*, 2002. Vol. 98, pp. 151–174.

[Dec00] Decreusefond, L. A Skohorod-Stratonovitch integral for the fractional Brownian motion. In *Proceedings of the 8th Workshop on Stochastic Analysis and Related Fields*, 2000.

[DH87] Davies, R. and Harte, D. Tests for Hurst effect. *Biometrika*, 1987. Vol. 74(1), pp. 95–101.

[Die02] Dieker, T. *Simulation of fractional Brownian motion*. Master's thesis, the Vrije Universiteit Amsterdam, 2002. http://www.cwi.nl/~ton/mastersthesis/index.html.

[DMR98] Dębicki, K., Michna, Z. and Rolski, T. On the supremum from Gaussian processes over infinite horizon. *Probability and Mathematical Statistics*, 1998. Vol. 8, pp. 83–100.

[DS89] Deuschel, J.-D. and Stroock, D. *Large Deviations*. Boston, Academic Press, 1989.

[DZ02] Dzhaparidze, K. and Zanten, J. v. *A series representation for fractional Brownian motion*. Tech. Rep. 2002-5, Department of Stochastics, Vrije Universiteit Amsterdam, 2002.

[Erl09] Erlang, A. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, 1909. Vol. 20.

[Fal94] Falconer, K. The multifractal spectrum of statistically self-similar measures. *Journal of Theoretical Probability*, 1994. Vol. 7, pp. 681–702.

[FGHW99] Feldmann, A., Gilbert, A., Huang, P. and Willinger, W. Dynamics of IP traffic: A study of the role of variability and the impact of control. In *Proceedings of the ACM/SIGCOMM'99*. Cambridge, MA, USA, 1999. Pp. 301–313.

[FGW98] Feldmann, A., Gilbert, A. and Willinger, W. Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic. In *Proceedings of the ACM/SIGCOMM'98*. Vancouver, Canada, 1998. Pp. 25–38.

[Fla92] Flandrin, P. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 1992. Vol. 38(2), pp. 910–917.

[GK02] Gaigalas, R. and Kaj, I. *Convergence of scaled renewal processes and a packet arrival model*. U.U.D.M. Report 2002:2, Uppsala University, January 2002. To appear in Bernoulli.

[GN96] Gripenberg, G. and Norros, I. On the prediction of fractional Brownian motion. *Journal of Applied Probability*, 1996. Vol. 33, pp. 400–410.

[HP99] Hüsler, J. and Piterbarg, V. Extremes of a certain class of Gaussian processes. *Stochastic Processes and Applications*, 1999. Vol. 83, pp. 257–271.

[Kah85] Kahane, J.-P. Sur le chaos multiplicatiff. *Ann. Sc. Math. Québec*, 1985. Vol. 9, pp. 105–150.

[Kah87] Kahane, J.-P. Positive martingales and random measures. *Chinese Annals of Mathematics*, 1987. Vol. 8B, pp. 1–12.

[Kah89] Kahane, J. Random multiplications, random coverings and multiplicative chaos. In Berkson, E., Peck, N. and Uhl, J., eds., *Analysis at Urbana I*, no. 137 in London Math. Soc. Lecture Notes. Cambridge University Press, 1989. Pp. 196–255.

[Kil03] Kilpi, J., 2003. Private communications.

[Kle02] Kleinrock, L. Creating a mathematical theory of computer networks. *Operations research*, 2002. Vol. 50(1), pp. 125–131.

[KMS01] Kulkarni, V., Marron, J. and Smith, F. *A cascaded on-off model for TCP connection traces*. UNC/OR/TR01-1, University of North Carolina, 2001.

[KN02] Kilpi, J. and Norros, I. Testing the Gaussian approximation of aggregate traffic. In *Proceedings of the 2nd Internet Measurement Workshop*. Marseille, France, 2002. Pp. 49–61.

[KP76] Kahane, J.-P. and Peyriére, J. Sur certaines martingales de Benoit Mandelbrot. *Advances in Mathematics*, 1976. Vol. 22, pp. 131–145.

[KSJ98] Konstantopoulos, T. and S.-J., L. Macroscopic models for long-range dependent network traffic. *Queueing Systems*, 1998. Vol. 28, pp. 215–243.

[LEWW95] Lau, W.-C., Erramilli, A., Wang, J. and Willinger, W. Self-similar traffic generation: The random midpoint displacement algorithm and its properties. In *1995 IEEE International Conference on Communications*. Seattle, USA, 1995. Pp. 466–472.

[LTWW93] Leland, W., Taqqu, M., Willinger, W. and Wilson, D. On the self-similar nature of Ethernet traffic. In *ACM SIGCOMM '93*, 1993. Pp. 183–193.

[LTWW94] Leland, W., Taqqu, M., Willinger, W. and Wilson, D. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 1994. Vol. 2(1), pp. 1–15.

[LVR97] Lévy Véhel, J. and Riedi, R. Fractional Brownian motion and data traffic modeling: The other end of the spectrum. In *Fractals in Engineering 97*. Springer, 1997. Pp. 185–202.

[LW91] Leland, W. and Wilson, D. High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection. In *Proceedings of IEEE INFOCOM'91*. Bal Harbour, Florida, 1991. Pp. 1360–1366.

[Man71] Mandelbrot, B. A fast fractional Gaussian noise generator. *Water Resources Research*, 1971. Vol. 7, pp. 543–553.

[Man72] Mandelbrot, B. Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence. In Rosenblatt, M. and Van Atta, C., eds., *Statistical models and turbulence*, no. 12 in Lecture notes in physics. Springer, 1972. Pp. 331–351.

[Man74] Mandelbrot, B. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 1974. Vol. 62, pp. 331–358.

[Mas99] Massoulié, L. Large deviations estimates for polling and weighted fair queueing service systems. *Advances in Performance Analysis*, 1999. Vol. 2(2), pp. 103–127.

[MN97] Mannersalo, P. and Norros, I. *Multifractal analysis of real ATM traffic: a first look*. COST257TD(97)19, VTT Information Technology, 1997.

[MRN03] Mannersalo, P., Riedi, R. and Norros, I. Multifractal products of stochastic processes: Part II, 2003. Under construction.

[MRRS02] Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Annals of Applied Probability*, 2002. Vol. 12, pp. 23–68.

[MS99] Massoulié, L. and Simonian, A. Large buffer asymptotics for the queue with FBM input. *Journal of Applied Probability*, 1999. Vol. 36(3), pp. 894–906.

[MU02] Mandjes, M. and van Uitert, M. Sample-path large deviation for tandem and priority queues with Gaussian inputs. CWI, Report PNA-R0221, 2002.

[Nar98] Narayan, O. Exact asymptotic queue length distribution for fractional Brownian traffic. *Advances in Performance Analysis*, 1998. Vol. 1(1).

[Nor93] Norros, I. Studies on a model for connectionless traffic, based on fractional Brownian motion. In Conference on Applied Probability in Engineering, Computer and Communication Sciences, INRIA/ORSA/TIMS/SMAI, Paris, June 16–18, 1993.

[Nor94] Norros, I. A storage model with self-similar input. *Queueing Systems*, 1994. Vol. 16, pp. 387–396.

[Nor95] Norros, I. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE Journal on Selected Areas in Communications*, 1995. Vol. 13(6), pp. 953–962.

[Nor99] Norros, I. Busy periods of fractional Brownian storage: a large deviations approach. *Advances in Performance Analysis*, 1999. Vol. 2(1), pp. 1–20.

[ÖS01] Östring, S. and Sirisena, H. The influence of long-range dependence on traffic prediction. In *Proceedings of ICC 2001*, 2001. Pp. 1000–1005.

[ÖSH01] Östring, S., Sirisena, H. and Hudson, I. Rate control of elastic connections competing with long-range dependent network traffic. *IEEE Transactions on Communications*, 2001. Vol. 49(6), pp. 1092–1101.

[Pat97] Patzchke, N. Self-conformal multifractal measures. *Advances in Applied Mathematics*, 1997. Vol. 19, pp. 486–513.

[Pax97] Paxson, V. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *Computer Communication Review*, 1997. Vol. 27, pp. 5–18.

[PF95] Paxson, V. and Floyd, S. Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 1995. Vol. 3, pp. 226–244.

[Pit96] Pitebarg, V. *Asymptotic methods in the theory of Gaussian processes and fields*. Providence, Rhode Island, USA, AMS, 1996.

[Pit01] Piterbarg, V. Large deviations of a storage process with fractional Brownian motion. *Extremes*, 2001. Vol. 4, pp. 147–164.

[PT01] Park, K. and Tuan, T. Multiple time scale TCP: harnessing proactivity in self-similar burstiness. In *Proceeding of ITC 17*, 2001. Pp. 421–432.

[RCRB99] Riedi, R., Crouse, M., Ribeiro, V. and Baraniuk, R. A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory*, 1999. Vol. 45(3), pp. 992–1018.

[Rie02] Riedi, R. Multifractal processes. In Doukhan, P., Oppenheim, G. and Taqqu, M., eds., *Long range dependence: theory and applications*. Birkhäuser, 2002. Pp. 625–715.

[RLV97] Riedi, R. and Lévy Véhel, J. *TCP traffic is multifractal: a numerical study*. Inria research report, no. 3129, Project Fractales, INRIA Rocquencourt, 1997.

[RMV97] Roberts, J., Mocci, U. and Virtamo, J., eds. *Broadband network teletraffic*, Vol. 1155 of *Lecture notes in computer science*. Berlin, Springer, 1997.

[RW00] Riedi, R. and Willinger, W. Toward an improved understanding of network traffic dynamics. In *Self-similar traffic and performance evolution*. Wiley, 2000. Pp. 507–530.

[SL02] Sang, A. and Li, S.-Q. A predictability analysis of network traffic. *Computer Networks*, 2002. Vol. 39(4), pp. 329–345.

[ST94] Samorodnitsky, G. and Taqqu, M. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* New York, Chapman and Hall, 1994.

[TTW97] Taqqu, M., Teverosky, V. and Willinger, W. Is network traffic self-similar or multifractal? *Fractals*, 1997. Vol. 5, pp. 63–73.

[TWS97] Taqqu, M., Willinger, W. and Sherman, R. Proof for a fundamental result in self-similar traffic modeling. *Computer Communications Review*, 1997. Vol. 27, pp. 5–23.

[Whi02] Whitt, W. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and their Application to Queues*. New York, Springer, 2002.

[Wis01] Wischik, D. Sample path large deviations for queues with many inputs. *Annals of Applied Probability*, 2001. Vol. 11, pp. 389–404.

[WTSW97] Willinger, W., Taqqu, M., Sherman, R. and Wilson, D. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 1997. Vol. 5(1), pp. 71–86.

Errata

Publication [4]

Figure 5. Class 1 and Class 2 traffic were not independent; the seeds of the random number generator were not changed.

Publication [6]

Page 408, Figure 4. *m* should read μ.

Appendices of this publication are not included in the PDF version. Please order the printed version to get the complete publication (http://www.vtt.fi/inf/pdf/)

Published by

Series title, number and report code of publication

VTT Publications 491
VTT–PUBS–491

Author(s)
Mannersalo, Petteri

Title

Gaussian and multifractal processes in teletraffic theory

Abstract

In this thesis, we consider two classes of stochastic models which both capture some of the essential properties of teletraffic. Teletraffic has two time regimes where profoundly different behavior and characteristics are seen. When traffic traces are observed at coarse resolutions, properties like self-similarity and long-range dependence are visible. In small time-scales, traffic exhibits complex scaling laws with much more spiky bursts than in coarser resolutions. The main part of the thesis is devoted to a large time-scale analysis by considering Gaussian processes and queueing systems with Gaussian input. In order to understand the small time-scale dynamics, first steps are taken towards general multifractal models offering a suitable basis for short time-scale teletraffic modeling.

The family of Gaussian processes with stationary increments serves as the traffic model for large time-scales. First, we introduce a fast and accurate simulation algorithm, which can be used to generate long approximate Gaussian traces. Moreover, the algorithm is also modified to run on-the-fly. Then approximate queue length distributions for ordinary, priority and generalized processor sharing queues are derived using a most probable path approach. Simulation studies show that the performance formulae appear to be quite accurate over the full range of buffer levels. Finally, we construct a semi-stationary predictor, which uses a constant variance function and mean rate estimation based on a moving average method. Moreover, we show that measuring the past of a process by geometrically increasing intervals is a good engineering solution and a much better way than equally spaced measurements.

We introduce a family of multifractal processes which belongs to the framework of *T*-martingales and multiplicative chaos introduced by Kahane. The family has many desirable properties like stationarity of increments, concave multifractal spectra and simple construction. We derive, for example, conditions for non-degeneracy, establish a power law for the moments and obtain a formula for the multifractal spectrum.

Keywords
Gaussian processes, multifractals, queueing systems, performance analysis, traffic modeling

Activity unit
VTT Information Technology, Otakaari 7 B, P.O.Box 1202, FIN–02044 VTT, Finland

ISBN			Project number
951–38–6036–1 (soft back ed.) 951–38–6037–X (URL:http://www.inf.vtt.fi/pdf/)			
Date	Language	Pages	Price
March 2003	English	44 p. + app. 109 p.	D
Name of project		Commissioned by	
Series title and ISSN		Sold by	
VTT Publications 1235–0621 (soft back ed.) 1455–0849 (URL: http://www.inf.vtt.fi/pdf/)		VTT Information Service P.O.Box 2000, FIN–02044 VTT, Finland Phone internat. +358 9 456 4404 Fax +358 9 456 4374	

VTT PUBLICATIONS 491 Gaussian and multifractal processes in teletraffic theory Petteri Mannersalo

Teletraffic has two time regimes where profoundly different behavior and characteristics are seen. When traffic traces are observed at coarse resolutions, properties like self-similarity and long-range dependence are visible. In small time-scales, traffic exhibits complex scaling laws with much more spiky bursts than in coarser resolutions. The main part of the this report is devoted to a large time-scale analysis by considering Gaussian processes and queueing systems with Gaussian input. In order to understand the small time-scale dynamics, first steps are taken towards general multifractal models offering a suitable basis for short time-scale teletraffic modeling.

Tätä julkaisua myy	Denna publikation säljs av	This publication is available from
VTT TIETOPALVELU PL 2000 02044 VTT Puh. (09) 456 4404 Faksi (09) 456 4374	VTT INFORMATIONSTJÄNST PB 2000 02044 VTT Tel. (09) 456 4404 Fax (09) 456 4374	VTT INFORMATION SERVICE P.O.Box 2000 FIN–02044 VTT, Finland Phone internat. +358 9 456 4404 Fax +358 9 456 4374

ISBN 951–38–6036–1 (soft back ed.)
ISSN 1235–0621 (soft back ed.)

ISBN 951–38–6037–X (URL: http://www.inf.vtt.fi/pdf/)
ISSN 1455–0849 (URL: http://www.inf.vtt.fi/pdf/)