
Traffic Characterisation and Modelling for Call Admission Control Schemes on Asynchronous Transfer Mode Networks

Stephen Bates



A thesis submitted for the degree of Doctor of Philosophy.

The University of Edinburgh.

-September 1997-

Abstract

Allocating resources to variable bitrate (VBR) teletraffic sources is not a trivial task because the impact of such sources on a buffered switch is difficult to predict. This problem has repercussions for call admission control (CAC) on asynchronous transfer mode (ATM) networks.

In this thesis we report on investigations into the nature of several types of VBR teletraffic. The purpose of these investigations is to identify parameters of the traffic that may assist in the development of CAC algorithms. As such we concentrate on the correlation structure and marginal distribution; the two aspects of a teletraffic source that affect its behaviour through a buffered switch.

The investigations into the correlation structure consider whether VBR video is self-similar or non-stationary. This question is significant as the exponent of self-similarity has been identified as being useful for characterising VBR teletraffic. Although results are inconclusive with regards to the original question, they do show that self-similar models are best able to capture the video data's behaviour.

The investigations into the marginal distributions are in two parts. The first considers applying a structured Markovian model to ATM data and demonstrates how model parameters can be estimated from measurable properties of teletraffic data. This has implications for parametric CAC. The second part considers the use of stable distributions in teletraffic characterisation and modelling. We show that several teletraffic datasets are heavy tailed and then develop a framework for the estimation of stable distribution parameters.

We finish by considering the effective bandwidths of stable distributions and models and by considering the effect of stable parameters on model behaviour. This is done in an attempt to develop a CAC algorithm based on the paradigms of self-similarity and stable distributions.

Declaration of Originality

I hereby declare that the research recorded in this thesis, and the thesis itself, is the original and sole work performed by the author while studying in the Department of Electrical Engineering at The University of Edinburgh.

Stephen Bates

Acknowledgements

Many people have help me get through the past three years. I'd like to thank the following.

- My parents for getting me this far and for all their love and support.
- My siblings Cazza and Shazza for being great.
- My friends in Edinburgh, I hope you all know who you are. Super thanks to John, Pete, Bri, Mary, Chris, Al, Dave, Dave, Richard, Catherine, Steve and Jane.
- Everyone in the signals and systems group for their advice whether it was technical or otherwise.
- Ganesh (The University of Edinburgh), Stan Zachary (Heriot Watt), James Hall, Simon Crosby and Co. (University of Cambridge Computing Labs), Raymond Russell, Fergal Toomey and Co. (DIAS, Dublin) and all other outside assistants.
- Steve McLaughlin for aiding and abetting.
- GEC Marconi Avionics for funding me for the last 3 years.

Contents

List of Figures	vii
Nomenclature	xv
Acronyms and Abbreviations	xvii
1 Introduction	1
1.1 Motivation	1
1.2 Layout of thesis	2
2 Call admission control on ATM networks	5
2.1 Introduction	5
2.2 The asynchronous transfer mode	5
2.2.1 Introduction	5
2.2.2 B-ISDN transfer mode	7
2.2.3 The ATM cell	8
2.2.4 Service classes	10
2.2.5 Adaption layers	11
2.2.6 Network topology	12
2.2.7 Network control and management	13
2.2.8 Asynchronous multiplexing and statistical gain	14
2.2.9 Traffic parameters	16
2.3 Traffic characterisation	16
2.3.1 Introduction	16
2.3.2 Effective bandwidths	17
2.3.3 Teletraffic modelling	21
2.4 The adaption of a structured Markovian model for ATM	22
2.4.1 Introduction	22

2.4.2	The structured Markovian model	23
2.4.3	A modelling example	27
2.4.4	Determining the higher moments of the model	28
2.4.5	Impact of long tail distributions	33
2.5	Discussion	35
2.6	Conclusions	36
3	Self-similarity teletraffic analysis and modelling	37
3.1	Introduction	37
3.2	Definitions and properties of self-similar time-series	38
3.3	Implications of self-similarity in B-ISDN networks	41
3.4	Measuring self-similarity in time-series	42
3.4.1	The rescaled adjusted range technique	43
3.4.2	The variance plot technique	43
3.4.3	The periodogram plot	44
3.4.4	The Whittle approximate MLE technique	45
3.4.5	Comparing the estimation techniques	45
3.5	Self-similar teletraffic models	46
3.5.1	Fractional Brownian motion and the Norros model	46
3.5.2	The auto-regressive integrated moving average model	48
3.5.3	The marginal distribution mapping model	48
3.5.4	Other models	50
3.6	Summary	51
4	Is VBR video non-stationary or self-similar? Implications for traffic characterisation	52
4.1	Introduction	52
4.2	The VBR video data	53
4.2.1	Introduction	53
4.2.2	The JPEG encoded data	53
4.2.3	The MPEG encoded data	54
4.3	The comparative data	56

4.3.1	The ON-OFF model	56
4.3.2	Auto-regressive models	57
4.3.3	The Norros self-similar model	59
4.3.4	Summary of model details	60
4.4	The wide sense stationary quotient	60
4.4.1	Formal definition	61
4.4.2	Results	62
4.4.3	Discussion	63
4.5	Hurst exponent dependence on N	66
4.5.1	Formal definition	66
4.5.2	Results	66
4.5.3	Discussion	66
4.6	Conclusions	69
5	Testing the Gaussian assumption for teletraffic traces	71
5.1	Introduction	71
5.2	The data	72
5.3	Data transform	73
5.4	Testing the data with the χ^2 -test	74
5.4.1	The χ^2 -test for normality	74
5.4.2	Results	75
5.5	Testing the data with the Shapiro-Wilk test	77
5.5.1	The Shapiro-Wilk test	77
5.5.2	Results	79
5.6	Considering LRD in the data	80
5.7	Qualitative study of the data distributions	81
5.7.1	Procedure	81
5.7.2	Results	81
5.8	Quantitative study of the data distributions	83
5.9	Conclusions	86

6	Stable distributions & stable parameter estimation	87
6.1	Introduction	87
6.2	Stable random variables	88
6.2.1	Stable parameters	88
6.2.2	The generalised central limit theorem	90
6.2.3	Normalising stable distributions	91
6.2.4	Fractional lower order moments	91
6.3	Generating stable random variables	91
6.4	The estimation of stable parameters	92
6.4.1	Estimation techniques	92
6.4.2	Summary	96
6.5	Estimating parameters for stable random variables	97
6.5.1	Procedure & results	97
6.5.2	Discussion	99
6.6	Conclusions	102
7	The analysis and modelling of teletraffic with stable distributions and models	103
7.1	Introduction	103
7.2	Estimating parameters for real data	104
7.2.1	The teletraffic data	104
7.2.2	Data transform	105
7.2.3	The estimation of the assumed Gaussian innovations	105
7.2.4	The estimation of the dataset pdfs	108
7.3	Modelling data with stable distributions	113
7.3.1	Introduction	113
7.3.2	Self-similar stable processes	113
7.3.3	The Norros model with symmetric stable innovations	114
7.3.4	The Norros model with asymmetric stable innovations	116
7.3.5	The marginal distribution mapping model	117
7.4	The effective bandwidth of stable distributions	118

7.4.1	The infinite moment generating function	118
7.4.2	The adapted moment generating function	119
7.4.3	Practical implications	121
7.5	A CAC scheme based on the online estimation of α	123
7.5.1	The effect of α on the queueing behaviour of the adapted Norros model	124
7.6	Conclusions	125
8	Conclusions & discussion	127
8.1	Future work	129
	References	131
A	Stable distribution estimation and modelling software	140
A.1	Overview	140
A.2	Installation	140
A.3	Usage	141
A.4	Disclaimer	141
B	An experiment to illustrate statistical gain	142
C	The JPEG and MPEG coding schemes	144
C.1	Introduction	144
C.2	JPEG	144
C.3	MPEG	145
D	Publications	147

List of Figures

1.1	<i>The desired system: The parameter set is estimated and using this and some function, $f()$ a prediction can be made for the service rate of the buffered switch.</i>	2
2.1	<i>A graph showing the holding time and bandwidth of the more popular broadband traffic types.</i>	6
2.2	<i>The ATM cell structure.</i>	8
2.3	<i>The ATM cell header format for (a) the user network interface (UNI) and (b) the network network interface (NNI).</i>	8
2.4	<i>A simplified example of the method used to route cells in ATM switches.</i>	9
2.5	<i>How the ATM header and the AAL are used to transfer data across ATM networks.</i>	11
2.6	<i>ATM network hierarchy</i>	12
2.7	<i>The topology of an ATM network</i>	13
2.8	<i>Statistical Gain - This comparison illustrates how asynchronous multiplexing can increase link utilisation. It is important to note that in the asynchronous case that the empty timeslots have not been reserved and may be utilised by other services.</i>	15
2.9	<i>How a typical ATM link might be utilised. The ABR services uses feedback to react to the changes in the VBR services and the UBR services fills the remaining space left by the other three.</i>	17
2.10	<i>A single server queue. $A[i]$ is a discrete arrival process, c is the service rate, B is the buffer size and $Q[i]$ is the size of the queue in the buffer.</i>	18
2.11	<i>The estimated effective bandwidth for the ON-OFF model (dashed line = the exact value).</i>	20
2.12	<i>The buffer occupancy plot estimated by Monte Carlo simulation (dashed line = the exact value).</i>	20
2.13	<i>State transition diagram for the structured Markovian model.</i>	24
2.14	<i>Solution Surface for $ABR=1Mbs^{-1}$ and $PBR=6Mbs^{-1}$.</i>	27
2.15	<i>Variance of the resultant traffic.</i>	29

2.16	<i>Time series graph for point a (low variance).</i>	29
2.17	<i>Time series graph for point b (high variance).</i>	30
2.18	<i>Cell inter-arrival diagram.</i>	30
2.19	<i>A uniform pdf of state occurrences with mean = 10</i>	32
2.20	<i>State recurrence pdfs for estimator and MC simulation for the uniform case.</i>	34
2.21	<i>Variance of the cell inter-arrival rate for the uniform case.</i>	34
2.22	<i>State recurrence pdfs for MC simulation for the poisson case.</i>	35
2.23	<i>Variance of the cell inter-arrival rate.</i>	35
3.1	<i>(a) The original self-similar process. (b) The process aggregated by a factor of 10 and scaled accordingly. (c) The process aggregated by a factor of 100 and scaled accordingly.</i>	39
3.2	<i>The marginal distribution mapping model: An illustrative example.</i>	49
4.1	<i>(a) A sample trace from the JPEG encoding of the motion picture Star Wars (b) the 200 bin histogram.</i>	53
4.2	<i>The first 108 points of the correlation function estimated from the JPEG data.</i>	54
4.3	<i>(a) A sample trace from the MPEG encoding of the motion picture Star Wars (b) the 200 bin histogram.</i>	54
4.4	<i>The first 108 points of the correlation function estimated from the MPEG data.</i>	55
4.5	<i>The first 108 points of the correlation function estimated from the MPEG data after grouping.</i>	55
4.6	<i>(a) The values for the Akaike Information Criterion for model order 1 to 30 (b) a 5000 point sample of the JPEG data generated by the model when fed with iid $N(0,1)$ innovations.</i>	58
4.7	<i>The correlation functions for the AR model trained with data from the MPEG video trace.</i>	59
4.8	<i>Measuring sample mean and variance of a sine wave.</i>	61
4.9	<i>Results for the WSS quotient for the unaggregated JPEG datasets.</i>	63
4.10	<i>Results for the WSS quotient for the JPEG datasets averaged into blocks of 12.</i>	63
4.11	<i>Results for the WSS quotient for the JPEG datasets averaged into blocks of 144.</i>	64
4.12	<i>Results for the WSS quotient for the unaggregated MPEG datasets.</i>	64

4.13	<i>Results for the WSS quotient for the MPEG datasets averaged into blocks of 12.</i>	65
4.14	<i>Results for the WSS quotient for the MPEG datasets averaged into blocks of 144.</i>	65
4.15	<i>Results for H_N for the unaggregated JPEG datasets.</i>	67
4.16	<i>Results for H_N for the JPEG datasets averaged into blocks of 12.</i>	67
4.17	<i>Results for H_N for the JPEG datasets averaged into blocks of 144.</i>	68
4.18	<i>Results for H_N for the unaggregated MPEG datasets.</i>	68
4.19	<i>Results for H_N for the MPEG datasets averaged into blocks of 12.</i>	69
4.20	<i>Results for H_N for the MPEG datasets averaged into blocks of 144.</i>	69
5.1	<i>(a) Sample trace of the converted data and (b) the 200 bin histogram of the converted dataset.</i>	74
5.2	<i>(a) The CDF of the transformed JPEG data compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.</i>	81
5.3	<i>(a) The CDF of the transformed MPEG data compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.</i>	82
5.4	<i>(a) The CDF of the transformed Ethernet data pAug.TL compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.</i>	82
5.5	<i>(a) The CDF of the transformed Ethernet data pOct.TL compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.</i>	83
5.6	<i>(a) The CDF of the transformed Ethernet data OctExt.TL compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.</i>	83
5.7	<i>The Hill estimate for JPEG, MPEG and Ethernet data.</i>	85
6.1	<i>Symmetric stable distributions with varying α. The point to note is that the distribution tails get heavier as $\alpha \rightarrow 0$.</i>	89
6.2	<i>Stable distributions with varying β ($\alpha = 1.5$).</i>	90
6.3	<i>The plots for calculating $\hat{\alpha}$ and $\hat{\beta}$ using Kogon's technique.</i>	95
6.4	<i>Results for $\hat{\alpha}$ when $\beta = 0$.</i>	98
6.5	<i>Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = -1$.</i>	98
6.6	<i>Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = -0.5$.</i>	99
6.7	<i>Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = 0.5$.</i>	99
6.8	<i>Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = 1$.</i>	100

6.9	<i>Results for $\hat{\alpha}$ when the FLOM based techniques are applied to a skewed distribution.</i>	100
6.10	<i>Results for \hat{c} when $k=0.1$.</i>	101
6.11	<i>Results for \hat{c} when $k=10$.</i>	101
7.1	<i>The CDF of the transformed JPEG data compared to the CDF of stable random variables generated using the estimates from Table 7.1.</i>	106
7.2	<i>The CDF of the transformed MPEG data compared to the CDF of stable random variables generated using the estimates from Table 7.2.</i>	107
7.3	<i>The CDF of the transformed MPEG GOP data compared to the CDF of stable random variables generated using the estimates from Table 7.3.</i>	108
7.4	<i>The CDF of the transformed Ethernet data compared to the CDF of stable random variables generated using the estimates from Table 7.4.</i>	109
7.5	<i>The CDF of the transformed differenced Ethernet data compared to the CDF of stable random variables generated using the estimates from Table 7.5.</i>	110
7.6	<i>The real and estimated stable pdfs for the JPEG data.</i>	110
7.7	<i>The real and estimated stable pdfs for the MPEG data.</i>	111
7.8	<i>The real and estimated stable pdfs for the MPEG GOP data.</i>	112
7.9	<i>The real and estimated stable pdfs for the differenced Ethernet data.</i>	113
7.10	<i>Samples from the original (a) JPEG, (b) MPEG GOP and (c) Ethernet teletraffic datasets (the y axis represents bits per frame or packet and the x axis represents frame or packet number).</i>	115
7.11	<i>Samples from the adapted Norros model (with symmetric stable innovations) trained with (a) JPEG, (b) MPEG GOP and (c) Ethernet teletraffic datasets (the y axis represents bits per frame or packet and the x axis represents frame or packet number).</i>	115
7.12	<i>Samples from the adapted Norros model (with asymmetric stable innovations) trained with (a) JPEG, (b) GOP MPEG and (c) Ethernet teletraffic datasets (the y axis represents bits per frame or packet and the x axis represents frame or packet number).</i>	117
7.13	<i>Samples from the marginal distribution mapping model for (a) JPEG and (b) GOP MPEG data (the y axis represents bits per frame and the x axis represents frame number).</i>	118
7.14	<i>The approximate pdf and adapted moment generating function for $\theta = 0.0001$ and $\alpha=1.0, 1.5$ and 2.0 using the IFT. The exact values for the $\alpha = 2$ case are also given.</i>	121

7.15	<i>The approximate pdf and adapted moment generating function for $\theta = 0.0001$ and $\alpha=1.0, 1.5$ and 2.0 using asymptotic expansion. The exact values for the $\alpha = 2$ case is also given.</i>	122
7.16	<i>The buffer occupancy curve for the truncated iid stable Norros model serviced at 10.1838 bits/s with the expected value from effective bandwidth theory (dashed line).</i>	122
7.17	<i>The buffer occupancy curve for the truncated iid stable Norros model serviced at 10.0518 bits/s with the expected value from effective bandwidth theory (dashed line).</i>	123
7.18	<i>The probability of buffer overflow for (a) $1.1 \leq \alpha \leq 1.5$ and (b) $1.6 \leq \alpha \leq 2.124$</i>	
7.19	<i>The estimated slope of the buffer overflow probability curves vs α.</i>	125

List of Tables

2.1	General B-ISDN service classes as defined by the ATM Forum.	10
2.2	Results from model simulations.	28
3.1	Results of the comparative study of the Hurst exponent estimators for a Gaussian self-similar process.	45
3.2	Results of the comparative study of the Hurst exponent estimators for MPEG data.	46
3.3	The Hurst exponent estimates of the output of the marginal distribution mapping model.	50
4.1	The summary of each model generated dataset's ability to capture as- pects of the video data.	60
5.1	Statistics of the datasets generated from the Bellcore trace pAug.TL . .	72
5.2	Statistics of the the datasets generated from the Bellcore trace pOct.TL	72
5.3	Statistics of the the datasets generated from the Bellcore trace OctExt.TL	73
5.4	Results of placing the JPEG data into bins to determine the χ^2 -test value.	76
5.5	Results of the χ^2 -test for the Bellcore trace pAug.TL	76
5.6	Results of the χ^2 -test for the Bellcore trace pOct.TL	76
5.7	Results of the χ^2 -test for the Bellcore trace OctExt.TL	77
5.8	Results of the adjusted SW test for the Bellcore trace pAug.TL	79
5.9	Results of the adjusted SW test for the Bellcore trace pOct.TL	79
5.10	Results of the adjusted SW test for the Bellcore trace OctExt.TL	79
5.11	Results of the adjusted SW test for the VBR video datasets.	80
5.12	Percentage of the SW tests that passed for true fGn traces.	80
5.13	Results for $\hat{\gamma}$ for the generalised Pareto technique.	86
6.1	Permissible ranges of estimation.	97
7.1	Results for the JPEG video data.	106
7.2	Results for the MPEG video data.	106
7.3	Results for the MPEG GOP video data.	107

7.4	Results for the original Ethernet activity data.	108
7.5	Results for the differenced Ethernet activity data.	108
7.6	Results for JPEG video data.	109
7.7	Results for the MPEG video data.	111
7.8	Results for the MPEG GOP video data.	111
7.9	Results for the Ethernet activity data.	112
7.10	Results for the differenced Ethernet activity data.	112
7.11	The model parameters for the adapted Norros model with symmetric innovations.	115
7.12	The model parameters for the adapted Norros model with asymmetric innovations.	116

Nomenclature

a	location parameter for a stable distribution
$A(t)$	arrival process
$\hat{A}(t)$	cumulative arrival process
$AIC(p)$	Akaike information criterion
B	buffer size
$B_H(t)$	fractional Brownian motion
c	service rate of a single server
C	capacity of a switch or link
$E()$	expected value
$f(x)$	probability density function
$F(x)$	cumulative density function
H	Hurst exponent
$H_{l,n}$	Hill estimate
$\text{Im}()$	imaginary component of a complex number
$M(\theta)$	moment generating function
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
N	size of a sample
$P()$	probability
$Q(t)$	queue size
$\text{Re}()$	real component of a complex number
$\text{VAR}()$	variance
W_N	wide sense stationary quotient
$X_H(t)$	fractional Gaussian noise
α	characteristic exponent for a stable distribution
β	index of skew for a stable distribution
$\delta(\theta)$	effective bandwidth
γ	scale parameter for a stable distribution
μ	average of a sample
σ^2	variance of a sample
Δ	maximum of a sample
∇	minimum of a sample
$\lambda(\theta)$	scaled cumulant generating function
θ	loss coefficient

$\rho(k)$ correlation function
 χ_k^2 chi-square distribution with k degrees of freedom

Notation convention

It is assumed that a continuous time process can be appropriately sampled to produce a discrete time process. For the discrete time process square brackets, $[]$, are used and the samples are indexed using $i \in Z$.

Acronyms and Abbreviations

AAL	ATM Adaption Layer
ABR	Available Bit Rate
ACF	AutoCorrelation Function
AEVT	Asymptotic Extreme Value Theory
ARMA	Auto-Regressive Moving Average
ARIMA	Auto-Regressive Integrated Moving Average
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband Integrated Services Distributed Network
BT	Burst Tolerance
CAC	Call (or Connection) Admission Control
CBR	Constant Bit Rate
CCIR	Centre for Communications Interface Research
CCITT	Comité Consultatif International Télégraphique et Téléphonique
CCSN/SS7	Common Channel Signalling Network
CDF	Cumulative Density Function
CDV	Cell Delay Variation
CI	Congestion Indicator
CLR	Cell Loss Ratio
CLP	Cell Loss Priority
CLT	Central Limit Theory
CRC	Cyclic Redundancy Check
CTD	Cell Transfer Delay
DAR	Discrete AutoRegressive
DQDB	Distributed Queue Dual Bus
ER	Explicit Rate
fBm	fractional Brownian motion
fGn	fractional Gaussian noise
FLOM	Fractional Lower Order Moment
FTP	File Transfer Protocol
GCLT	Generalised Central Limit Theory
GCRA	Generic Cell Rate Algorithm
GFC	Generic Flow Control
GOP	Groups of Pictures

ISDN	Integrated Services Distributed Network
ITU-T	International Telecommunications Union - Telecoms Standards Sector
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LDT	Large Deviation Theory
LRCS	Long Range Correlation Structure
LRD	Long Range Dependence
MAN	Metropolitan Area Network
MCR	Minimum Cell Rate
MMPP	Markov Modulated Poisson Process
MPEG	Motion Picture Experts Group
NaN	Not a Number
NNI	Network Network Interface
NRT	Non-Real Time
PCI	Protocol Control Information
PCR	Peak Cell Rate
PSD	Power Spectral Density
PTI	Payload Type Identifier
QoS	Quality of Service
RNG	Random Number Generator
RT	Real Time
SAR	Segmentation And Reassembly
SCR	Sustained Cell Rate
SRCS	Short Range Correlation Structure
SRD	Short Range Dependence
SW	Shapiro-Wilk
TES	Transform-Expand-Sample
UNI	User Network Interface
VBR	Variable Bit-Rate
VCI	Virtual Channel Identifier
VOD	Video on Demand
VPI	Virtual Path Identifier
WAN	Wide Area Network
WSS	Wide Sense Stationary

Chapter 1

Introduction

1.1 Motivation

Overloading a digital network results in the loss and/or delay of information. This leads to a degradation in the quality of service (QoS) and is to be avoided. However whilst not wishing to overload a network, it is desirable to load it as close to the maximum as possible. This maximises the utilisation of the network and the profits made (if network usage is charged for). Therefore some mechanism is required that can determine the impact of traffic upon a network. This can take the form of traffic shaping [1], traffic policing [2] or call admission control [3].

In this thesis we concentrate on call admission control. This is a control mechanism for the asynchronous transfer mode (ATM) and its task is to decide whether a new traffic source should be allowed onto the network. It is therefore a function of the present state of the network, the effect the new source would have on the network state and the QoS demands of the new source. Basically the call admission control scheme must be able to answer the following question.

If I allow this traffic onto the network can I maintain its QoS demands whilst still maintaining all existing traffic's QoS demands?

If the answer to the above question is yes then the new traffic can be admitted to the network. However if the answer is no then the traffic must be refused entry to the network. Although the concept of a call admission control scheme is simple, implementation is much more complex. The complicating factors that this thesis addresses are highlighted below.

- We need to know what characteristics of a traffic source are relevant to its impact on a network and find ways of measuring them.

- We have to determine a link between these characteristics and the actual network performance. Only then can we use the estimates of the characteristics to determine how the traffic will impact on the network.

In order to simplify the problem mentioned in the first point above researchers have modelled the network as a single server queue. In this way the correlation structure and marginal distribution of a traffic source have been identified as the two factors that affect the queuing behaviour. This can be argued heuristically in that the marginal distribution controls the traffic source output values (in bits per second or cells per second etc.) whilst the correlation structure controls whether output values are correlated (i.e. are larger values likely to be grouped together?). However the relationship between correlation and distribution, and their effect on the single server queue, is non-trivial for all but the most basic of traffic sources [4].

The long-term aim of the work initiated in this thesis would be to develop a system similar to that in Figure 1.1.

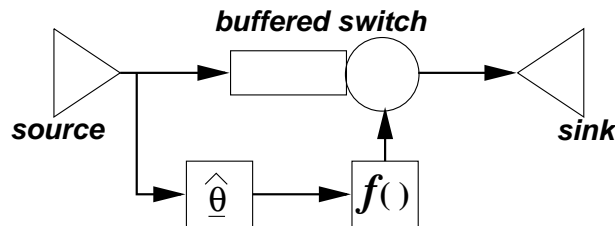


Figure 1.1: *The desired system: The parameter set is estimated and using this and some function, $f()$ a prediction can be made for the service rate of the buffered switch.*

To solve this system we must identify the parameters that make up the vector θ , devise suitably accurate and speedy estimate procedures for the parameters in θ and determine $f()$. This function's solution should be the minimum switch speed that maintains the source's QoS parameters. Although the complete development of such a system is beyond the scope of this thesis we will return to it in Chapter 8 and discuss the progress we have made.

1.2 Layout of thesis

The remainder of this thesis is divided into 7 chapters, the organisation of which is explained in this section.

Chapter 2 introduces the features of ATM which are relevant to this thesis. It attempts to show how ATM is able to cater with a wide range of traffic thanks to its use of service classes. We also introduce the concept of traffic characterisation with regards to CAC. The characterisation parameters identified by the ATM Forum (the body responsible for the specification of ATM) are introduced and we extend this by discussing effective bandwidths and parametric characterisation. We conclude Chapter 2 with some novel work which attempts to characterise ATM traffic using a structured Markovian model.

In Chapter 3 we introduce self-similarity and explain why it has had significant impact on research into teletraffic analysis and modelling. The properties of a self-similar process are discussed as are procedures for estimating a coefficient of self-similarity. We conclude by introducing some self-similar models which are used later in this thesis.

In Chapter 4 we apply two novel non-parametric characterisation measures to several VBR video datasets and a series of model datasets. In doing this we attempt to ascertain the correlation structure of the video data. We are not able to determine whether the data is self-similar or stationary. However we do conclude that self-similar models are best able to capture the data's behaviour.

In Chapter 5 we shift our focus to the other important traffic characteristic with regards to call admission control, the marginal distribution of the source. Much of the work to date has concentrated on self-similar models with a Gaussian innovation process. We question the validity of this assumption and present evidence that conclusively shows that this assumption fails for a wide range of teletraffic. In addition, we show the real data innovations are more impulsive than the Gaussian case. We reinforce this result by applying the Hill estimate and the generalised Pareto model which demonstrate that several of the datasets are heavy tailed.

Chapter 6 introduces the stable distribution and describes the properties that are relevant for teletraffic analysis. If these distributions are to be applied successfully we must be able to estimate stable parameters for real data. Several estimation techniques are available but they have never been compared directly, nor has their application to real data been studied in any depth. We rectify this in Chapter 6 and the first part of Chapter 7.

After considering the estimation of stable parameters for real data, Chapter 7 considers stable self-similar models. These tie together the first and second parts of the thesis by taking the correlation results from Chapter 4 and the stable distribution work from

Chapters 5 and 6. This Chapter concludes by considering the framework for two call admission control schemes which use stable distributions. The first examines the effective bandwidths of stable models whilst the second is concerned with the effect of the stable parameter α on queueing behaviour.

Finally, in Chapter 8, we summarise the main contributions of this thesis, draw conclusions from the work presented herein and discuss several areas of future work.

Call admission control on ATM networks

2.1 Introduction

In this chapter we introduce the Asynchronous Transfer Mode (ATM). This is an emerging broadband network switching technology and has been designed to cater for the networking needs of the early 21st century. We begin by explaining why there is a need for ATM and how ATM differs from existing network transfer modes.

For ATM to be successful we will need traffic characterisation schemes that allow networks to make call admission control (CAC) decisions. This is because ATM networks employ statistical multiplexing to achieve a high utilisation of network resources. Traffic characterisation and statistical multiplexing are discussed in Sections 2.2.8 and 2.3.

In the final part of this chapter we present novel work based on developing a structured Markovian model for ATM modelling and CAC. In Section 2.4.2 we link the model parameters to traffic characteristics and in Section 2.4.4 we go on to develop expressions for the moments of the models. We finish with a discussion on this work in Section 2.5 and draw conclusions in Section 2.6.

2.2 The asynchronous transfer mode

2.2.1 Introduction

Historically telecommunication networks have always been dedicated to a specific service e.g., television and telephony. This methodology has led to the development of separate networks each with its own management and maintenance costs. Each network

is well suited for the traffic it was intended to carry but implementing new technologies over an older network involves compromise and a reduction in the benefits that could be passed onto the customer e.g., facsimile over the telephony network.

The Comité Consultatif International Télégraphique et Téléphonique (CCITT) (now the International Telecommunications Union - Telecoms Standards Sector (ITU-T)) recognised the problem of dedicated networks and developed ISDN, an Integrated Services Distributed Network. Now there was a network that was not dedicated to any one service but could transfer voice, video or electronic data with speeds of up to 2 Megabits per second (Mbs^{-1}). This one network could replace many of the networks in existence and hence reduce overall running costs.

As ISDN was being developed and implemented the CCITT was already aware that a successor would be needed. They recognised that new services were being developed that required very high bandwidths and that ISDN would not be able to transfer this volume of traffic. Some of the examples of broadband traffic are given in Figure 2.1.

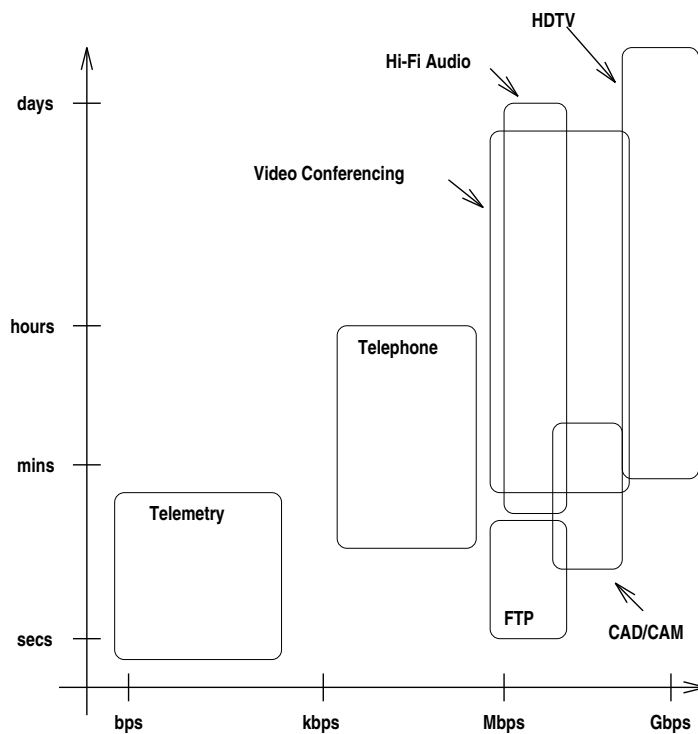


Figure 2.1: A graph showing the holding time and bandwidth of the more popular broadband traffic types.

A network able to cope with broadband traffic would have to have a bandwidth of hundreds of Megabits per second and with the emergence of optic fibres and associated technologies such bandwidths were not inconceivable. So a new network was proposed

that could carry the broadband services as well as those already in existence on ISDN.

Broadband ISDN (B-ISDN) is that new network and Asynchronous Transfer Mode (ATM) is the technology that has been chosen as the transfer mode for B-ISDN. Sub-section 2.2.2 explains why ATM was chosen and the remainder of Section 2.2 describes the properties of this technology.

2.2.2 B-ISDN transfer mode

The CCITT had to choose a transfer mode for Broadband ISDN that would be responsible for the routing and switching of information across the network. Such a mechanism would have to be fast to cope with the high speeds the network was expected to work at. The mechanism would also have to be flexible to be able to transfer both high bandwidth and low bandwidth services simultaneously.

There were two transport modes already in existence which possessed some of the properties that B-ISDN required.

- *Circuit Mode* - In circuit mode the connection between start and end points is established before any data is transferred. This requires a complicated connection setup protocol but once the connection is established information can be transferred at very high speeds.

The bandwidth of a circuit mode connection is fixed at call setup so dynamic bandwidth allocation is not possible and as the variety of channel bandwidths increases, implementation becomes more complicated.

- *Packet Mode* - Packet mode is suitable for data communications over low quality networks (e.g., X.25) because it incorporates complex error detection procedures. However, these require software switching which reduces the speed of the network. Data is sent in blocks called packets which are only transmitted when information is being transferred, this makes it suitable for heterogeneous traffic.

The CCITT recognised that circuit mode could provide the speed required for B-ISDN while packet mode was better suited to transfer the heterogeneous traffic. The solution was to propose a new transfer mode that incorporated the features of circuit and packet modes and which facilitated the high speed transfer of heterogeneous traffic likely on such networks. This target technology has developed into ATM.

ATM is a *cell switching* technology where data is packed into fixed length blocks called

cells. A complicated call connection protocol ensures that these cells can be transferred at high speed. In addition, by controlling the number of cells transmitted per unit time per customer ATM is able to transfer different services simultaneously.

2.2.3 The ATM cell

The ATM cell has a fixed length of 53 bytes. This is split into a header field of 5 bytes and a data field of 48 bytes. The choice of 53 bytes was not an arbitrary one but was the compromise value chosen after a conflict of interests between those planning to utilise ATM for telephony switching and those planning to utilise ATM for data switching. The pro telephony switching companies were interested in small cell sizes to reduce the jitter introduced by grouping the data into cells. On the other hand the pro data switching companies preferred larger cell sizes to reduce the overhead of the cell header [5].

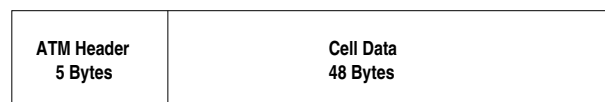


Figure 2.2: *The ATM cell structure.*

The 5 byte ATM cell header can be broken down as shown in Figure 2.3.

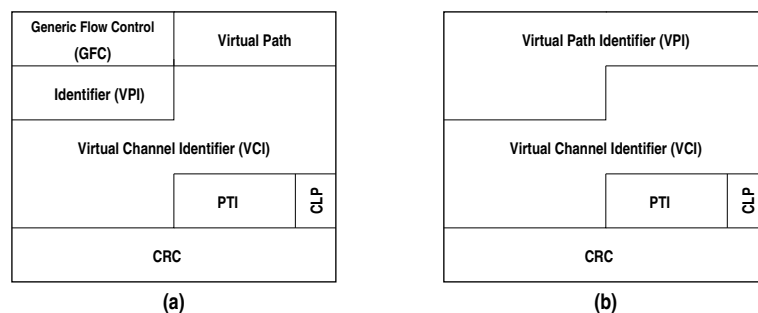


Figure 2.3: *The ATM cell header format for (a) the user network interface (UNI) and (b) the network network interface (NNI).*

The header is divided into different fields associated with routing, network management and error correction.

- **Generic Flow Control** - This 4 bit label is only present in cells at the user network interface (UNI). It provides a unique code which allows a cell to negotiate with a shared access network.
- **Virtual Path Identifier** - Since ATM is a connection orientated mode each

cell must contain some sort of routing information to ensure it can be correctly directed through each switching node. The VPI is the first of two parts of a cell's routing code. It defines a general area for the destination and not a precise location.

- **Virtual Channel Identifier** - The second part of the code provides information on routing at a more precise level. The VPI and VCI are used in a look-up-table within each ATM switch.

The VPI and VCI are reallocated on a link by link basis meaning each switch need only ensure the identifiers on its links are unique. A much larger network can now be constructed than would be otherwise possible with only 2^{28} unique combinations of VPI and VCI.

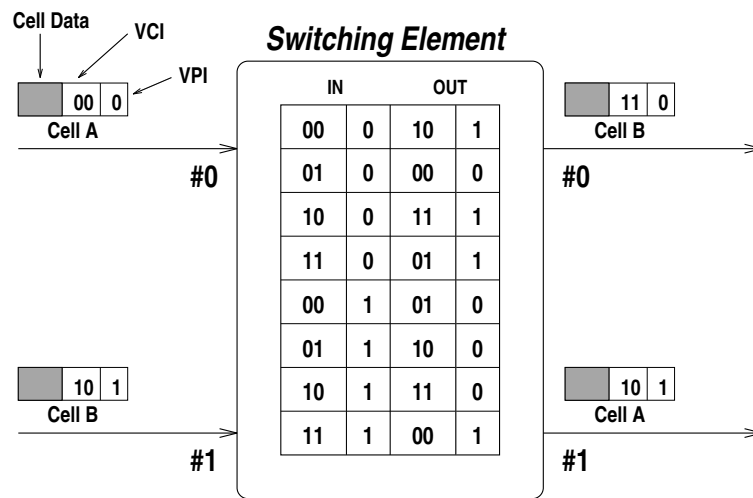


Figure 2.4: A simplified example of the method used to route cells in ATM switches.

Figure 2.4 illustrates the switch process using a simplified model with a VPI and VCI of lengths 1 bit and 2 bits respectively. The look-up-table enables the node to switch the cells correctly and change the cells' VPIs and VCIs to their new values.

- **Payload Type** - The payload field is used to distinguish between customer data and network generated cells concerned with administration.
- **Cell Loss Priority** - The CLP field is just 1 bit in length and is used as a flag to mark those cells which are of the lowest priority. These cells are the first to be discarded when congestion occurs.

The CLP is also used to flag cells of customers abusing their contract with the network. For example if a customer has a contract for 10Mbs^{-1} peak output but is actually outputting at 10.01Mbs^{-1} then the extra cells can have their CLP bit

set to one. Any cells that are flagged will be discarded if necessary but may pass unhindered if the network is lightly loaded.

- **Header CRC** - The error correction is only computed over the header of the cell. The polynomial used is $x^8 + x^2 + x + 1$ which can correct a single error and detect many multiple bit errors.

2.2.4 Service classes

The ATM Forum attempted to find some classification system for the services expected on B-ISDN. In this way they could identify the necessary transfer requirements for a few broad traffic classes and pack user data in an efficient manner for each class. They decided upon four major service classes and these are given in Table 2.1.

Service Class	Description
Constant Bit Rate (CBR)	This class is used for circuit emulation, the source must send blank cells if necessary to maintain a constant bandwidth.
Variable Bit Rate (VBR)	There are two types of VBR service. The first is real time (RT) services which are delay sensitive. These might include compressed video and voice. The second VBR type is non-real time (NRT) and an example of this might be video mail.
Available Bit Rate (ABR)	This class was developed for normal computer data, whether it be text or binary. This class works by using as much bandwidth as is left on the network. So as congestion increases, ABR bandwidths are reduced.
Unspecified Bit Rate (UBR)	This class is similar to ABR and is expected to transfer the same types of services but it gives no bounds on cell loss or delay. It is often referred to as a "best effort" class.

Table 2.1: General B-ISDN service classes as defined by the ATM Forum.

In order to facilitate the efficient transfer of all these services several methods are used to pack data into the user 48 bytes of an ATM cell. These methods are termed the ATM adaption layers (AALs) and each AAL has a different segmentation and reassembly (SAR) scheme that performs the packing and unpacking in a manner optimised for a specific service class (see Section 2.2.5).

2.2.5 Adaption layers

Information passed over an ATM network has to be packed into the data fields of ATM cells. This packing is carried out in accordance with an ATM adaption layer (AAL). Five AALs exist and they differ to suit the different forms of services that must be provided.

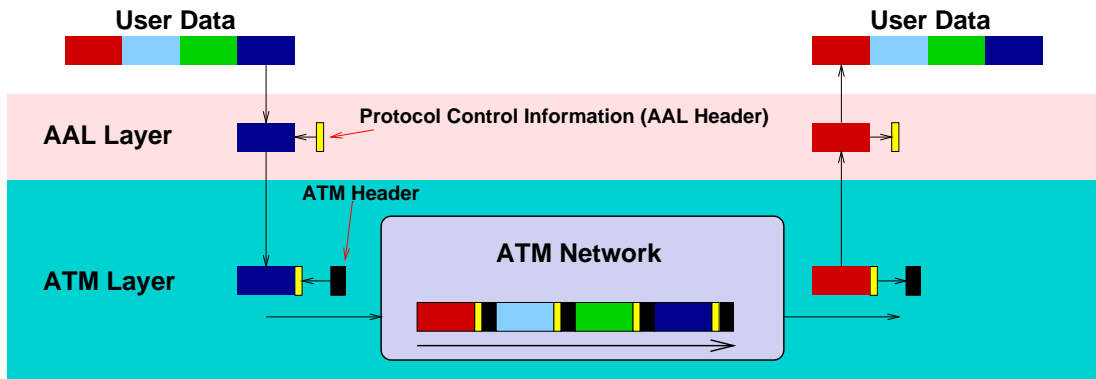


Figure 2.5: *How the ATM header and the AAL are used to transfer data across ATM networks.*

Figure 2.5 illustrates the two layer packing process, each with its own overhead. The AAL layer overhead is the protocol control information (PCI) and its length varies on the AAL number. The ATM header is a fixed length and contains information on routing, error detection and cell priority.

- **AAL0** - This is the null case where the user data is placed into the full 48 bytes available. The only operations performed by the SAR layer is to split the data into 384 bit blocks and to put these blocks back together again at the far end.
- **AAL1** - Designed for CBR services this AAL places a single byte of information into the protocol control information PCI, leaving 47 bytes for user data. This PCI contains a sequence counter so that lost or mis-inserted cells can be detected. As this AAL may have to package RT data, each cell is time stamped with a synchronous residual time stamp (SRTS), to facilitate CDV reduction when reassembling the data at the destination.
- **AAL2** - AAL2 was designed to be efficient at packing VBR services into ATM cells. The main concern was to ensure that synchronisation of data between endpoints is maintained in the case of RT services. In addition the layer has the ability to detect lost cells and errors within the cell. The ability to carry out

these tasks requires space in the PCI so for AAL2 it is 3 bytes long, leaving 45 bytes for user information.

- **AAL3/4** - These two adaption layers were initially designed to pack and unpack connection orientated and connectionless computer data services. However they were merged when it was realised the requirements of both types of service were very similar. However in both cases a PCI of 4 bytes is required. Many considered this PCI to be too large and overly complex for most data services and developed a simpler alternative, AAL5.
- **AAL5 - The simple and efficient layer (SEAL)** - AAL5 is expected to be used more often than AAL3/4 because it requires no PCI header at all leaving all 48 bytes per cell for user data. Instead AAL5 uses part of the cell header to mark cells that contain the beginning and the end of user packets. This AAL is distinctly different from the others in that it can affect the ATM layer and therefore does not conform to the model in Figure 2.5. However it was decided that this infringement of dependency was outweighed by the increase in user data space and the reduction in AAL layer complexity.

2.2.6 Network topology

An ATM network has a structured hierarchy, the transport medium (e.g., optic fibre) provides a transport service for the virtual path (VP) and the VP provides a transport service for the virtual channel (VC). It is the VC that provides a unidirectional transport mechanism for the cells.

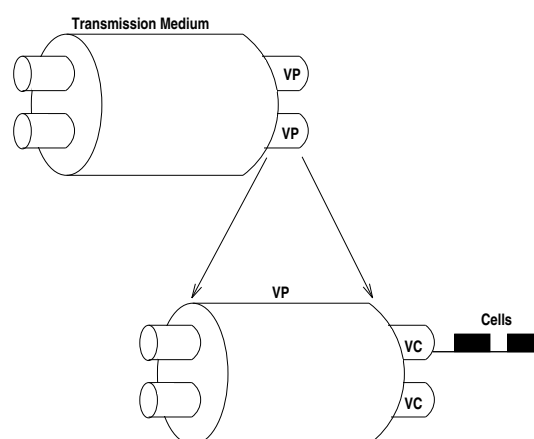


Figure 2.6: *ATM network hierarchy*

Traffic enters an ATM network via a User Network Interface (UNI) which can be

connected to a computer network, a video conference facility or any other B-ISDN client. Once the virtual channel connection is established information can be transported to its destination routed by the Network Network Interfaces (NNI).

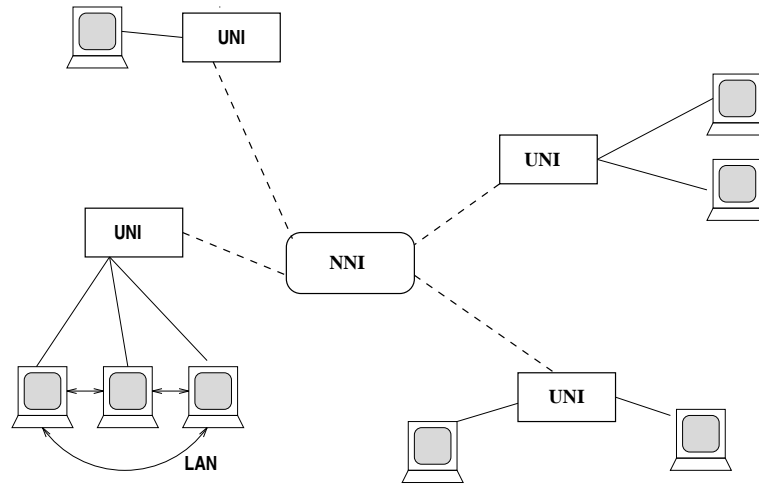


Figure 2.7: *The topology of an ATM network*

2.2.7 Network control and management

As well as concerning itself with managing the network at the cell level, there are other management issues that ATM must consider within a practical network. These issues include quality of service (QoS), tariffing and network control.

- **Quality of Service** - When a customer forms a contract with a network provider there will be a specification of the customer's expected quality of service (QoS). This QoS figure will be guaranteed by the network provider and if poorer performance is provided the supplier is in breach of contract. The QoS can be specified using parameters like guarantee of availability of bandwidth, tolerances on cell delay variation (CDV) and limits on cell delay and cell loss rate.
- **Tariffing** - When a customer uses a network service some charge should be incurred, however charging for unused services is unfair. The network provider must ensure that a customer's bandwidth is monitored correctly and charged accordingly. The same monitoring that checks QoS could also be responsible for tariffing.
- **Control** - Three levels of control have been identified in ATM networks.
 1. **Network Control** - This level of control is concerned with traffic routing and link capacity assignment.

2. **Call Control** - This is the level of control that we consider in this thesis. It is concerned with allowing new customers onto the network without increasing cell loss and/or delay to an unacceptable level (i.e. call admission control (CAC)).
3. **Cell Control** - Cells flagged with their CLP bit can be discarded if necessary. If the network becomes congested then cell control attempts to sacrifice the low priority cells whilst saving the high priority cells.

2.2.8 Asynchronous multiplexing and statistical gain

In synchronous multiplexing networks a connection is allocated a fixed bandwidth, B_i , for a connection of type i . If N_i users have connections of type i on the same link and the total capacity of the link is C then the remaining bandwidth B_r is easily calculated,

$$B_r = C - \sum_i N_i B_i. \quad (2.1)$$

The case of $i = 1$ (i.e. single channel capacity) is presented as case 1 in Figure 2.8. The CBR service (on channel B) uses all its timeslots on the multiplexed channel but both the VBR services waste timeslots (the grey slots). This is because synchronous schemes allocate the space according to the channel size and make no allowance for the amount of bandwidth actually used by the service. This scheme is bandwidth efficient when the mean transmission rate is close to the peak transmission rate but this will not be the case for much of the traffic on B-ISDN [6]. It is possible to transfer bursty services efficiently over ISDN using a service called frame relay. This involves grouping the data into packets and transferring these using ISDN, however this grouping introduces delay and is therefore inappropriate for real time (RT) services.

ATM has to manage a combination of CBR and VBR services with a wide range of bandwidth characteristics and with varying sensitivities to data loss and delay. The standardising bodies could have chosen a synchronous multiplexing scheme with a number of different sized channels (i.e. $N > 1$). However in the case of very bursty services (such as remote database access) the channel would be seriously under-utilised for the majority of the time. This scheme is presented as case 2 in Figure 2.8 and is comparable to using combinations of the basic rate interface on ISDN to transfer services with peak rates in excess of 64 kb/s.

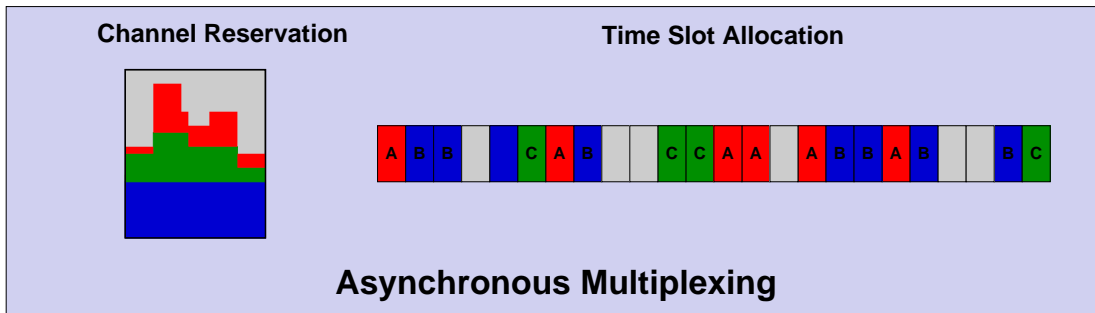
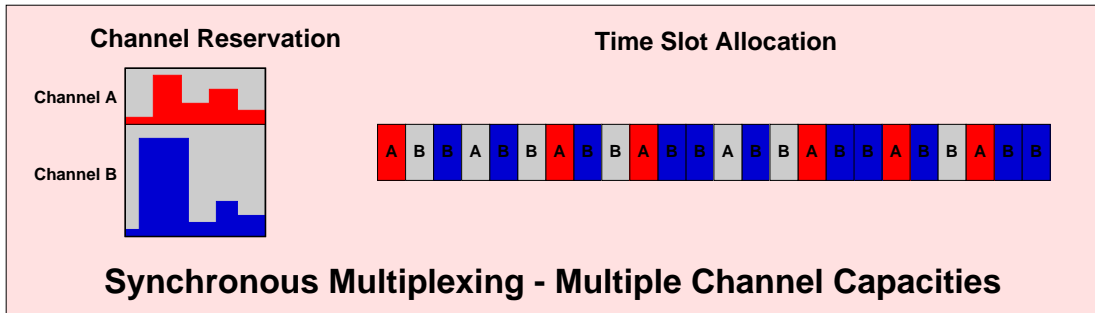
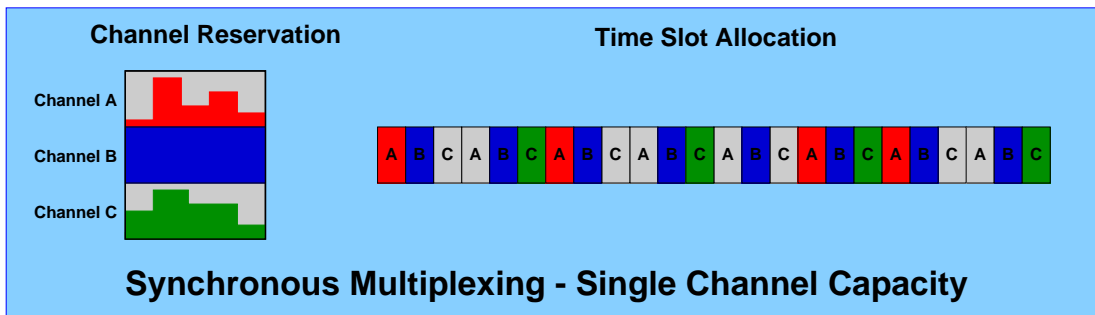


Figure 2.8: *Statistical Gain* - This comparison illustrates how asynchronous multiplexing can increase link utilisation. It is important to note that in the asynchronous case that the empty timeslots have not been reserved and may be utilised by other services.

Instead ATM uses asynchronous multiplexing to achieve statistical gain which allows network providers to utilise bandwidth more efficiently (case 3 in Figure 2.8). In asynchronous multiplexing timeslots are allocated to a channel only when they are required. Therefore when a service is transmitting a burst it will be allocated more timeslots than those services that are between bursts. This is evident by the fact that no strict timeslot allocation scheme is adhered to in case 3 of Figure 2.8 (case 1 uses ...ABCABCABC... and case 2 uses ...ABBABBABB...). Statistical gain is achieved by noting that the probability of all channels transmitting at their peak rate simultaneously is small. In the diagram this gain is represented by the unallocated timeslots (in grey). These slots have not been reserved for a channel (remember ATM uses virtual connections with no fixed bandwidth) so they can be used by another service on this link. A simple experiment that illustrates statistical gain is given in Appendix B.

2.2.9 Traffic parameters

In order to estimate the bandwidth of a service over the duration of a virtual connection ATM requires parameters from the users. As well as parameters that describe the traffic stream, the network also needs information concerning the service's tolerance on cell loss, delay and jitter (CDV). The parameters that the ATM Forum have specified in UNI standards are given below.

Peak Cell Rate (PCR)	The maximum rate at which cells are transmitted. It can also be thought of as the inverse of the minimum inter-cell arrival time.
Sustained Cell Rate (SCR)	The average rate at which cells are transmitted (inverse of the mean inter-cell arrival time).
Minimum Cell Rate (MCR)	The minimum rate of cell transmission.
Cell Loss Ratio (CLR)	The ratio of lost cells over total number of transmitted cells.
Cell Transfer Delay (CTD)	The time taken for a cell to cross the network.
Cell Delay Variation (CDV)	The variance of CTD.
Burst Tolerance (BT)	The maximum number of cells allowable in a burst at the PCR.

2.3 Traffic characterisation

2.3.1 Introduction

If ATM is to be able to allocate resources to network users whilst maintaining the QoS requirements of both new and existing calls some form of bandwidth allocation scheme is required. However efficient use of resources will only be achieved when this allocation is as close as possible to the minimum required to maintain QoS demands.

There has been a lot of discussion about how to characterise traffic on ATM networks. If we think about how a link will be used, with regard to the four service classes, network providers might hope to achieve something similar to Figure 2.9. The CBR services are simple to characterise using the parameters in Section 2.2.9. ABR schemes use feedback to ensure they maintain a suitable transmission level, whilst UBR fills the last few percent of the link capacity. So the main concern of this thesis is VBR services, which will include MPEG video, compressed voice and computer communications.

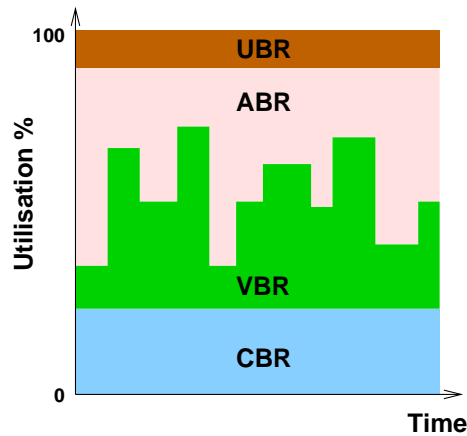


Figure 2.9: *How a typical ATM link might be utilised. The ABR services uses feedback to react to the changes in the VBR services and the UBR services fills the remaining space left by the other three.*

The danger with regards to network performance is that many VBR RT services will try and transmit data at their PCRs simultaneously. The total bandwidth may then exceed the available link capacity and hence create overload and loss. This is the down side of the statistical gain we introduced in Section 2.2.8.

A lot of work has been done to try to fit the VBR streams to models because by doing this it is possible to solve capacity problems with the model and assume that the result will be the same for real services (see Appendix B for a simple example of this principle). Over the last few years these models have become more realistic but also more complex.

One problem is that there are no closed form solutions for the overflow probabilities of the more complex models so they have to be estimated by simulation. Techniques such as effective bandwidths have therefore become popular as they allow loss estimates to be made on complex traffic streams without the need for lengthy simulations. Instead a source's effective bandwidth can be estimated directly from the traffic stream. In the next section we introduce effective bandwidths and show how these can be used to characterise teletraffic.

2.3.2 Effective bandwidths

The effective bandwidth of a VBR source is a value that lies somewhere between its SCR and PCR and is related to the network performance by a loss coefficient [7]- [11]. Work carried out by Hsu and Walrand [12]-[14] and Dublin Institute of Advanced Studies

and Cambridge University Computing Lab [15]-[19] suggest that they have potential for CAC on ATM.

The effective bandwidth of a VBR source is a value between the mean and peak cell transmission rate of that source. However this value is affected by the correlation structure of the source and a loss parameter that is chosen to match the QoS demands of that source. If the source is then serviced at its effective bandwidth it will conform to the QoS constraints imposed. In addition if several sources are serviced simultaneously at one switch and all are serviced at their effective bandwidths, then their QoS demands will not be violated [9], [20].

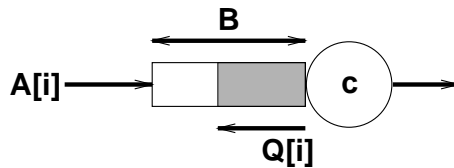


Figure 2.10: A single server queue. $A[i]$ is a discrete arrival process, c is the service rate, B is the buffer size and $Q[i]$ is the size of the queue in the buffer.

We can state this more formally. Consider the single server queue in Figure 2.10. The VBR video service is a discrete time stochastic arrival process, $A[i]$, and $Q[i]$ is the queue size of the buffer at time-slot i . If the queue is serviced at a constant rate, c then the queue-length at any time-slot can be calculated using the recursion

$$Q[i] = (Q[i-1] + A[i] - c, 0)^+, \quad (2.2)$$

where $(a, b)^+$ is defined as the maximum out of a and b . Now define the scaled cumulant generating function

$$\Lambda(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E[\exp(\theta(A[1] + \dots + A[n]))]. \quad (2.3)$$

If certain assumptions about the arrival process [20] (i.e. it is stationary and mixing) and the stability of the queue (i.e. $E(A[i]) < c$) are met then we can say

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log P(Q[i] > B) = -\theta^*, \quad (2.4)$$

where $\theta^* = \sup\{\theta > 0 : \Lambda(\theta) < \theta c\}$. We then define the effective bandwidth of the arrival process,

$$\delta(\theta) = \frac{\Lambda(\theta)}{\theta}. \quad (2.5)$$

So the theory of effective bandwidths links an arrival process to the probability of buffer overflow via some constant. For this reason it has the potential to be employed in CAC schemes for ATM networks.

To illustrate how effective bandwidths can be used to predict bandwidth requirements of stochastic sources consider the following example.

An ON-OFF model is a source that generates p ATM cells when it is ON and 0 cells when it is OFF. The source is ON with probability ψ and we use the source to model the number of cells/frame of an VBR video encoding. Therefore each frame is iid and either has the value of p bits or 0 bits. We matched the model to a VBR video encoding and obtained parameter values $p = 473$ cells and $\psi = 0.068$. The effective bandwidth of the ON-OFF model is simple to calculate and can be given as

$$\frac{\Lambda(\theta)}{\theta} = \frac{1}{\theta} \log [1 + \psi(e^{\theta p} - 1)]. \quad (2.6)$$

By generating sample data using the ON-OFF model we estimated the effective bandwidth for a variety of θ using (2.3) and by plotting the buffer occupancy probability we measured the observed θ (using (2.4)).

We chose $\theta = 1 \times 10^{-3}$ which for a buffer size of 1000 cells equates to a loss rate of approximately 0.368 when serviced at the effective bandwidth. This is too high for a realistic ATM scenario but reduces the amount of simulation required to verify the accuracy of effective bandwidths which is the purpose of this example.

The exact effective bandwidth of the ON-OFF source is 40.303 cells/s which is greater than the mean of the model (32.164 cells/s). The estimated effective bandwidth for varying sizes of sample sizes, N is plotted in Figure 2.12.

It is clear that (2.3) provides a good estimate of the exact value of the effective bandwidth if we allow for the convergence of the algorithm. Next we wished to demonstrate that if the ON-OFF source is serviced at a rate equal to its effective bandwidth then the QoS requirement given in (2.4) will be met. We used 100000 arrivals to simulate a simple queuing scenario and plotted the negative of the natural logarithm of the buffer occupancy probability versus the buffer size (Figure 2.11). The slope of this curve was estimated to be 0.95238×10^{-3} which is within five percent of the expected value.

This simple example illustrates how effective bandwidths can be used to determine resource requirements on an ATM network. Unfortunately most of the traffic expected

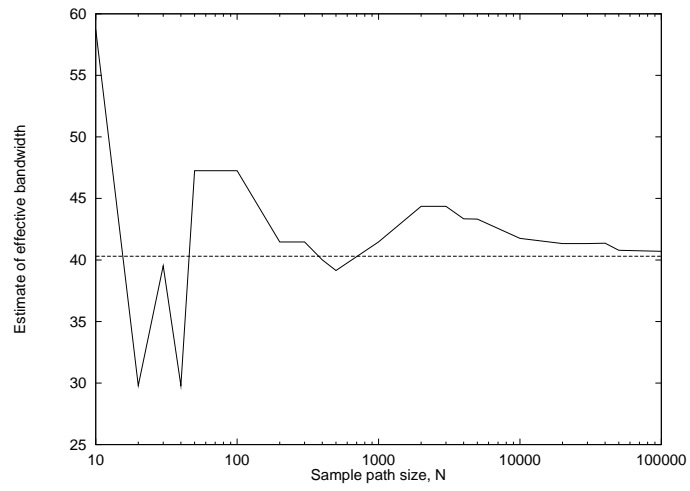


Figure 2.11: *The estimated effective bandwidth for the ON-OFF model (dashed line = the exact value).*

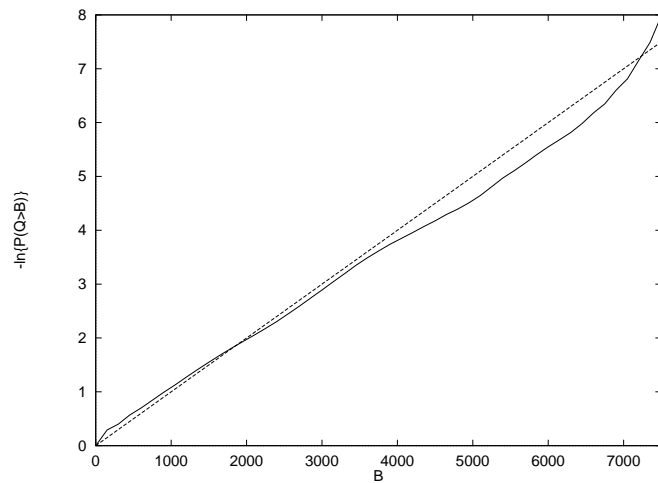


Figure 2.12: *The buffer occupancy plot estimated by Monte Carlo simulation (dashed line = the exact value).*

on these networks is unlikely to conform to an ON-OFF model, so exact expressions for the effective bandwidth (as in (2.6)) are not available. The Dublin Institute of Advanced Studies and Cambridge Computing Laboratories have published several studies into estimating effective bandwidth on-line for real data [15], [16]. Their results are positive and seem to suggest that, if applied with care, effective bandwidths have the potential to be applied to CAC schemes. For recent results on measurement based effective bandwidth estimation for CAC see the paper by Gibbens and Kelly [21].

2.3.3 Teletraffic modelling

Teletraffic models are important for several reasons. Firstly they allow designers to investigate how a network will behave under loads and conditions that are not normally encountered. Secondly they are less expensive and more flexible than experimental test-bed networks and thus speed up the research process for novel network technologies. Thirdly they can be used for parametric CAC schemes.

The argument for parametric CAC is as follows: If a model can be found that matches the salient characteristics of the data and if there exists a direct relationship between the model's parameters and its performance on an ATM network then the model can be used to estimate the network performance of real ATM services. In practice we would attempt to estimate the model parameters from the real data and then calculate the network performance by assuming that the model and data performance will be similar. The problematic part of this paradigm is finding models that are complex enough to match the salient characteristics of the data yet simple enough to enable estimation of performance parameters.

As an example the ON-OFF model discussed in Section 2.3.2 is a parametric model. It has the advantage that a closed form expression for its effective bandwidth is known (equation (2.6)) but from experiments [22] and real traffic studies [23] we know that such a model is not a realistic representation of real VBR video data.

In Section 2.3.3.1 we discuss the research that has been carried out in this area whilst in Section 2.4 we concern ourselves with the work we carried out in adapting structured Markovian models for ATM.

2.3.3.1 Parametric CAC for ATM: A discussion

A great deal of work has been carried out in the area of ATM modelling. This is because parametric CAC would be much easier to implement and much more reliable if a tractable model can be found that accurately represents the data. Some of the models that have been applied to this problem include Markov modulated Poisson processes [24]-[26]. Elwalid and Mitra considered a more general range of Markov processes (both fluid and point processes) and managed to compute the effective bandwidth of these sources [27]. Kelly calculated the effective bandwidth for several models in [10] and summarised the history of effective bandwidths to date. Melamed and associates are advocates of the TES (Transform-Expand-Sample). This model was used in an attempt to characterise VBR video [28] as well as being used to evaluate network performance of teletraffic data [29].

Although CAC control schemes have been developed for many kinds of models the problem of applying such techniques to real data remains problematic. In the effective bandwidth example in Section 2.3.2 the model was simple and generated an iid arrival process. In real teletraffic scenarios we know that signals possess more complicated correlation structures. Non zero correlations cause problems because they mean that bursts can occur in the data, therefore successive large arrival values may occur causing buffer overflow and a degradation in QoS. In Chapters 3 and 4 we investigate this correlation problem in greater detail.

2.4 The adaption of a structured Markovian model for ATM

2.4.1 Introduction

In this section we present a structured Markovian model [30] and illustrate how such a model can be extended to conform to popular ATM traffic parameters. We go on to show how the model can be adapted to suit a variety of traffic types while still adhering to the traffic parameters that are known *a priori*.

We produce a solution surface for traffic with common average bit-rate (ABR) and peak bit-rate (PBR) and show how different traces with the same common characteristics

can be generated by choosing different points on the solution surface.

2.4.2 The structured Markovian model

Consider a Markovian model with state vector \underline{S} and N types of traffic. Then for each type of traffic let there be M possible states. The size of \underline{S} is $(N \times M)$ and the transition matrix for \underline{S} , \mathbf{Q}_T , has dimension $(N \times M)$ by $(N \times M)$.

Consider a model with 4 states per traffic type (i.e. $M = 4$). These are defined as state O_i (null), state A_i , state T_i and state E_i . The null state is common to all traffic types so $O_i = O_j \{i, j \in 1, 2, \dots, N\}$ and only one null state, O , is required. State E_i is the transmission state for traffic type i and it is only in this state that cell transmission occurs. Using the nature of the model proposed in [30] we can introduce sparsity. This is due to the structure of the model which forbids certain state transitions. We can define

$$\underline{P}_0 \equiv [\rho_{OO} \quad \rho_{OA_1} \quad \dots \quad \rho_{OA_N}], \quad (2.7)$$

where \underline{P}_0 is a vector containing the transition probabilities from state O . For each of the traffic types to be defined there exists a transition matrix $\mathbf{P}_i \{i : i \in 1, 2, \dots, N\}$,

$$\mathbf{P}_i \equiv \begin{bmatrix} \rho_{OO} & \rho_{OA} & 0 & 0 \\ 0 & \rho_{AA} & \rho_{AT} & 0 \\ 0 & 0 & \rho_{TT} & \rho_{TE} \\ \rho_{EO} & \rho_{EA} & \rho_{ET} & 0 \end{bmatrix}. \quad (2.8)$$

Each state transition has an execution time associated with it. These times can be represented by N matrices, $\mathbf{T}_i \{i \in 1, 2, \dots, N\}$,

$$\mathbf{T}_i \equiv \begin{bmatrix} \tau_{OO} & 0^\dagger & 0^* & 0^* \\ 0^* & \tau_{AA} & 0^\dagger & 0^* \\ 0^* & 0^* & \tau_{TT} & \tau_{TE} \\ 0^\dagger & 0^\dagger & 0^\dagger & 0^* \end{bmatrix}. \quad (2.9)$$

0^* - Corresponding state transition has zero probability.

0^\dagger - Corresponding state transition has no time factor associated with it.

With the transition vector (2.7), matrices (2.8) and with the associated time matrices (2.9), we have described the whole model. Now we wish to determine how we can relate the parameters of the model to the characteristics of real traffic.

We can represent (2.8) in diagrammatic form. The resultant state transition diagram is often easier to interpret than a matrix because the structure of the model can be seen (see Figure 2.13).

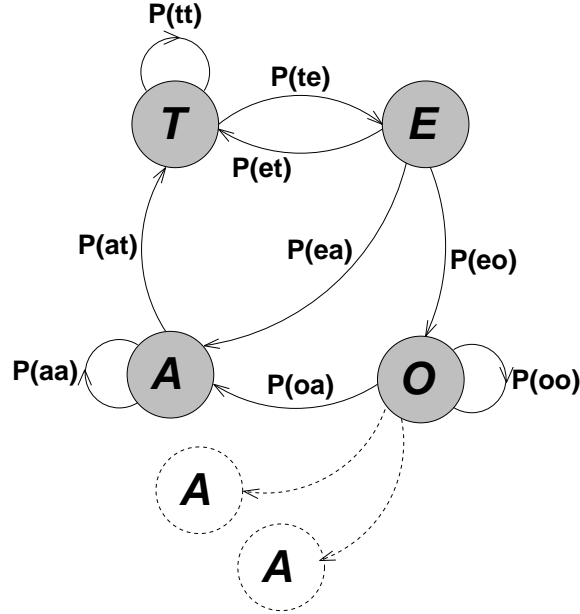


Figure 2.13: *State transition diagram for the structured Markovian model.*

Upon examination it is obvious that the smallest possible loop between two cell transmissions is the path $\dots \rightarrow T \rightarrow E \rightarrow T \rightarrow \dots$. If PCR is defined as the peak cell transmission rate of the service then it is possible to say

$$\text{PCR} = \frac{1}{\tau_{te}}. \quad (2.10)$$

The PBR and PCR can be linked by the fact that ATM cells are of a constant length and that, for a particular service, the cell overhead will be known *a priori*. Let ψ be defined as the number of data bytes in an ATM cell of the service type under consideration. Then

$$\text{PCR} = \frac{\text{PBR}}{8\psi}, \quad (2.11)$$

similarly

$$\text{SCR} = \frac{\text{ABR}}{8\psi}. \quad (2.12)$$

For a constant bitrate service $\text{PCR} = \text{SCR}$ and the model can completely describe the service with no additional calculations (assuming that cell delay variation is negligible). However for a variable bit rate service $\text{SCR} < \text{PCR}$ and this implies that the model must modulate cell transmission using other state transitions with times associated with them. From (2.9) we know that these transitions must be $A \rightarrow A$ and $T \rightarrow T$. If we define $\kappa(S)$ as an entry into state S , followed by n recurrent loops in state S before exiting state S ; where ($S : S \in A, T$) and ($n : n \in 1, 2, \dots$). Then define $\vartheta(S)$ as the expected time of $\kappa(S)$ in seconds. It is possible to say that the SCR must be a function of PCR, $\vartheta(A)$ and $\vartheta(T)$. From [30] and [31] we know that the state transitions of the model are geometrically distributed and that the expected number of occurrences of a state, S , before a transition is

$$E(\kappa(S)) = \frac{1}{1 - \rho_{ss}}. \quad (2.13)$$

Accordingly, it is possible to state

$$\vartheta(A) = E(\kappa(A)) \cdot \tau_{aa} = \frac{\tau_{aa}}{1 - \rho_{aa}}, \quad (2.14)$$

$$\vartheta(T) = E(\kappa(T)) \cdot \tau_{tt} = \frac{\tau_{tt}}{1 - \rho_{tt}}. \quad (2.15)$$

The next step in the derivation is to consider the model in state E . A transition must occur to either state O , state A or state E . A transition into state O does not affect the SCR, however, it is linked to the model parameter ρ_{eo} . If T_{av} is defined to be the average duration of an active period then,

$$T_{av} = \frac{1}{\text{SCR} \rho_{eo}}. \quad (2.16)$$

The two other possible state transitions from state E modulate the cell transmission rate. The loop which commences with state A can be defined as A_l and the loop commencing with state T as T_l . Due to the nature of the model

$$\vartheta(A_l) = \vartheta(A) + \vartheta(T) + \frac{1}{P_{ctr}}, \quad (2.17)$$

and

$$\vartheta(T_l) = \vartheta(T) + \frac{1}{P_{ctr}}. \quad (2.18)$$

The fact that the distributions are geometric in nature means we can form an expression for the expected number of A_l loops before a transition to T_l loops (assuming ρ_{eo} is negligible or ρ_{ea} and ρ_{et} are conditional)

$$E(\kappa(A_l)) = \frac{1}{1 - \rho_{ea}} = \frac{1}{\rho_{et}}. \quad (2.19)$$

Using a similar method for T_l yields the result

$$E(\kappa(T_l)) = \frac{1}{1 - \rho_{et}} = \frac{1}{\rho_{ea}}. \quad (2.20)$$

The expected behaviour of the model therefore is to perform $E(\kappa(A_l))$ A_l loops followed by $E(\kappa(T_l))$ T_l loops before repeating itself. Over a large number of iterations it is possible to use the expected value to form an equation for SCR. Over one period the number of cells transmitted is

$$E(\kappa(A_l)) + E(\kappa(T_l)) = \frac{1}{\rho_{et}} + \frac{1}{\rho_{ea}} \quad (2.21)$$

and the time taken to transmit this number of cells is

$$\begin{aligned} E(\kappa(A_l))\vartheta(A_l) + E(\kappa(T_l))\vartheta(T_l) = \\ \frac{1}{\rho_{et}}\vartheta(A_l) + \frac{1}{\rho_{ea}}\vartheta(T_l). \end{aligned} \quad (2.22)$$

The SCR can be found using (2.21) and (2.22)

$$\text{ABR} = \frac{1}{\rho_{ea}\vartheta(A_l) + (1 - \rho_{ea})\vartheta(T_l)} \quad (2.23)$$

The solution of (2.23) for a known SCR and PCR is a 3 dimensional surface with $\vartheta(A)$, $\vartheta(T)$ and ρ_{ea} as the variables. A model that conforms to the demanded service parameters can be constructed from any point on the solution surface although parameters that have not been defined such as variance will vary. It is the purpose of Section 2.4.3 to investigate how these parameters behave when solutions are chosen from across the surface and whether the model can be made to fit traffic characteristics by using this surface.

2.4.3 A modelling example

Consider an ATM traffic service with an ABR of 1Mbs^{-1} , PBR of 6Mbs^{-1} and an average holding time, T_{av} , of 600sec. In Figure 2.14 a solution surface for this service on a 600Mbs^{-1} ATM network is given.

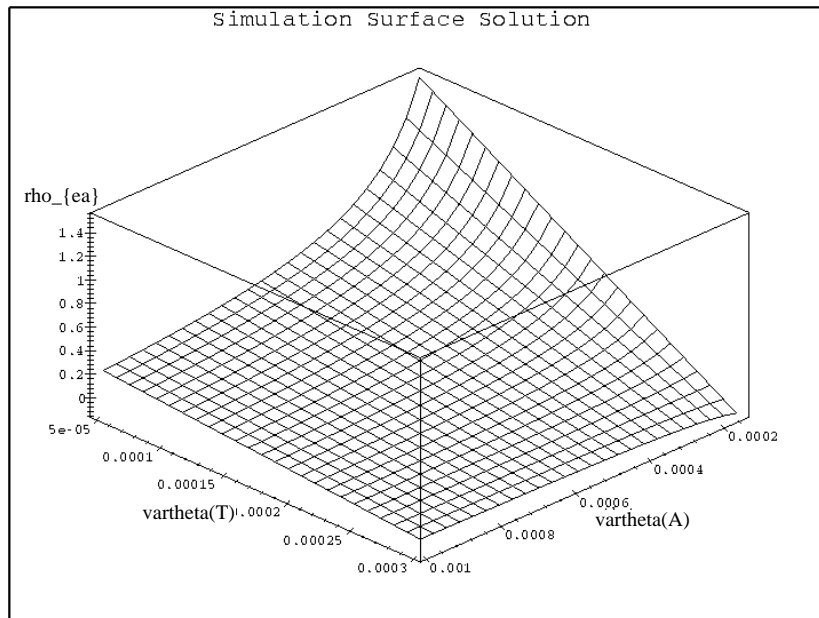


Figure 2.14: *Solution Surface for $ABR=1\text{Mbs}^{-1}$ and $PBR=6\text{Mbs}^{-1}$.*

Describing PBR and ABR does not uniquely define the behaviour of this model and every point on the surface supplies different parameter values for the model. The first and second moments (SCR and σ^2) of the distribution of the traffic were estimated for points across the surface, and the PCR was noted.

The mean of the model under the different input parameters varied by less than 9% which implies that the model has a reasonable degree of stability. The length of the simulations and the introduction of errors due to truncation probably contributed to this 9%. However the variance of the traffic changed by 5000% over the same range of input parameters suggesting that the model is able to produce a wide variety of traffic.

Figure 2.15 plots the variance for the simulations recorded in Table 2.2. Traffic possessing the largest variance is produced when $\vartheta(T)$ is small and $\vartheta(A)$ is large. As $\vartheta(T)$ is increased the effect of $\vartheta(A)$ on the variance is reduced.

Figures 2.16 and 2.17 were produced using points *a* and *b* from Table 2.2. Point *a* has

Inputs			Outputs		
$\vartheta(T)$	$\vartheta(A)$	ρ_{ea}	SCR	σ^2	PCR
5.0e-5	2.5e-4	0.919	2986	1552	4800
5.0e-5	4.0e-4	0.573	3000	3315	5100
5.0e-5	6.0e-4	0.383	2993	5495	6000
5.0e-5	8.0e-4	0.289	2981	7587	6800
5.0e-5	1.0e-3	0.229	3001	9784	7000
5.0e-5	2.0e-3	0.115	3016	20658	9200
5.0e-5	4.0e-3	0.055	3114	42016	9900
5.0e-5	6.0e-3	0.038	3039	57721	10600
1.5e-4	1.5e-4	0.870	2978	1081	4400
1.5e-4 ^a	2.0e-4	0.650	2984	1405	4400
1.5e-4	4.0e-4	0.320	3001	2645	5000
1.5e-4	6.0e-4	0.210	3022	3926	5600
1.5e-4	8.0e-4	0.160	3007	5099	5500
1.5e-4	1.0e-3	0.130	2995	6260	5700
1.5e-4 ^b	2.0e-3	0.060	3094	11763	6300
2.5e-4	4.0e-5	0.750	2980	1532	4400
2.5e-4	1.0e-4	0.290	2985	1651	4500
2.5e-4	2.0e-4	0.150	2980	1806	4600
2.5e-4	4.0e-4	0.074	2986	2069	4800
2.5e-4	6.0e-4	0.051	2976	2345	4700
2.5e-4	8.0e-4	0.037	2987	2597	4700
2.5e-4	1.0e-3	0.031	2975	2709	4800
2.5e-4	2.0e-3	0.002	3232	2296	5200

Table 2.2: Results from model simulations.

a low variance while point b has a high variance. However it is worth noting that both sets of traffic have the same coefficient of burstiness which is defined as

$$\frac{PCR}{SCR} \tag{2.24}$$

Due to the nature of the model the probability of traffic generated at point a transmitting at the PCR is very much smaller than for point b .

2.4.4 Determining the higher moments of the model

Next we determine the first and second moments of our model in terms of the model parameters. We then wish to determine whether these values are valid using Monte Carlo simulation of the model.

Consider an ATM service that produces traffic in accordance with an unknown distri-

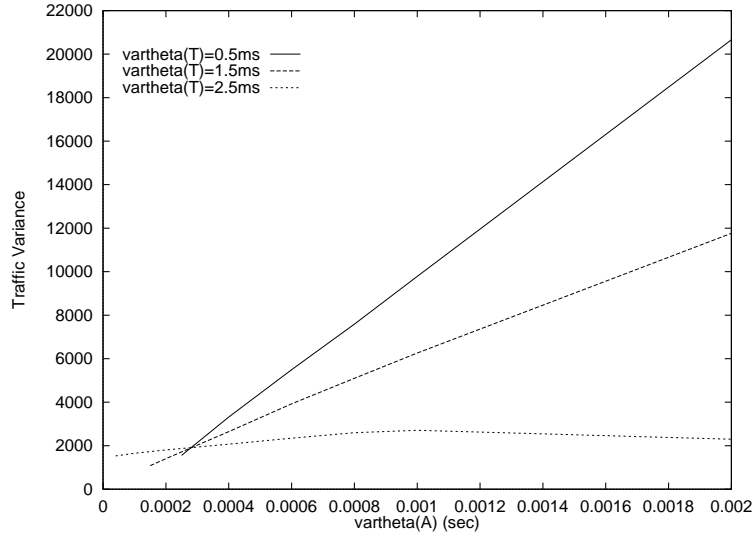


Figure 2.15: *Variance of the resultant traffic.*

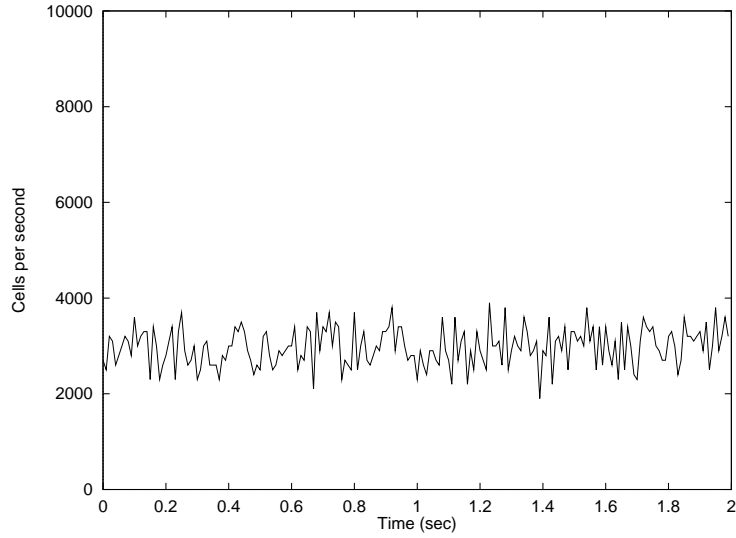


Figure 2.16: *Time series graph for point a (low variance).*

bution. We can represent the production of cells and inter-arrival times as in Figure 2.18.

The first moment of the distribution of inter-arrival times, $M_T^{(1)}(0)$, can be calculated using

$$M_T^{(1)}(0) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N T_i. \quad (2.25)$$

We know that any T_i is modelled by either an A_l loop or a T_l loop. Let l equal the number of A_l loops in N measurements of T_i . If we define $\tau_{A_l}^{(i)}$ ($\tau_{T_l}^{(i)}$) to be the time association with the i^{th} inter-arrival time generated by an A_l (T_l) loop then it is possible

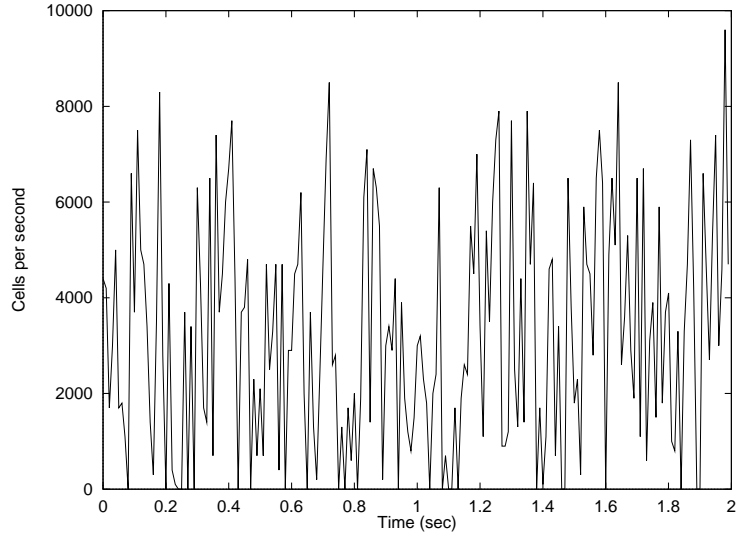


Figure 2.17: *Time series graph for point b (high variance).*

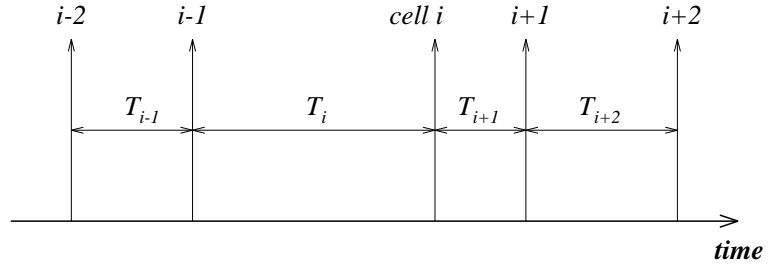


Figure 2.18: *Cell inter-arrival diagram.*

to rewrite (2.25) so that the A_l and T_l loops are separated,

$$M_T^{(1)}(0) = \lim_{N \rightarrow \infty} \frac{1}{N} \left(\sum_{i=1}^l \kappa(A_l) \tau_{A_l}^{(i)} + \sum_{j=l+1}^N \kappa(T_l) \tau_{T_l}^{(j)} \right). \quad (2.26)$$

However as $N \rightarrow \infty$,

$$\sum_{i=1}^l \kappa(A_l) \tau_{A_l}^{(i)} \rightarrow l \vartheta(A_l), \quad (2.27)$$

$$\frac{l}{N} \rightarrow \rho_{ea} \big|_{\rho_{eo}=0}. \quad (2.28)$$

This is because the proportion of A_l loops to the total number of loops is dependent on the probability of entering an A_l loop after transmitting a cell. Similarly for T_l loops,

as $N \rightarrow \infty$,

$$\sum_{i=l+1}^N \kappa(T_l) \tau_{T_l}^{(i)} \rightarrow (N-l)\vartheta(T_l), \quad (2.29)$$

$$\frac{N-l}{N} \rightarrow \rho_{et} \mid_{\rho_{eo}=0}. \quad (2.30)$$

For the remainder of this section we assume that ρ_{ea} and ρ_{ea} are the conditional probabilities i.e.,

$$\rho_{ea} + \rho_{et} = 1. \quad (2.31)$$

Therefore the first moment of the cell inter-arrival time distribution can be written as

$$M_T^{(1)}(0) = \rho_{ea}\vartheta(A_l) + (1 - \rho_{ea})\vartheta(T_l). \quad (2.32)$$

The sustained cell transmission rate, SCR, was derived in Section 2.4.2 using a different approach and was found to be

$$\text{SCR} = \frac{1}{\rho_{ea}\vartheta(A_l) + (1 - \rho_{ea})\vartheta(T_l)}, \quad (2.33)$$

$$= \frac{1}{M_T^{(1)}(0)}. \quad (2.34)$$

The next step in the examination of the model is to derive an equation for the second moment of the cell inter-arrival rate and then find an expression for the variance, σ^2 , using,

$$\sigma^2 = M_T^{(2)}(0) - (M_T^{(1)}(0))^2. \quad (2.35)$$

Consider Figure 2.18 and recall that,

$$M_T^{(2)}(0) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (T_i)^2. \quad (2.36)$$

However due to the nonlinearity of the squared term it is necessary to consider a different approach to that used for $M_T^{(1)}(0)$. Let $P(n_a)$ be defined as the probability of n_a occurrences of state A in the generation of one cell inter-arrival time and let $P(n_t)$ be defined in a similar manner for state T. To continue the derivation of $M_T^{(2)}(0)$ it is convenient to consider a uniform example and then develop a more general result from this.

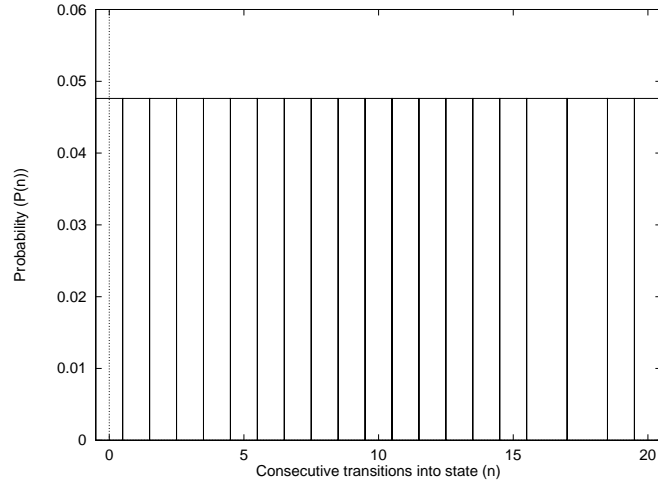


Figure 2.19: A uniform pdf of state occurrences with mean = 10

The plot for $P(n)$ is given in Figure 2.19 and from this it is possible to separate (2.36) into all the possible combinations of $\kappa(A)$ and $\kappa(T)$ that could result in T_i ,

$$\begin{aligned}
 M_T^{(2)}(0) &= \frac{l}{N} \sum_{n_a=0}^{20} \sum_{n_t=0}^{20} \left[(n_a \tau_{aa} + n_t \tau_{tt} + \frac{1}{\text{PCR}})^2 \frac{1}{20} \frac{1}{20} \right] \\
 &+ \frac{N-l}{N} \sum_{n_t=0}^{20} \left[(n_t \tau_{tt} + \frac{1}{\text{PCR}})^2 \frac{1}{20} \right]. \tag{2.37}
 \end{aligned}$$

For a more general case define N_a as the maximum number of recurrent transitions for state A and N_t as the maximum number of recurrent transitions for state T . Then we can rewrite (2.37) in a more general form,

$$\begin{aligned}
 M_T^{(2)}(0) &= \frac{l}{N} \sum_{n_a=0}^{N_a} \sum_{n_t=0}^{N_t} \left[(n_a \tau_{aa} + n_t \tau_{tt} + \frac{1}{P_{ctr}})^2 P(n_a) P(n_t) \right] \\
 &+ \frac{N-l}{N} \sum_{n_t=0}^{N_t} \left[(n_t \tau_{tt} + \frac{1}{P_{ctr}})^2 P(n_t) \right]. \tag{2.38}
 \end{aligned}$$

Equation (2.38) is referred to as the $M_T^{(2)}(0)$ estimator. It is obvious that as N_a and N_t increase the amount of computation required to solve (2.38) increases.

The first and second moments of the distribution function have been derived in terms of the parameters of the model and these results can be generalised into an expression

for the p th moment of the inter-arrival time,

$$\begin{aligned}
M_T^{(p)}(0) &= \frac{l}{N} \sum_{n_a=0}^{N_a} \sum_{n_t=0}^{N_t} \left[(n_a \tau_{aa} + n_t \tau_{tt} + \frac{1}{P_{ctr}})^p P(n_a) P(n_t) \right] \\
&+ \frac{N-l}{N} \sum_{n_t=0}^{N_t} \left[(n_t \tau_{tt} + \frac{1}{P_{ctr}})^p P(n_t) \right].
\end{aligned} \tag{2.39}$$

2.4.5 Impact of long tail distributions

ATM traffic models often employ long tailed distributions to simulate their bursty nature. This has implications for the application of large deviation theory (see [32]) which is concerned with the occurrence of rare events and is useful for cell loss ratio and effective bandwidth estimation in ATM networks (e.g., [24], [25], [14]). In the structured Markovian model long tail distributions can be introduced using the state recurrence pdfs we introduced above.

Consider the case when $P(n_a) = P(n_t) = P(n)$ where $P(n)$ is a Poisson distribution. As a result of the distribution's nature, $P(n)$, never actually reaches 0, even for large n . This implies that the $M_T^{(2)}(0)$ estimator is unable to calculate a value for the second moment over the entire range of $n : P(n) > 0$. There are two possible solutions to this problem though both compromise the accuracy of the estimator to some degree.

The first solution to the problem involves calculating the contribution from each n until $\sum_{i=0}^n P(i) = Z \{Z : 0 \leq Z < 1\}$ and ignores the contribution of larger n . The second solution attempts to approximate the effect of larger n by assigning values to particular large n in such a way as to maintain $M_T^{(1)}(0)$. The second method does not ignore the rare events due to n with small probabilities and this may prove useful for effective bandwidth modelling [25].

Consider a variable bit-rate B-ISDN service with a sustained cell transmission rate, SCR, and a peak cell transmission rate PCR. From Section 2.4.2 we know that a surface solution for this service is formed in the parameters ρ_{ea} , $\vartheta(A)$ and $\vartheta(T)$.

In the first simulation $P(n_a)$ and $P(n_t)$ were assumed to be equal and uniform with a mean of 10 (Figure 2.20). The comparison between the Monte Carlo variance and the estimator variance is given in Figure 2.21.

The next series of simulations assumed $P(n_a)$ and $P(n_t)$ were Poisson distributions

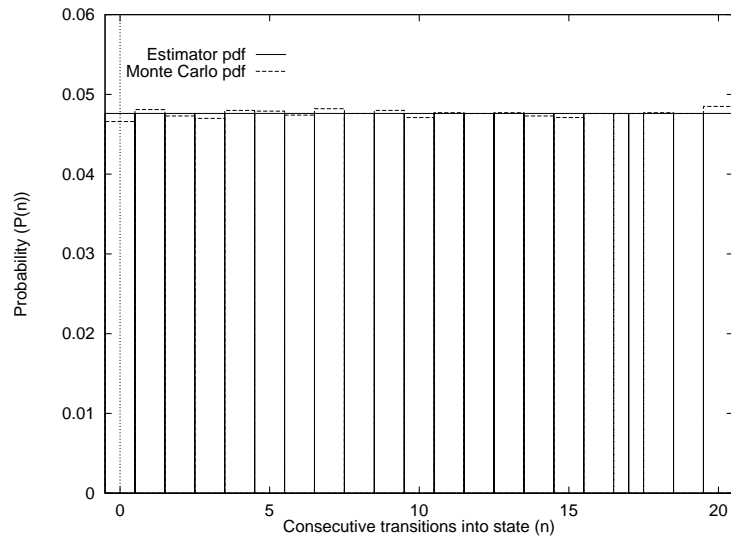


Figure 2.20: *State recurrence pdfs for estimator and MC simulation for the uniform case.*

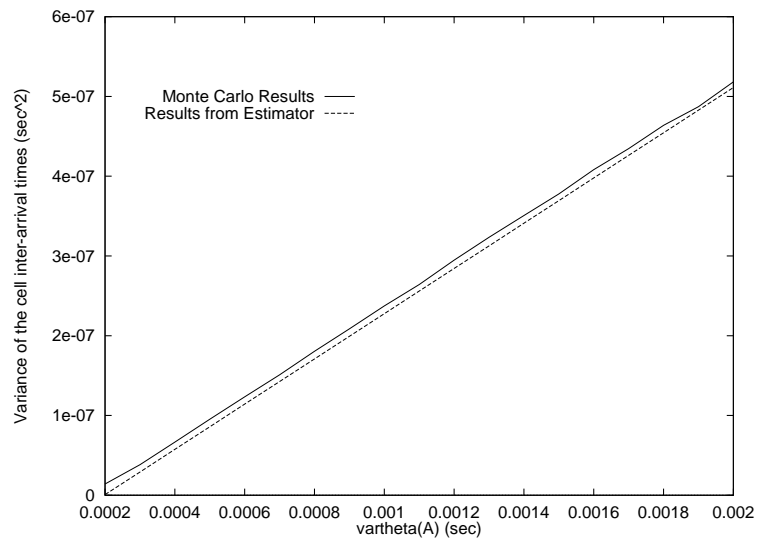


Figure 2.21: *Variance of the cell inter-arrival rate for the uniform case.*

(Figure 2.22). Three cases were considered for the estimator and the results are presented in Figure 2.23.

In the first case Z was set to 0.99 producing an accurate estimation of the variance. However such a large value for Z may be undesirable when computational resources are limited. In case two $Z = 0.9$ and the approximation was much poorer, however by using the poisson approximation (case three) a much better estimation was obtained (Z^* in Figure 2.23).

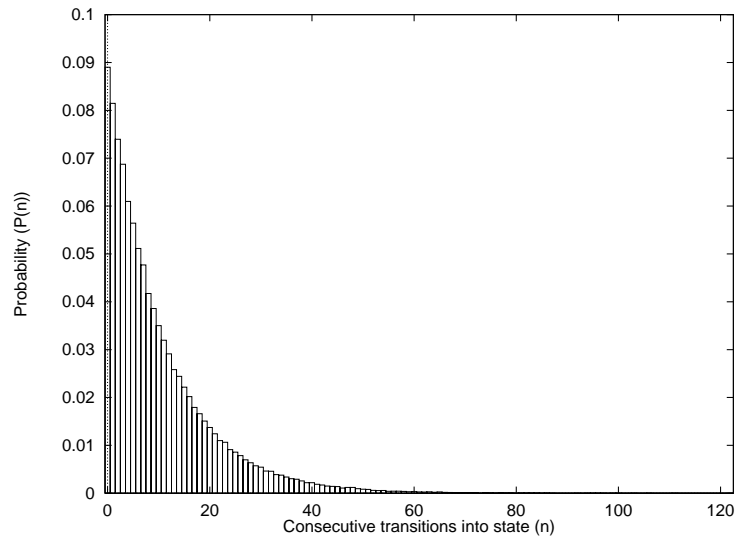


Figure 2.22: *State recurrence pdfs for MC simulation for the poisson case.*

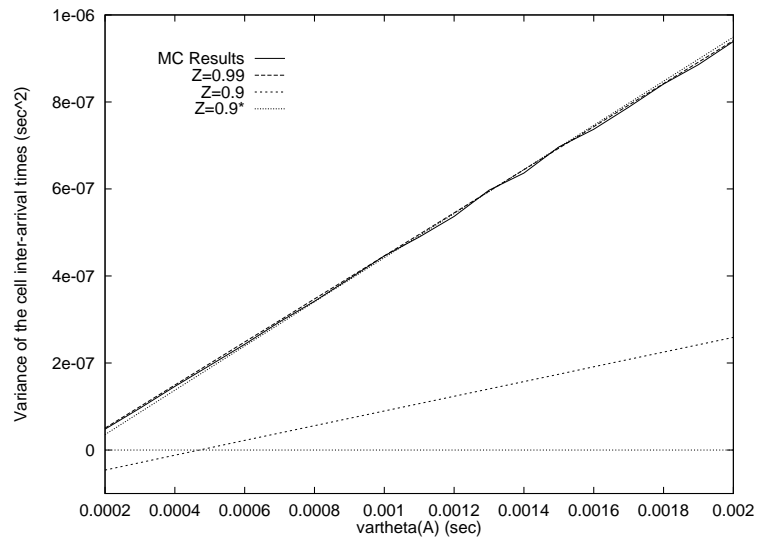


Figure 2.23: *Variance of the cell inter-arrival rate.*

2.5 Discussion

As it stands the structured Markovian model is able to capture the behaviour of a wide range of traffic types. In addition, the parameters of the model can be directly related to traffic parameters. However at present this model has no correlation structure and this is unrealistic. Each inter-cell arrival time is calculated with no regard to the previous behaviour of the model. It may be possible to introduce correlation structure by introducing some form of memory into the state transition probability matrix.

Using this model for parametric CAC is not an unrealistic proposition but before this can be done the effective bandwidth of the model is required.

The investigation into the model concluded at this point because of the arguments concerning the correlation structure of real teletraffic. In [33] Leland and other claimed that some forms of teletraffic exhibit behaviour that traditional models, including Markovian ones, are unable to capture. These results have such major repercussions for teletraffic modelling and CAC that they became the focus of the remainder of the work presented in this thesis.

2.6 Conclusions

In this chapter we have presented some background material on the asynchronous transfer mode (ATM) and some results on our investigations into developing models for parametric CAC schemes for ATM. ATM has the potential to cater for the diverse, multimedia based services that companies are hoping to introduce over the next 5 to 10 years. For example BT are experimenting with Video On Demand (VOD) and the Center for Communications Interface Research (CCIR) at Edinburgh University are designing systems for home shopping that incorporate voice, video, animation and data. Such a range of heterogenous source types, each with their own demands on QoS, posed a problem for all pre-ATM networks. We have seen how ATM can cater with these services using different service classes.

The main problem this thesis addresses is concerned with the VBR service class which is expected to route and manage bursty RT and NRT services. We want to investigate how the network can make use of measures of the traffic to estimate how much capacity that traffic will require to maintain the desired QoS. Specifically, effective bandwidths have been identified as a promising paradigm for creating both parametric and non-parametric CAC schemes.

Towards the end of this chapter we have presented some novel work by introducing a structured Markovian model. We showed how this model's parameters can be matched to properties of an ATM service. We have derived an expression for estimating the p th moment of the cells inter-arrival time distribution function and we have presented results proving that the expression performs well for at least $p = 1$ and 2. Unfortunately the model has no correlation structure which does limit its potential for CAC. In addition no work into deriving a closed-form expression for the effective bandwidth of this model has yet been carried out. This remains an area for further investigation.

Self-similarity teletraffic analysis and modelling

3.1 Introduction

Beniot Mandelbrot once posed the following question to a group of mathematicians. He asked them “How long is the coastline of Britain?”. It is reported that his colleagues treated this question with contempt and told him to look up the answer in an encyclopedia. But Mandelbrot declared that the answer in any encyclopedia would be wrong! He had observed that no matter at what scale he observed the coastline of Britain the outline was always a rugged one. This argument held for scales of tens of miles down to fractions of an inch. For example the rugged features of the north Scottish coast could be repeated (on a smaller scale) in a rockpool in Dorset or (on an even smaller scale) on a pebble in Wales. If one extrapolated this similarity to the smallest possible scale of measurement then surely the coastline must be huge, possibly even infinite.

The similarity Mandelbrot observed is a common and well understood occurrence in Geology termed scale invariance. It is for this reason geological photographs often have objects such as rulers, coins or people included in them, i.e. no indication of scale is inherent in the subject matter.

Although scale invariance was appreciated as long ago as the 1600s, the mathematics to explain it were not developed until the 1960s when Mandelbrot and associates developed the theory of fractal geometry [34]-[39].

In the 1950s Hurst was studying river level data recorded from the Nile in an attempt to solve the regularisation problem of river flows [40], [41]. He knew that the Nile exhibited trends in levels, i.e. there were often several years of low river levels followed by years of annual floods (the Joseph effect [42]). However overall the river data appeared to

have no underlying trends, merely the superposition of many short-term ones at a very wide range of frequencies.

In order to solve the regularisation problem he developed a reservoir system that would buffer the water. By storing water when the levels were high and releasing it when the levels were low a result could be obtained. While determining the optimum size for this reservoir he discovered the river level data was self-similar in nature. Self-similarity is to the time domain what scale invariance is to the spatial domain i.e. statistical similarity over a wide range of scales.

Self-similarity has since been found in econometrics [43], [44], hydrology [40], meteorology [45] and telecommunications [46]. In 1994 the seminal study of self-similarity in network traffic was published, [33]. The importance of this, with regards to teletraffic analysis and modelling, is discussed in Section 3.3. For now we begin by defining the self-similar process and introducing some of its properties.

3.2 Definitions and properties of self-similar time-series

In Section 3.1 we introduced the concept of a self-similar time-series in a qualitative fashion. In this section the term is defined and the salient properties of a self-similar time series are introduced. Two definitions exist for long memory processes (processes which possess long-range dependence (LRD)), the first relates to the correlation structure of the process and the second to the power spectral density (PSD). The two definitions are equivalent [45].

Definition 3.1: *Let X_t be a stationary process with a correlation function, $\rho(k)$, for which there exists a number $\alpha \in (0,1)$ and a constant $c_\rho > 0$ such that*

$$\lim_{k \rightarrow \infty} \frac{\rho(k)}{[c_\rho k^{-\alpha}]} = 1$$

Then X_t is a stationary process with long-memory (long range dependence).

Definition 3.2: Let X_t be a stationary process with a PSD, $f(\omega)$, for which there exists a number $\beta \in (0, 1)$ and a constant $c_f > 0$ such that

$$\lim_{\omega \rightarrow \infty} \frac{f(\omega)}{[c_f |\omega|^{-\beta}]} = 1$$

Then X_t is a stationary process with long-memory (long range dependence).

These definitions imply that all long memory processes possess an ACF that decays in proportion to $k^{-\alpha}$, i.e. in an algebraic fashion. This is the feature that has a direct impact on the queue performance of such traffic, a topic we shall discuss further in Section 3.3.

Self-similar processes are inherently related to LRD processes in that all self-similar processes possess LRD. The definition of the self-similar process is given below.

Definition 3.3: Let X_t be a stochastic process with continuous time parameter, t . X_t is called self-similar with self-similarity parameter, H , if for any positive stretching factor c , the rescaled process with time scale ct , $c^{-H}Y_{ct}$, is equal in distribution to the original process Y_t .

We can demonstrate the scaling property of a self-similar process in a simple experiment. Consider a 100000 point sample trace of a self-similar process with $H = 0.8$. A 1000 point extract from this trace is shown in Figure 3.1(a). If a stretching factor of 10 is used then a 10000 point process is obtained by grouping the original data into blocks of 10 and scaling the total of each block by $10^{-0.8} \approx 0.1584$. A 1000 point extract of the resultant process is given in Figure 3.1(b). A similar procedure is carried out on this trace to yield the 1000 point trace in Figure 3.1(c). These three traces appear statistically similar.

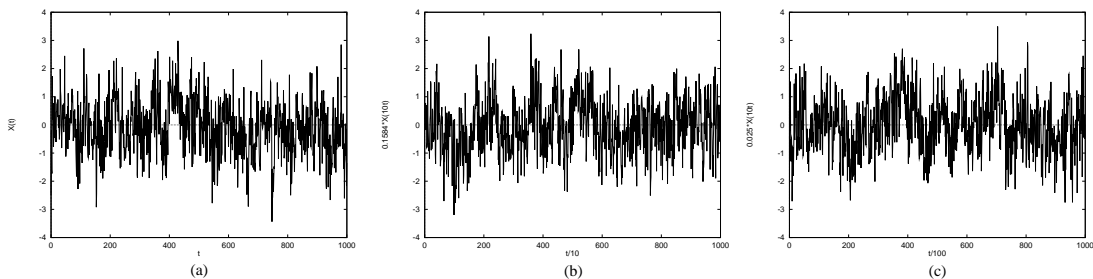


Figure 3.1: (a) The original self-similar process. (b) The process aggregated by a factor of 10 and scaled accordingly. (c) The process aggregated by a factor of 100 and scaled accordingly.

In addition to statistical similarity after scaling, these processes possess more quantitative properties.

1. They possess a hyperbolically decaying auto-correlation function of the form,

$$\rho(k) \simeq k^{(2H-2)} L(k) \text{ as } k \rightarrow \infty, \quad (3.1)$$

where $L(t)$ is a slowly varying function at infinity (i.e. $\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1$ for all $x > 0$). Therefore the autocorrelation function is unsummable, i.e.

$$\sum_k \rho(k) = \infty. \quad (3.2)$$

This infinite sum is the definition for *long range dependency* so all self-similar signals are long-range dependent. The repercussions of this for time-series can be extreme because the cumulative effect over a large range of lags can be significantly different to that of a short range dependent (SRD) process (e.g. Poisson, Markov or auto-regressive (AR) process). Since teletraffic analysis has relied heavily on SRD models in the past the implications of LRD in teletraffic are potentially severe. This is because LRD can produce long bursts of larger than average traffic levels which overflow buffers and cause loss and/or delay.

2. The variance of the sample mean decays more slowly than the number of points in the sample. For example if

$$X^{(m)} = \frac{1}{m} \sum_{i=1}^m X_i \quad (3.3)$$

then

$$\text{VAR}(X^{(m)}) \propto m^{(2H-2)}. \quad (3.4)$$

This implies that sample statistics such as mean and variance will be slow to converge, especially when $H \rightarrow 1$. This has repercussions for all measures of self-similar processes and will have to be considered when estimating statistics in later sections of this thesis.

3. The power spectrum obeys a $1/f$ type law close to the origin,

$$f(\omega) \sim c_f |\omega|^{1-2H} \text{ (as } |\omega| \rightarrow 0). \quad (3.5)$$

This is why self-similar and long range dependent processes are sometimes termed *1/f-noise* [47].

(3.1), (3.4) and (3.5) are linked by the parameter H which is called the Hurst exponent. The Hurst exponent of a self-similar time-series can lie between 0.5 and 1. The closer H is to 1 the more self-similar the time-series; this manifests itself as a slower decay of the autocorrelation function (as implied by (3.1)).

3.3 Implications of self-similarity in B-ISDN networks

Evidence to suggest that self-similarity does occur in teletraffic scenarios exists for Ethernet LAN and WAN [33], [48], Common Channel Signalling Network (CCSN/SS7) [18], Integrated Services Digital Network (ISDN) [49] and VBR video [50]. Although counter claims to these results have been made, [17], in this section we wish to consider the implications of self-similarity on network performance.

If self-similarity does exist in broadband teletraffic then [20], [33], [51]-[55] show that it will have a major impact on resource allocation and network dimensioning. This is because the results in these papers prove that LRD has a major impact on the queuing behaviour of teletraffic. For example in [20] Duffield and O'Connell show that for a single server queue with a large buffer, a process with LRD is more likely to overflow that buffer than a similar SRD process. More formally, for a stable queue with a deterministic service rate and a stationary SRD arrival process the probability of the queue length, Q , exceeding a given buffer size, b , is given by the buffer occupancy relationship,

$$\mathbf{P}(Q > b) \approx \exp(-\delta b), \quad (3.6)$$

for some positive constant δ . Duffield and O'Connell showed that this relationship did not hold for processes with LRD. When a self-similar process is fed through a stable queue the buffer occupancy relationship is given by

$$\mathbf{P}(Q > b) \approx \exp(-kb^c). \quad (3.7)$$

Where $k > 0$ and $0 \leq c \leq 1$ are constants that depend on the parameters of the self-similar process. This is a Weibull distribution which decays at a much slower rate than

the exponential in (3.6). In addition to this mathematical result other papers have used simulations to form similar conclusions (see [33], [51]-[55]).

However [56] and [57] also report on investigations into queue behaviour when a server is fed with LRD traffic and concluded that, for realistic buffer sizes and QoS figures, the LRD had very little impact. These results seem to contradict the previous results. One explanation is that the smaller buffer sizes used in the latter two papers reduce the memory of the system hence negating the LRD effect. Since the queue cannot “see” further back in time than the last “buffer full” or “buffer empty” event, the long range correlation structure of the input traffic is irrelevant. Although these results are not conclusive they do suggest that good queuing system design may negate the effect of LRD.

In addition to its impact on the buffers in a queuing environment, LRD would also have significant repercussions for the deployment of call admission control (CAC) schemes, especially effective bandwidths. This is because effective bandwidths rely on the assumption that the source is stationary and mixing (i.e. its autocorrelations sum to some finite value). We know from property 1 in Section 3.2 that this is not true for self-similar processes. In theory the effective bandwidth of a source that is self-similar will always be positive infinity. In practice this result will not occur because we would have to measure an infinite amount of data to record the result. This leads to a dilemma. Are some teletraffic types inherently self-similar and if so how does that affect the practical implementation of effective bandwidths for CAC schemes? As mentioned above some studies have concluded that, in practice, LRD has little effect on buffer queues; perhaps the same result will be found for effective bandwidths. In the mean time self-similar analysis and modelling remains an interesting and controversial area of research within teletraffic engineering.

3.4 Measuring self-similarity in time-series

As we saw in Section 3.2 a good measure of self-similarity is the Hurst exponent, H . In this section we introduce techniques whereby H can be estimated for a time-series. Most of these techniques exploit the properties of the time-series to develop heuristic approaches to estimating H . However the final method (the Whittle estimator) is an approximate MLE technique and as such can provide us with confidence limits on the estimate of H .

3.4.1 The rescaled adjusted range technique

The rescaled adjusted range statistic (R/S -statistic) method was employed by Hurst in his solving of the Nile river regulation problem [40], [41], [45]. If our time-series is approximated by an appropriately sampled discrete process, $\{X_i\}_{i=1,2,3,\dots,n}$ then let

$$Y_j = \sum_{i=1}^j X_i \quad (3.8)$$

for $j = 1, 2, \dots, n$ and

$$\bar{X}_j = \frac{1}{j} \sum_{i=1}^j X_i. \quad (3.9)$$

The next step is to find

$$R(j) = \max_{1 \leq i \leq j} (0, Y_i - i\bar{X}_j) - \min_{1 \leq i \leq j} (0, Y_i - i\bar{X}_j). \quad (3.10)$$

If we define

$$S(j) = \frac{1}{j-1} \sqrt{\sum_{i=1}^j (X_i - \bar{X}_j)^2} \quad (3.11)$$

and plot

$$\log E\left(\frac{R(j)}{S(j)}\right) \text{ vs } \log j \quad (3.12)$$

over $j = 1, \dots, n$ then we can estimate H as the slope of this plot.

Problems with the R/S -statistic include choosing the optimal range of j over which to perform the least-squares fit for \hat{H} . For small j the long-range behaviour is unapparent but for large j the number of points over which the expectation must be calculated is small.

3.4.2 The variance plot technique

One of the properties of a self-similar process is that the variance of the estimate of the mean decays to zero slower than for SRD processes (equation (3.4) in Section 3.2). By determining the rate of this decay it is possible to estimate a value for H . A procedure

by which this can be done is the variance plot technique [45].

Assume the data $\{X\}$ is of length N . Then divide the data into n_k blocks, each of length k and find the mean of each block, i.e

$$\overline{X}_i^{(k)} = \frac{1}{k} \sum_{j=i*k}^{(i+1)*k} X_j \quad i = 0, 1, \dots, n_k. \quad (3.13)$$

Also define

$$\overline{X}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} \overline{X}_i^{(k)} \quad (3.14)$$

and

$$s^2(k) = \frac{1}{n_k - 1} \sum_{k=1}^{n_k} (\overline{X}_i(k) - \overline{X}(k))^2. \quad (3.15)$$

It is possible to determine \hat{H} by noting

$$\log s^2(k) = (2H - 2) \log k + c, \quad (3.16)$$

where c is a constant. A SRD process will have a variance slope plot of -1 whereas a LRD process will have a slope between 0 and -1.

3.4.3 The periodogram plot

In Section 3.2 we saw that for a self-similar process the PSD possesses a pole at $\omega = 0$, i.e.

$$f(\omega) \sim c_f |\omega|^{1-2H} \quad (\text{as } |\omega| \rightarrow 0) \quad (3.17)$$

By plotting $\log f(\omega)$ vs $\log |\omega|$ it is possible to estimate H . However this linear relationship only holds for $|\omega|$ close to 0 so determining the range over which to perform the linear regression is critical.

3.4.4 The Whittle approximate MLE technique

The MLE technique for estimating H is described in Chapter 5 of [45]. It is formulated by maximising the log likelihood function with respect to the parameter vector. However the procedure is very computationally intensive, especially for long datasets and it can be unstable when $H \approx 1$. Instead Whittle's approximation to the MLE method is employed [58]. One disadvantage of this method is that it is only optimal for Gaussian self-similar processes, which is often not a valid assumption for real teletraffic data.

3.4.5 Comparing the estimation techniques

In order to evaluate the performance of the estimation techniques we applied them to a dataset with a known Hurst exponent. We generated a 16384 point long sample trace with $H = 0.8$ using a fractional Gaussian noise (fGn) algorithm. We then tested this trace with three of the techniques discussed above; the R/S -statistic technique, the variance plot technique and the Whittle technique. The results were recorded for $N = 1024$ and $N = 16384$.

N	$H_{R/S}$	H_{var}	H_{Whittle}	95% C.L. for H_{Whittle}
1024	0.8375	0.7228	0.7781	0.7373 - 0.8188
16384	0.8083	0.7670	0.8011	0.7909 - 0.8113

Table 3.1: Results of the comparative study of the Hurst exponent estimators for a Gaussian self-similar process.

The two heuristic techniques estimate H with accuracy to within 10% for $N = 1024$ and 5% for $N = 16384$. The Whittle estimate included the correct value of H within its 95% confidence limits in both cases. So the estimation techniques all work well for a Gaussian self-similar process.

Next we want to consider how well the techniques perform when applied to real data. The data in question is a sample of an MPEG ¹ encoding of the motion picture Star Wars. This is a typical representation of a VBR video service on an ATM network.

These results are less sanguine than those for the Gaussian self-similar process. Firstly we do not know what the true value of H is or even if the dataset is truly self-similar. The results of the heuristic estimation techniques suggest that the data is self-similar

¹See Appendix B for a description of the MPEG video coding standard

N	$H_{R/S}$	H_{var}	H_{Whittle}	95% C.L. for H_{Whittle}
1024	0.8853	0.7119	0.9897	0.9480 - 1.000
16384	0.8377	0.8147	0.9954	0.9850 - 1.000

Table 3.2: Results of the comparative study of the Hurst exponent estimators for MPEG data.

with H approximately equal to 0.82. However both these techniques assume the data is stationary and this may not be the case. If not then the value of H is meaningless with regard to its defined sense [17]. The question of stationarity versus self-similarity is investigated further in Chapter 4.

The estimates produced by the Whittle technique are very high and do not agree with the estimates obtained by the heuristic approaches. This is not surprising since the data is highly non-Gaussian in nature.

3.5 Self-similar teletraffic models

The evidence that some forms of teletraffic possess LRD and the need to investigate how such LRD might affect networks mean self-similar teletraffic models are required. Since self-similarity had been investigated in other areas for some time the framework for such models is available. The best known self-similar process is fractional Brownian motion (fBm) [34]-[36] from which Norros developed a teletraffic model [53]. We discuss both fBm and the Norros model in the next section. Another form of self-similar model that we consider is the auto-regressive integrated moving average (ARIMA) process [45]. The third model that we consider is based on the marginal distribution mapping model [52]. In Section 3.5.4 we discuss some of the other self-similar teletraffic models.

3.5.1 Fractional Brownian motion and the Norros model

Fractional Brownian motion (fBm) is a Gaussian self-similar process with stationary increments. If we define $B_H[t]$ as a fBm process with Hurst exponent H and variance σ^2 then:

1. $B_H[0] = 0$ a.s.
2. $B_H[t]$ is Gaussian.

3. $B_H[t]$ has stationary increments.
4. $E[B_H[t] - B_H[s]] = 0$.
5. $\text{VAR}[B_H[t] - B_H[s]] = \sigma^2|t - s|^{2H}$
6. $E[B_H[t]B_H[s]] = \frac{\sigma^2}{2} \{|t|^{2H} + |s|^{2H} - |t - s|^{2H}\}$.

This process was first introduced by Kolmogorov [59] but it was Mandelbrot who suggested it was anything more than an interesting mathematical phenomenon [35]. The stationary incremental process of fBm is termed fractional Gaussian noise (fGn). If we define the discrete fGn process with Hurst exponent H as $X_H[t]$ then

$$X_H[i] = B_H[i] - B_H[i - 1] \quad (3.18)$$

When modelling fBm it is more common to generate fGn and then construct fBm from it [52], [60].

Norros has used fBm to construct a Gaussian self-similar teletraffic model [53], [61]. The definition of the Norros model is as follows.

Definition 3.4: *The continuous time process $\hat{A}(t)$ $t \in (0, \infty)$ is defined as*

$$\hat{A}(t) = mt + \sqrt{am}B_H(t)$$

where $m > 0$ and $a > 0$ are constants and $B_H(t)$ is a continuous fractional Brownian motion process with Hurst exponent H .

$\hat{A}(t)$ is a cumulative arrival process i.e., $\hat{A}(t_1) \geq \hat{A}(t_2)$ if $t_1 > t_2$ for all $t_1 > 0, t_2 \geq 0$. m is the mean arrival rate of the process and a is a scaling term.

The Norros model can also be written in the discrete form,

$$\hat{A}[i] = mi + \sqrt{am}B_H[i]. \quad (3.19)$$

Advantages of this model are its parsimonious nature (3 parameters) and the fact that the Gaussian assumption allows us to derive closed form expressions for lower bounds on buffer occupancy probabilities. Disadvantages include the dependence on assuming real teletraffic has a Gaussian marginal distribution and the fact that no direct estimation method for a exists.

3.5.2 The auto-regressive integrated moving average model

The ARMA process has been understood for many years and can be expressed in the following manner. If B is the backshift operator, $\phi(x)$ and $\psi(x)$ are polynomials of order p and q respectively and ϵ_t is some noise process then the ARMA(p, q) model is the stationary solution of

$$\phi(B)X_t = \psi(B)\epsilon_t. \quad (3.20)$$

The ARIMA(p, d, q) model is an adaptation of this and has an additional parameter d ,

$$\phi(B)(1 - B)^d X_t = \psi(B)\epsilon_t. \quad (3.21)$$

The fractional ARIMA model is simply an ARIMA model with $-0.5 < d \leq 0.5$. If $d \in (0, 0.5]$ then the ARIMA model possesses LRD and will produce self-similar traffic data.

The ARIMA process has the advantage that it can capture short range correlation structure (SRCS) and long range correlation structure (LRCS) simultaneously, which the Norros model fails to do.

3.5.3 The marginal distribution mapping model

The Norros model permits the modelling of teletraffic with a Gaussian marginal distribution. However in many cases the assumption of normality may not be a valid one and we may wish to experiment with other distributions. For example in [52] Garrett and Willinger consider the heavy tailed Pareto distribution. However in order to produce a self-similar process with such a distribution they were required to use a marginal distribution mapping model. This model has the advantage that it can be used to generate self-similar data with any arbitrary marginal distribution.

If we assume $F_d(\cdot)$ is the cumulative probability density function (CDF) of the desired output marginal distribution then we can obtain output data, $Y[i]$, by the mapping

$$Y[i] = F_d^{-1}(F_{N(\mu, \sigma^2)}(X_H[i])). \quad (3.22)$$

$X_H[t]$ is fGn with a suitably chosen Hurst exponent and $F_{N(\mu, \sigma^2)}$ is the CDF of $N(\mu, \sigma^2)$.

To demonstrate the marginal distribution model we used it to generate 1024 points of data matched to the marginal distribution of a JPEG encoding of the motion picture Star Wars. From [50] we know $H = 0.80$ approximately.

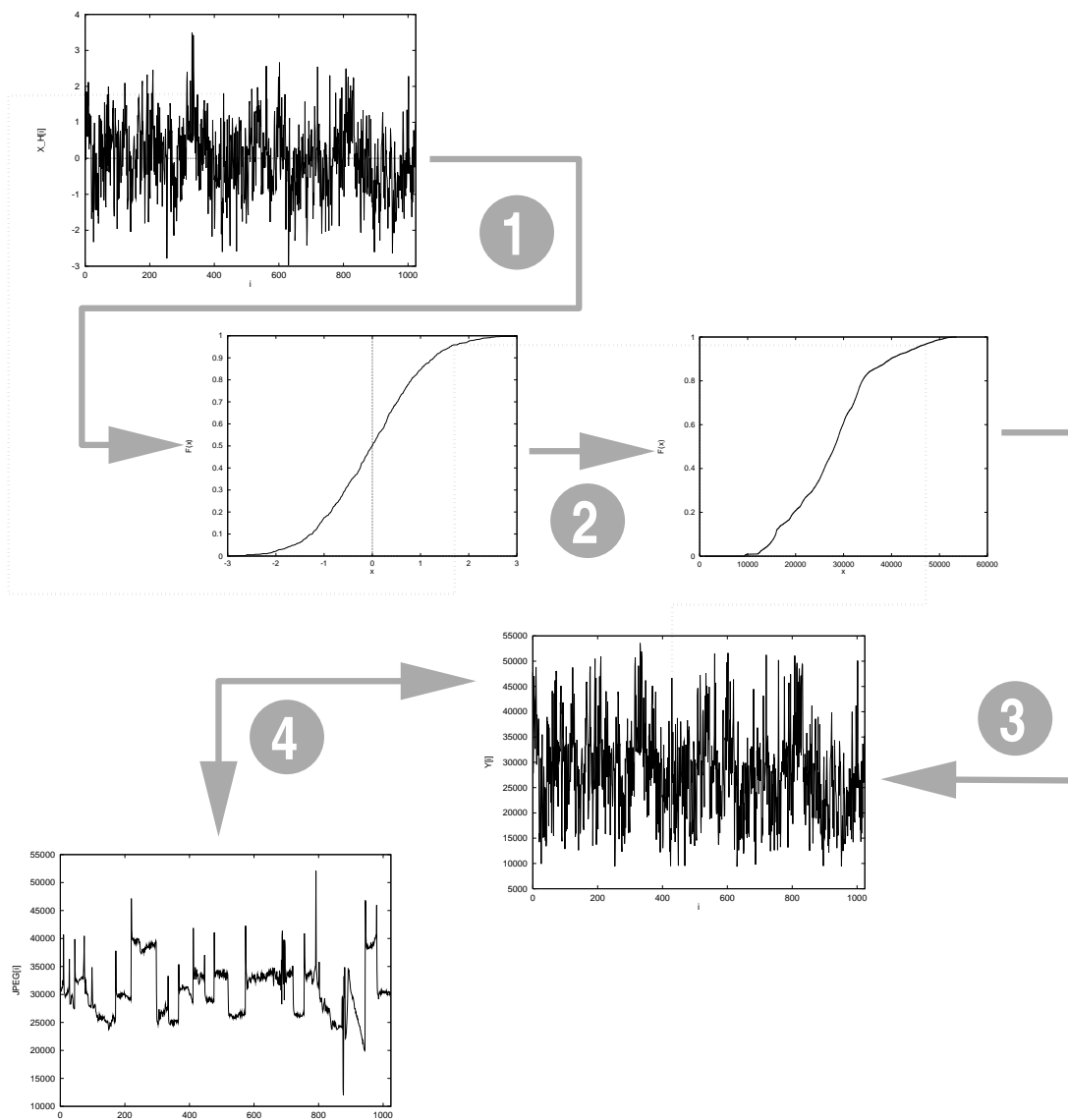


Figure 3.2: *The marginal distribution mapping model: An illustrative example.*

The modelling technique can be broken down into the four stages illustrated in Figure 3.2.

- **Step 1:** Obtain the CDF of the fGn process. This can be done either by empirical estimation or by estimating the mean (μ) and variance (σ^2) and using the CDF of $N(\mu, \sigma^2)$.
- **Step 2:** Obtain the CDF of the output process by empirical estimation from available data. Now we can map between the CDF of the input process and the

CDF of the output process.

- **Step 3:** Taking the inverse of the CDF function to obtain the output value.
- **Step 4:** Compare model output with available data to evaluate the performance.

The dashed line in Figure 3.2 represents the mapping process for a single point. This point occurs at $i = 420$ where $X_{0.8}[420] = 1.80198$. At this value of $X_{0.8}$ the CDF is approximately 0.95. By looking at the CDF of the desired output process we see this equates to $Y[420] = 46000$. By repeating this process for all i we can generate the output data.

N	$H_{R/S}$	H_{var}	H_{Whittle}	95% C.L. for H_{Whittle}
1024	0.6274	0.7721	0.7906	0.7498 - 0.8142

Table 3.3: The Hurst exponent estimates of the output of the marginal distribution mapping model.

The Hurst exponent estimates of the model output are recorded in Table 3.3. These estimates suggest that H has not been affected by the mapping process. It is also interesting to note that the Whittle estimate performs well in this case even though the data is no longer Gaussian but a non-linear mapping of a Gaussian process.

If we compare the marginal distribution model output with a sample of the JPEG trace we were attempting to model we notice a difference. This is due to the fact that the marginal distribution mapping model takes no account of the SRCS which is apparent in the JPEG data. In addition the range of the model output seems larger than the sample of real data. This may be due to the fact that the model is similar in distribution to the entire JPEG series and any local variations can not be mimicked by the model. Nonstationary data will pose a similar problem for this model because the resultant distribution will combine the distributions of the stationary segments of the data.

3.5.4 Other models

Willinger *et al.* developed a self-similar model that utilises the fact that aggregating SRD processes can result in a self-similar process [45]. They concentrate on ON-OFF models with holding times drawn from a heavy tailed distribution [55]. The argument behind this is that teletraffic is often formed by the aggregation (or multiplexing) of many heavy tailed SRD processes. In fact this has been argued to be the reason for LRD in teletraffic.

Another family of self-similar teletraffic are the self-similar point processes developed by Ryu [62]. These processes are developed from fractal ON-OFF processes which modulate the rate function of a Poisson process. In [56] Ryu and Elwalid combined the model with a SRD discrete auto-regressive (DAR) process to investigate the effect of a process with both LRCS and SRCS on queue behaviour.

3.6 Summary

Self-similar processes have led to a resurgence of interest in teletraffic analysis and modelling. The validity of traditional queuing results have been questioned and the emergence of high speed heterogenous networks has fuelled research into even the most fundamental of queuing problems. Therefore understanding self-similarity and how it can affect networks is of utmost importance.

In this chapter we have seen how self-similar processes behave and explained how this may impact on resources in a network environment. We have discussed how we may be able to estimate a measure of self-similarity, the Hurst exponent, using both heuristic and approximate MLE techniques. In addition, we have introduced the self-similar models that we will be considering in this thesis.

Is VBR video non-stationary or self-similar? Implications for traffic characterisation

4.1 Introduction

In Chapter 3 we introduced the concept of a self-similar time-series and showed how such a property has direct impact on a signal's autocorrelation function (ACF). We also discussed the implications for network CAC schemes if teletraffic data such as LAN packet data, VBR video data, DQDB MAN data and ISDN data are self-similar.

However there is a real possibility that the assumption that the data is stationary is violated by teletraffic data. If this is the case then the Hurst exponent has no meaning in relation to its defined sense and the data cannot be said to be self-similar. In [17] the arguments for non-stationary data were based on visual examination and it would therefore seem appropriate to extend this by placing some sort of quantitative figure on the extent to which a signal may be considered stationary ¹.

To do this we consider two VBR video encodings of the motion picture Star Wars. The first is a joint photographic experts group (JPEG) encoding and the second is a motion picture experts group (MPEG) encoding ². We then generate comparative datasets with a variety of correlation structures and apply two novel non-parametric tests to the real and the comparative data. The tests are the wide-sense-stationary quotient (WSS quotient) and the H_N statistic. The results of the tests are discussed in Sections 4.4.3 and 4.5.3 and we draw conclusions in Section 4.6.

¹The work was first presented in [63] at the European Simulation Multiconference 1996 in Budapest where it received a special acknowledgement.

²See Appendix C for an explanation of the JPEG and MPEG video CODEC schemes

4.2 The VBR video data

4.2.1 Introduction

By choosing two different coding schemes of the same motion picture we hope to be able to separate statistics introduced by the coding scheme from those inherent in the motion picture itself. Both of these datasets are available from the Bellcore FTP site at ftp.bellcore.com and both consist of 171000 integer values. These values are the number of bits per frame for the motion picture and at a playback rate of 24 frames per second represent almost two hours of video.

4.2.2 The JPEG encoded data

For illustrative purposes the bits/frame values for the first 1200 frames of the JPEG video data are given in Figure 4.1(a).

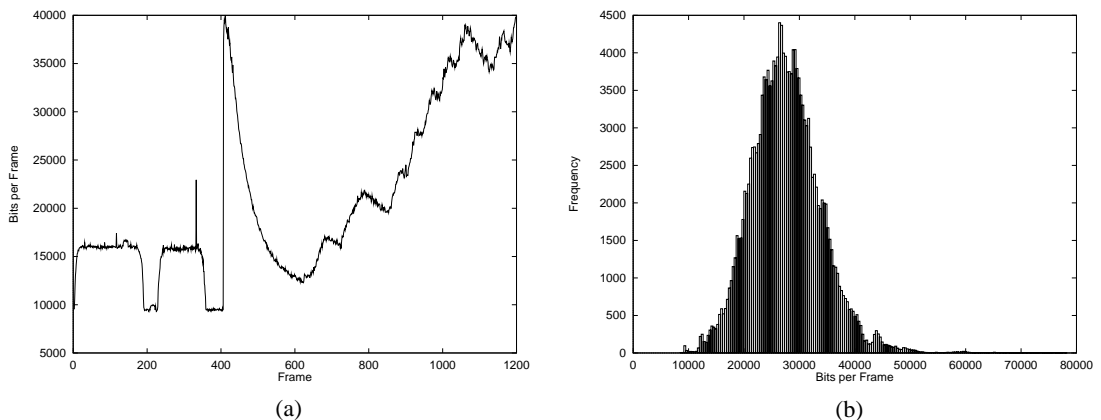


Figure 4.1: (a) A sample trace from the JPEG encoding of the motion picture *Star Wars* (b) the 200 bin histogram.

A 200 bin histogram, constructed from the data is given in Figure 4.1(b) and some of the statistical properties of the trace are given below.

$$\begin{aligned} N &= 171000 & \sigma &= 6254.2 & \nabla &= 8622 \\ \mu &= 27791.2 & \Delta &= 78459 \end{aligned}$$

The correlation function is given in Figure 4.2.

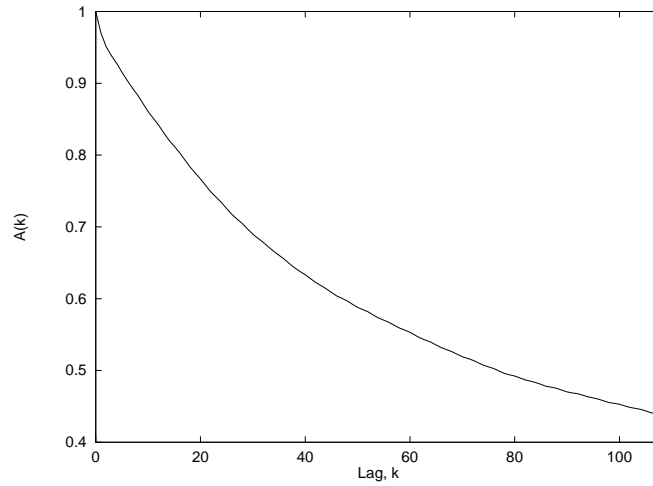


Figure 4.2: *The first 108 points of the correlation function estimated from the JPEG data.*

4.2.3 The MPEG encoded data

For illustrative purposes the bits/frame values for the first 1200 frames of the MPEG data are given in Figure 4.3(a).

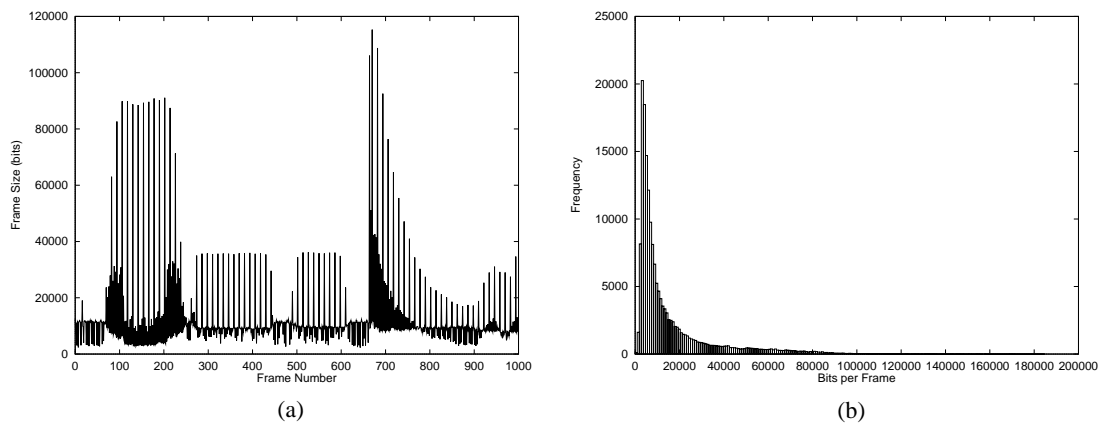


Figure 4.3: (a) *A sample trace from the MPEG encoding of the motion picture Star Wars* (b) *the 200 bin histogram.*

A 200 bin histogram, constructed from the data is given in Figure 4.3(b) and some statistics of the trace are given below.

$$\begin{aligned}
 N &= 171000 & \sigma &= 18164.95 & \nabla &= 476 \\
 \mu &= 15599.18 & \Delta &= 185267
 \end{aligned}$$

The MPEG encoding scheme introduces a strong periodicity that is not present in the JPEG data. This is immediately apparent if we examine the autocorrelation function of the MPEG data (Figure 4.4).

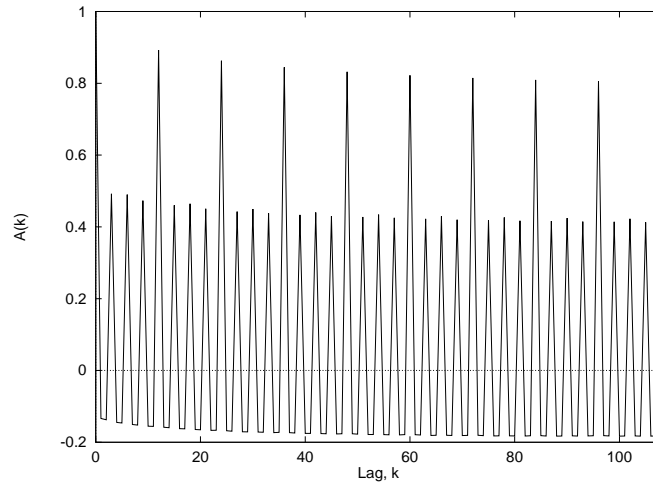


Figure 4.4: *The first 108 points of the correlation function estimated from the MPEG data.*

To remove this periodicity from the data we grouped the MPEG data into blocks of 12 frames. Each block is termed a group of pictures (GOP). The autocorrelation function of the GOP data is given in Figure 4.5.

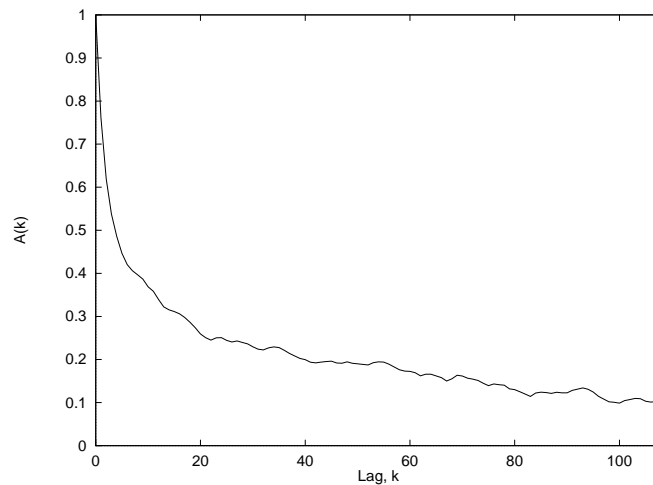


Figure 4.5: *The first 108 points of the correlation function estimated from the MPEG data after grouping.*

The salient statistics of the grouped MPEG data are given below.

$$\begin{aligned}
 N &= 14250 & \sigma &= 5514.3 & \nabla &= 6458 \\
 \mu &= 14959.4 & \Delta &= 78720
 \end{aligned}$$

4.3 The comparative data

The purpose of this chapter is to test the hypothesis *Is VBR video teletraffic stationary?*. In order to do this two tests were developed; these are introduced later in this chapter. We also require a range of model generated data to (i) confirm the performance of the tests and (ii) compare the performance of the real data which has assumed characteristics with the model data which has known characteristics.

We know that self-similarity in a signal is intrinsically linked to its autocorrelation function by equation (3.1). Therefore we concentrated on constructing models that generate data with a range of ACF structures.

4.3.1 The ON-OFF model

The ON-OFF model generates frames of either size p bits or 0 bits in accordance with the output of a Random Number Generator (RNG) and a threshold value, a , i.e.

$$X[t] = \begin{cases} p & \text{if RNG}[t] \geq a \\ 0 & \text{if RNG}[t] < a. \end{cases} \quad (4.1)$$

$$E(X) = ap \text{ and } \text{VAR}(X) = a(1 - a)p^2.$$

The model parameters for the JPEG data were $a = 0.952$ and $p = 29198$. The MPEG data was too bursty for the ON-OFF model to capture so we applied it to the GOP data and calculated $a = 0.880$ and $p = 16999$. The statistics of the output data are given below.

ON-OFF statistics for the JPEG data

$$\begin{aligned} N &= 171000 & \sigma &= 6269.3 & \nabla &= 0 \\ \mu &= 27783.3 & \Delta &= 29198 \end{aligned}$$

ON-OFF statistics for the GOP MPEG data

$$\begin{aligned} N &= 14250 & \sigma &= 5582.3 & \nabla &= 0 \\ \mu &= 14909.0 & \Delta &= 16999.0 \end{aligned}$$

This model has the short-comings that it can only assume two values of frame size (and

one of these is the unrealistic value of zero). However it is useful in this study as we know that there is no correlation between frames.

4.3.2 Auto-regressive models

If the video data is LRD it will possess both short and long range correlation structure. Therefore it was necessary to develop a model with the same characteristic. The model we used is an auto-regressive integrated moving average (ARIMA) model, as introduced in Section 3.5.2. In this case though we simplified the model by choosing a MA order of 0, leaving only the AR part. The reason for this was that parameter estimation for an ARMA process involves solving a system of highly non-linear equations. By using an AR process we could accurately determine the model order and values of the AR components that matched the data best.

An AR(p) process can be written as

$$x[i] = - \sum_{k=1}^p a[k]x[i-k] + u[i], \quad (4.2)$$

where $\{u[i]\}$ is some independent Gaussian noise source [64]. We used the Akaike Information Criterion (AIC) [65] to estimate the model order. The AIC is defined as

$$AIC(k) = N \ln \hat{p}_k + 2k \quad (4.3)$$

where \hat{p}_k is the estimate of the noise variance for the k th order AR model and N is the number of training points. The AIC, when trained with 500 points of the video data (mean removed prior to training), yielded a model order of 3 for the JPEG data and 12 for the MPEG data. The AIC values for model orders 1 to 30 for the JPEG data are given in Figure 4.6(a).

The parameters of the AR model can be found using the Yule-Walker equations [64]. For the JPEG data they were calculated to be

$$\begin{aligned} a[1] &= -0.564635 & a[2] &= -0.319230 & a[3] &= -0.110245 \\ \sigma^2 &= 1983281 \end{aligned}$$

A sample trace of the model output (mean reinserted) is given in Figure 4.6(b).

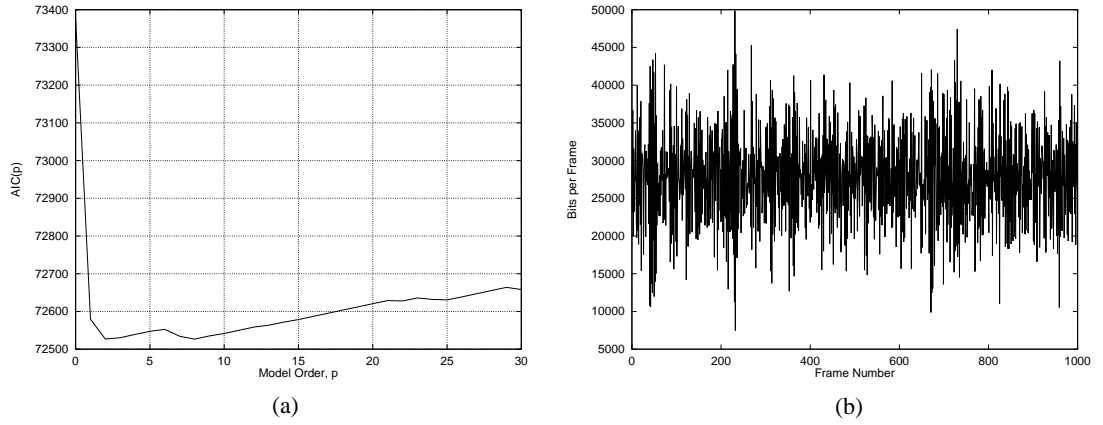


Figure 4.6: (a) The values for the Akaike Information Criterion for model order 1 to 30 (b) a 5000 point sample of the JPEG data generated by the model when fed with iid $N(0, 1)$ innovations.

To achieve the LRD we desire, it is possible to replace the iid $N(0, 1)$ noise with a unit variance fGn process with $H = 0.80$, thus creating an ARIMA model with the MA order set to zero. Therefore we generated two datasets, one with $H = 0.50$ and the other with $H = 0.80$. The statistics for each are given below.

Non LRD dataset for the JPEG data

$$N = 171000 \quad \sigma = 6324.6 \quad \nabla = 6.26$$

$$\mu = 27989.7 \quad \triangle = 66965$$

LRD dataset for the JPEG data

$$N = 171000 \quad \sigma = 6522.7 \quad \nabla = 2037.5$$

$$\mu = 27789.8 \quad \triangle = 48042.1$$

The AR and ARI models were also constructed for the MPEG data. The ACF of the AR model data is given in Figure 4.7. The large components at multiples of 12 lags are still apparent though they are not as marked as for the real MPEG data.

The statistics for the MPEG trained auto-regressive models are given below.

Non LRD dataset for the MPEG data

$$N = 171000 \quad \sigma = 18161.1 \quad \nabla = -69844.6$$

$$\mu = 16520.4 \quad \triangle = 97929.7$$

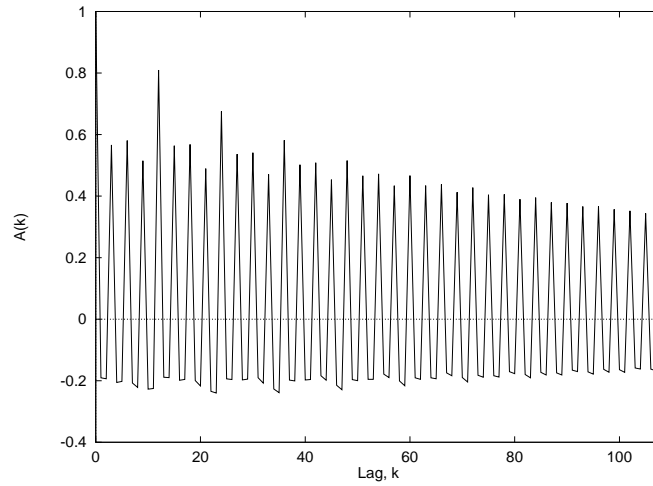


Figure 4.7: *The correlation functions for the AR model trained with data from the MPEG video trace.*

LRD dataset for the MPEG data

$$\begin{aligned}
 N &= 171000 & \sigma &= 18164.0 & \nabla &= -66709.8 \\
 \mu &= 15600.0 & \Delta &= 95042.3
 \end{aligned}$$

It is worth noting that the models generate negative traffic when trained with the MPEG data. These values must be truncated to produce a realistic traffic stream.

4.3.3 The Norros self-similar model

The Norros model was introduced in Section 3.5.1. To recap, we can generate a discrete arrival stream (in this case bits/frame) in accordance with

$$A[i] = m + kX_H[i]. \quad (4.4)$$

Where m is the mean encoded value, k is some scaling constant and $X_H[i], i = 0, 1, 2, \dots$ is a fractional Gaussian noise process [53], [61], [35]. In order to estimate the value of k

1. Find the size, mean, variance and Hurst exponent of each VBR video dataset.
2. Generate a trace of fGn with Hurst exponent and length matched to that of the JPEG data.
3. Calculate the variance of the fGn trace and use this and the variance of the video data to estimate k .

The models were constructed with the following parameters

Norros model parameters for the JPEG data

$$N = 171000 \quad m=27791$$

$$H=0.803 \quad k=6254$$

Norros model parameters for the MPEG data

$$N = 171000 \quad m=18697.7$$

$$H=0.801 \quad k=18104.0$$

4.3.4 Summary of model details

In total we have discussed two real datasets and eight model datasets. We chose the models to produce data that matched either the video data short range correlation structure (SRCS) and/or its long range correlation structure (LRCS). Table 4.1 summarises the degree to which each data model matches the correlation structure of the video data.

Model Name	SRCS	LRCS
ON-OFF	x	x
AR	√	x
Norros	x	√
ARI	√	√

Table 4.1: The summary of each model generated dataset’s ability to capture aspects of the video data.

In the next section we go on to develop some novel methods for testing the stationary nature of the datasets and to determine whether the results can help us determine more about the nature of the real data.

4.4 The wide sense stationary quotient

Heuristically, we understand that a wide-sense stationary time-series is one that has the same mean and variance over its entire existence. However, in reality, testing for this in a finite sample, like the video data is non-trivial. This is because statistical measures such as the mean and variance are normally defined as a limit as dataset size

$(N) \rightarrow \infty$. As an example consider a perfect sine wave, $X(t) = \sin \omega t + \theta$. We know that this is a stationary process but if we measure the mean and variance over small periods with respect to the period of the wave then we could conclude that the process was non-stationary. This effect is illustrated in Figure 4.8.

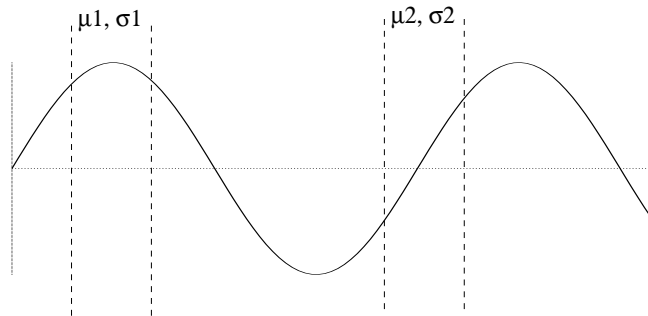


Figure 4.8: *Measuring sample mean and variance of a sine wave.*

The sample means (μ_1 and μ_2) and variances (σ_1^2 and σ_2^2) are considerably different but we know that they are measures of a stationary process. It is only by increasing the size of the sample that we might discover the process is stationary.

The wide sense stationary quotient (WSS quotient) performs mean and variance comparisons using hypothesis testing to test for equality.

4.4.1 Formal definition

Consider a dataset split into S independent segments, each of length N . If the mean of each segment is denoted as $\hat{\mu}_i$, and the variance as $\hat{\sigma}_i^2$, $i = 1, 2, \dots, S$ it is possible to test for the equality of the means and the variances between any two blocks i and j (see [66] pp. 230-231). The statistical test for the equality of means is the T test,

$$T = (\hat{\mu}_i - \hat{\mu}_j) \left(\frac{N-1}{\hat{\sigma}_i^2 + \hat{\sigma}_j^2} \right)^{\frac{1}{2}}, \quad (4.5)$$

where T has a Student t distribution with $v = 2N - 2$ degrees of freedom. The statistical test for the equality of variances is the F test,

$$F = \frac{\hat{\sigma}_i^2}{\hat{\sigma}_j^2}, \quad (4.6)$$

where F has a F distribution with $v_1 = N - 1$ and $v_2 = N - 1$ degrees of freedom.

If the total number of frames in the dataset is D . Then define

$$S_N = \frac{D}{N} \quad (4.7)$$

also define,

$$p_{ij} = \begin{cases} 1 & \text{if } |T| \leq t_{v, \frac{\alpha}{2}} \text{ and } F_{v_1, v_2, \frac{1-\alpha}{2}} \leq F \leq F_{v_1, v_2, \frac{\alpha}{2}}, \\ 0 & \text{otherwise,} \end{cases} \quad (4.8)$$

where the T and F tests are performed on blocks i and j . $t_{v, \frac{\alpha}{2}}$ is the critical value for the T test and $F_{v_1, v_2, \frac{1-\alpha}{2}}$ and $F_{v_1, v_2, \frac{\alpha}{2}}$ are the critical values for the F test. Then the wide sense stationary (WSS) quotient can be defined as

$$W_N = \frac{2}{S_N(S_N - 1)} \sum_{i=1}^{S_N-1} \sum_{j=i+1}^{S_N} p_{ij}. \quad (4.9)$$

For both the equality of means test and the equality of variances test the confidence level was chosen to be 95%. Therefore in 5% of cases where the means (variances) are equal the test will say they are not. So even for stationary data we expect 5% of the mean (variance) tests to fail. Since these tests are independent, in the worst case 10% of all stationary tests will fail for stationary data. By this reckoning we expect a WSS quotient as low as 0.9 for stationary data.

4.4.2 Results

When analysing VBR video data Beran [50] recommended using aggregate forms of the data where the time-series X_i is replaced with $X_k^{(m)}$. $X^{(m)}$ represents the original data averaged into non-overlapping blocks of size m . i.e.

$$X_k^{(m)} = \frac{1}{m} (X_{km-m+1} + \dots + X_{km}). \quad (4.10)$$

The reasoning for this is that the averaging reduces the effect of SRCS thus increasing the accuracy of the Hurst exponent estimators. Therefore we want to investigate the performance of the data at both low and high levels of aggregation. We considered $m = 1, 12$ and 144 and the results are presented in Figures 4.9 - 4.14.

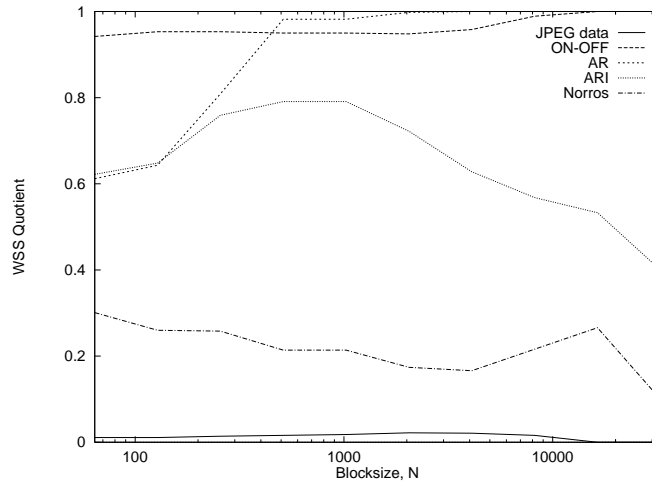


Figure 4.9: Results for the WSS quotient for the unaggregated JPEG datasets.

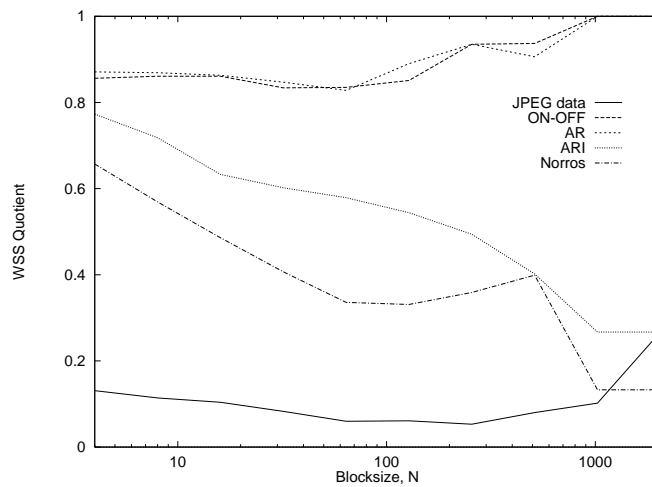


Figure 4.10: Results for the WSS quotient for the JPEG datasets averaged into blocks of 12.

4.4.3 Discussion

The WSS quotient for the JPEG data was low for all block sizes and all levels of aggregation. From this we might conclude that we cannot accept the hypothesis that this trace is stationary. However consider the ARI and the Norros results, these models also have consistently low WSS quotients but the models are known to be stationary. In these cases the self-similarity must be affecting the WSS quotient, hence we must conclude that the test cannot differentiate between non-stationary and self-similar data.

When there is no aggregation the AR model begins with a WSS quotient that is both low and of a similar value to the ARI model. However as the blocksize is increased the quotient rises for the AR data but falls for the ARI data. This result suggests that the

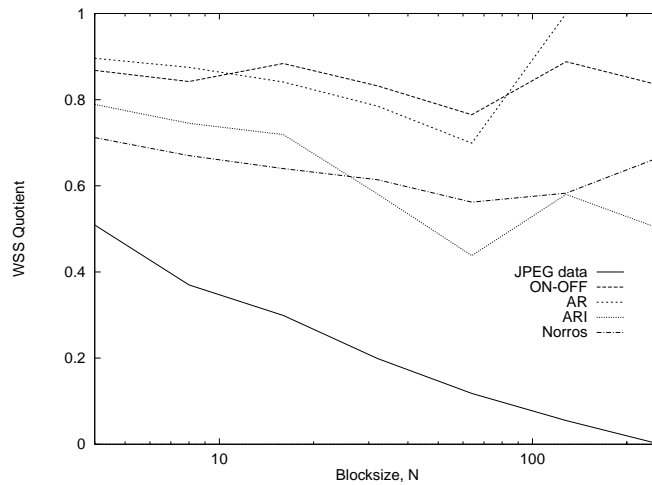


Figure 4.11: Results for the WSS quotient for the JPEG datasets averaged into blocks of 144.

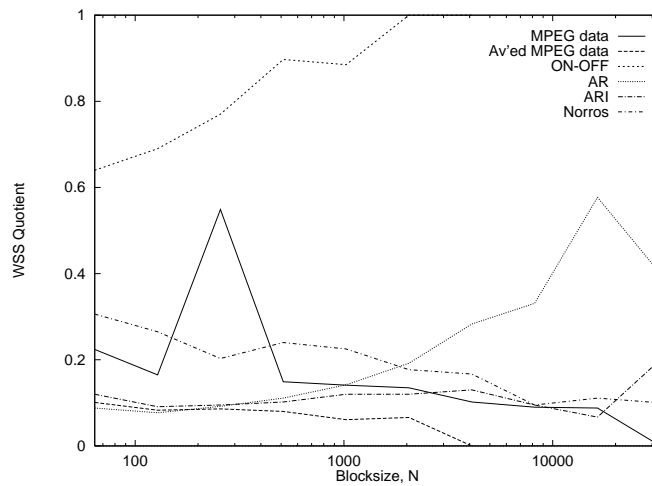


Figure 4.12: Results for the WSS quotient for the unaggregated MPEG datasets.

test can differentiate between data with LRD and that without, providing the block size is large enough.

With aggregation the WSS quotient for the JPEG data remains small (always less than 0.5). The self-similar models also retain a low WSS quotient after aggregation although their values lie in a higher range (always less than 0.8). The AR model possesses a WSS quotient in the region of 0.9 for all blocksizes when the aggregation is 12 or 144.

The MPEG data and the averaged MPEG data both possess low WSS quotients for all levels of aggregation. Again the self-similar models also have low WSS quotient but in addition the AR data also has a low WSS quotient. This value only increases when the level of aggregation and the blocksizes are large. The JPEG AR model did not perform like this and the difference may be due to the MPEG AR model having a

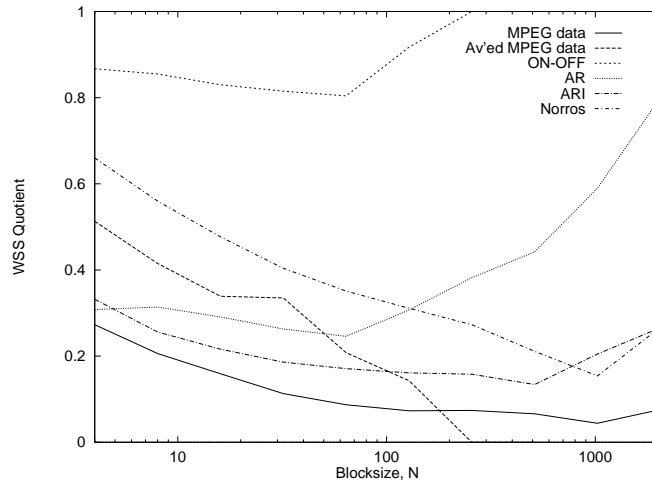


Figure 4.13: Results for the WSS quotient for the MPEG datasets averaged into blocks of 12.

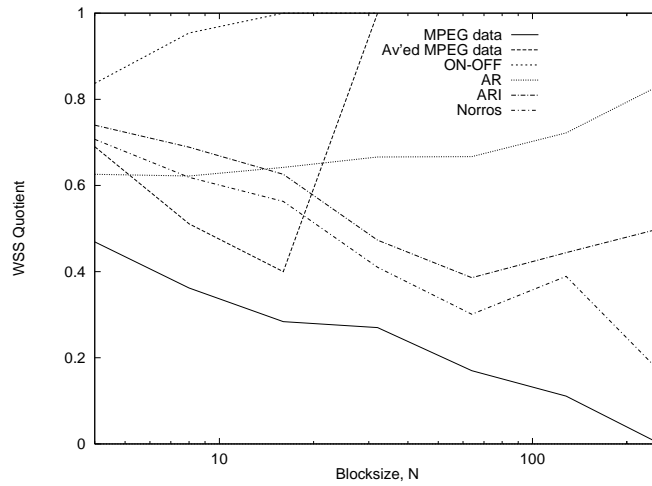


Figure 4.14: Results for the WSS quotient for the MPEG datasets averaged into blocks of 144.

much stronger SRCS.

The problem with the WSS quotient is that it assumes independence between the blocks of data it is trying to compare. The self-similar data has a wide range of random trends running through it and these bias the mean and variance estimates.

In summary the test fails to differentiate between non-stationary and self-similar data. However the test could differentiate between data with no correlation structure, that with SRCS structure and that with LRCS structure.

4.5 Hurst exponent dependence on N

In Section 4.4 it was shown that the WSS quotient was able to differentiate between datasets that possessed no dependency, short range dependency and long range dependency. However it did not assist in determining whether the VBR video data was stationary or not. In this section the effect of blocksize on the Hurst exponent estimate is investigated.

We know that the Hurst exponent is a good measure of self-similarity and that it can be estimated in a number of ways (see Section 3.4). We also know that the Hurst exponent estimate of a truly self-similar process will be constant no matter how much of the data is considered (provided convergence occurs). So by estimating the Hurst exponent for blocks of the data we can investigate its correlation structure.

4.5.1 Formal definition

Consider the S segments of a time series, each of length N . The Hurst exponent, H , can be estimated in each segment S_i , $i = 1, 2 \dots \frac{D}{N}$, using R/S analysis [39]. If the estimate obtained from the i^{th} block is defined as \hat{H}^i then let

$$\hat{H}_N = \frac{N}{D} \sum_{i=1}^{\frac{D}{N}} \hat{H}^i, \quad (4.11)$$

for suitable N . Assuming N is long enough to give reasonable convergence of the \hat{H}^i then \hat{H}_N should be independent of N for a stationary process.

4.5.2 Results

H_N was plotted against blocksize (N) for the MPEG, JPEG and model datasets using aggregation levels of 1, 12 and 144. The results are plotted in Figures 4.15 - 4.20.

4.5.3 Discussion

The H_N for the JPEG datasets was consistent with respect to N for the aggregation levels of 1 and 12. In both cases the value of H was high (in the range 0.8-1.0) but varied by less than 10% as N was varied. However when the aggregation level was 144

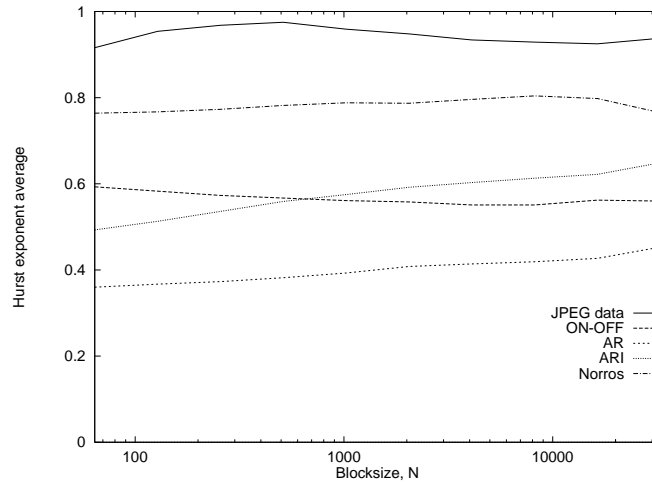


Figure 4.15: Results for H_N for the unaggregated JPEG datasets.

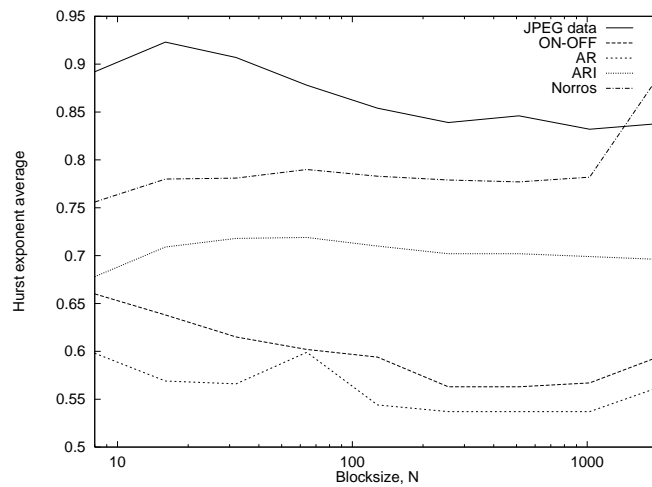


Figure 4.16: Results for H_N for the JPEG datasets averaged into blocks of 12.

the variation of H_N was over 15%. However since the block sizes for the highest level of aggregation were between 8 and 256 it is highly probable that some of this variation is due to non-convergence of the Hurst exponent estimator.

The MPEG data behaved in the opposite manner; the H_N for the case of no aggregation varied by over 50% and variation decreased as aggregation increased. After aggregation the H_N was estimated to be in the range 0.80-0.85 for all N . An explanation for this result is that the SRCS influences the Hurst exponent estimation technique when N is small but as N is increased the SRCS becomes less important and it is the LRCS that affects the estimator. This would account for H_N being small for small N but increasing as N increases.

The H_N estimates for the averaged MPEG data were more uniform with varying N . This is because the SRCS has been partially removed by the aggregating process. In

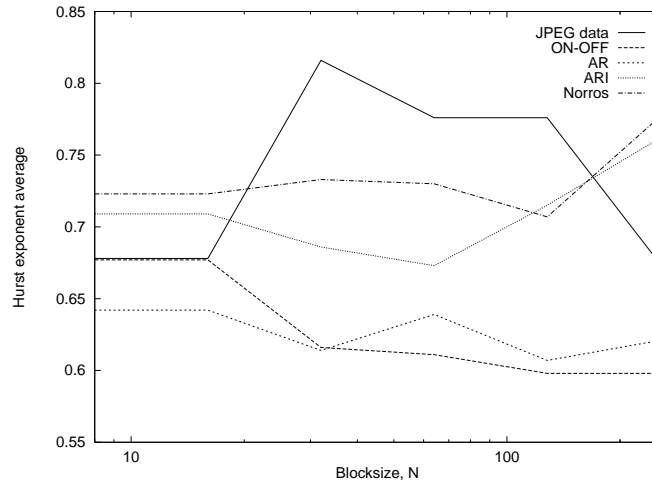


Figure 4.17: Results for H_N for the JPEG datasets averaged into blocks of 144.

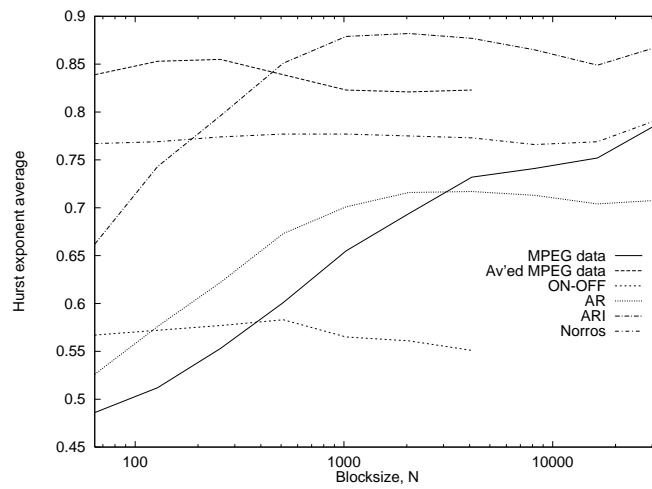


Figure 4.18: Results for H_N for the unaggregated MPEG datasets.

all cases the Hurst exponent estimate was in the range 0.75-0.90.

One interesting result is that the ARI model is the only one that captured the H_N behaviour of the unaggregated MPEG data. The ARI model also had a H_N result that increased with N , probably because it too has a strong SRCS and LRCS.

Another point to note is that the ON-OFF and AR models have consistently low values of H_N for all N and all aggregation levels. In some cases H_N did reach a value of greater than 0.6 for these processes which suggests the estimates of H_N are not that accurate (to within ± 0.1).

The Norros model has H_N consistent and close to the expected value of 0.8 for all levels of aggregation and blocksize. This is true for both the MPEG and JPEG Norros data. The ARI model though has a H_N that is often less than 0.8. In the JPEG unaggregated

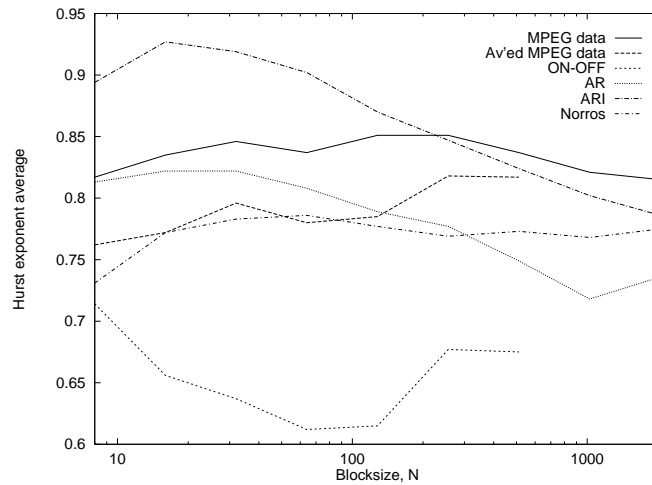


Figure 4.19: Results for H_N for the MPEG datasets averaged into blocks of 12.

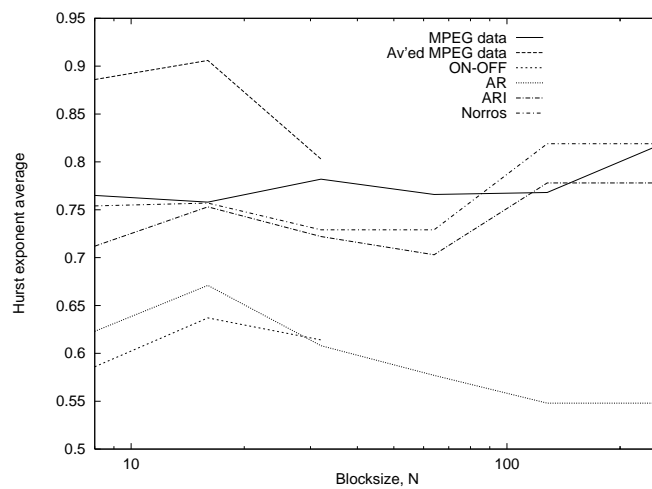


Figure 4.20: Results for H_N for the MPEG datasets averaged into blocks of 144.

case H_N barely exceeds 0.6 though this may be due to the SRCs discussed above.

4.6 Conclusions

The purpose of this chapter was to investigate whether the JPEG and MPEG VBR video traces were non-stationary or self-similar. The reason for the investigation was to determine whether traffic characterisation for CAC schemes would be required to take into account the fact that LRD exists in the data.

We applied two non-parametric tests to our real datasets and a variety of suitably constructed model datasets. From the results of these tests we can draw the following conclusions about the data.

1. The WSS quotient is able to differentiate between iid data, data with a SRCS and data with a LRCS. However it is not able to differentiate between non-stationary and self-similar data.
2. The H_N measure was able to differentiate between data with a SRCS, data with a LRCS and data with both (the ARI model).
3. The ARI model was the only one that obtained similar values for H_N for the unaggregated MPEG data. This result suggests the MPEG data possesses a SRCS and LRCS.
4. The self-similar models approximated the behaviour of the real data to a greater extent than the other models with regards to the measures considered in this chapter.

The tests cannot tell us whether the video data is non-stationary or self-similar but it does reveal that self-similar models can be used to generate data that behaves like the VBR video teletraffic.

Testing the Gaussian assumption for teletraffic traces

5.1 Introduction

In Chapter 4 we presented results that attempted to ascertain whether VBR video data was nonstationary or self-similar. The conclusion of that work was that differentiating between the two was not a trivial task since the LRD of self-similar signals can give the impression that a signal has moments that vary with time. However, without ascertaining the underlying nature of the VBR video processes, we were able to show that it was the self-similar models that best captured the behaviour of the real traffic.

Chapter 4 investigated the time-domain (correlation structure) characteristics of real teletraffic traces. Now we wish to turn our attention to the marginal distribution of teletraffic. In [61] Norros applied his model to Bellcore Ethernet traces and among the results he commented

The most obvious difference [with the trace OctExt.TL] is that the overall mean rate is much lower than the mean deviations of averages over intervals even in the order of magnitude of seconds. The distribution of the local rate is thus strongly non-Gaussian, which indicates that the traffic is not aggregated from sufficiently many independent streams to allow the applicability of a Gaussian model.

In the remainder of this chapter we test the Gaussian assumption for the Norros model using the data discussed in Section 5.2. We do this by transforming the data into what the Norros model assumes to be fGn and then test this using two tests for normality (the χ^2 test and the Shapiro and Wilk test).

We then qualitatively and quantitatively assess the marginal distributions of the transformed teletraffic data and suggest a possible alternative to the Gaussian distribution.

5.2 The data

The datasets used in this chapter fall into two categories. The first category is VBR video and the two datasets used in Chapter 4 are considered. These datasets are a JPEG and MPEG encoding of the motion picture Star Wars. The salient statistics of the datasets can be found in Sections 4.2.2 and 4.2.3.

The other category is packet data from a LAN. The datasets are constructed from the Bellcore Ethernet log files [33] available via anonymous FTP from ftp.bellcore.com. Each file (pAug.TL, pOct.TL and OctExt.TL) consists of one million rows of two columns of data. The first column contains an accurate time stamp of the arrival time of the packet and the second column contains the packet size. We constructed four datasets from each log file where each observation represented the number of bytes sent across the network per timeslot. We choose a range of timeslots from 0.01 seconds to 100 seconds and denoted a dataset constructed with the timeslot Δt as $\{W_i^{\Delta t}\}_{i=0,1,\dots}$. The salient statistics of the Ethernet datasets are given in Tables 5.1 - 5.3.

pAug.TL					
	N	μ	σ	∇	Δ
$W^{0.01}$	314283	1381.8	2226.9	0	13550
$W^{0.1}$	31428	13818	13074.9	0	95702
W^1	3142	138177	85615.3	6795	569707
W^{10}	314	1378705	579303	411448	3521340

Table 5.1: Statistics of the datasets generated from the Bellcore trace pAug.TL

pOct.TL					
	N	μ	σ	∇	Δ
$W^{0.01}$	130390	3060.3	3472.7	0	13195
$W^{0.1}$	13039	30603	17147	0	95090
W^1	1303	306009	123041	27915	777544
W^{10}	130	3046074	934187	825621	5628218

Table 5.2: Statistics of the the datasets generated from the Bellcore trace pOct.TL

OctExt.TL					
	N	μ	σ	∇	Δ
W^1	122798	1142.2	2836.1	0	76880
W^{10}	12279	11422	22435	98	330895
W^{100}	1227	114220	177954	2984	1546032
W^{1000}	122	1140350	1231826	75645	7207948

Table 5.3: Statistics of the the datasets generated from the Bellcore trace OctExt.TL

5.3 Data transform

Recall from Section 3.5.1 that the Norros model generates a discrete cumulative arrival stream, $\hat{A}[i]$, in accordance with

$$\hat{A}[i] = mi + \sqrt{am}B_H[i], \quad i = 0, 1, 2, \dots \quad (5.1)$$

Where m is the mean arrival size per time unit, a is a variance parameter and $B_H[i]$ is a discrete Brownian motion process with Hurst exponent H , [35], [67]. This model has been studied in some detail and is employed in teletraffic modelling due to its parsimonious nature and its ability to emulate the apparent self-similar nature of real traffic [53], [61].

Since $\hat{A}[n]$ is a cumulative arrival process we can obtain the arrival process by differencing (5.1) over k and $k - 1$,

$$A[k] = m + \sqrt{am}X_H[k]. \quad (5.2)$$

$X_H[k]$ is a discrete fractional Gaussian noise (fGn) process [67] and has a marginal distribution that is Gaussian in nature. If the real dataset, $\{X[i]\}_{i=1,2,\dots,n}$, is assumed to be the output of a Norros model then we can transform it into an assumed scaled fractional Gaussian noise process, $\sqrt{am}X_H[k]$. If the data truly conforms to a Norros model then we would expect $X[i] - m \sim N(0, am)$. Figure 5.1(a) gives a sample trace of $X[i] - m$ for the JPEG data trace whilst Figure 5.1(b) gives the 200 bin histogram of $X[i] - m$. Upon observation it is obvious that the distribution of $X[i] - m$ is approximately symmetric and bell-shaped but it is not apparent if it is normal.

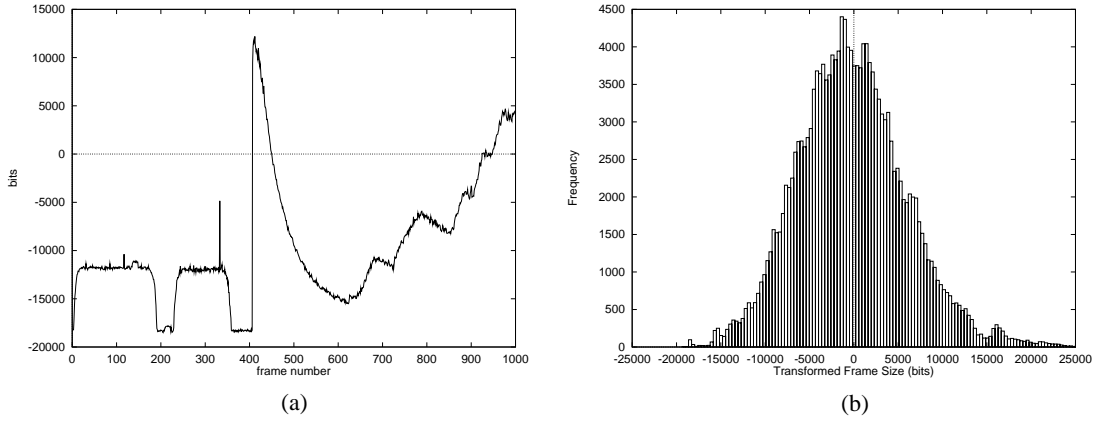


Figure 5.1: (a) Sample trace of the converted data and (b) the 200 bin histogram of the converted dataset.

5.4 Testing the data with the χ^2 -test

Since determining the nature of a distribution from a finite sample of random variables is classed as a hypothesis test it seems natural to apply a goodness of fit test. One of the best known goodness of fit tests is the χ^2 -test (chi-square) [68], [69]. In this section we introduce the χ^2 -test for normality and apply it to the datasets mentioned in Section 5.2. The results are presented and discussed in Section 5.4.2

5.4.1 The χ^2 -test for normality

If X_1, X_2, \dots, X_n are iid random variables we want to test whether $X_i \sim N(\frac{x-\mu}{\sigma})$. We assume both the mean and variance of the distribution to be unknown, hence the maximum likelihood estimates of μ and σ^2 are

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad (5.3)$$

and

$$\bar{s}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (5.4)$$

respectively.

Next we wish to group the data into r bins over the interval $-\infty < x < \infty$, i.e $x_0 = -\infty$, $x_r = \infty$ and $\text{bin1} = (x_0, x_1]$, $\text{bin2} = (x_1, x_2]$, \dots , $\text{bin}r = (x_{r-1}, x_r]$. Then we denote $\underline{v}^* = (v_1^*, v_2^*, \dots, v_r^*)$ to be the frequency vector obtained by grouping the

data into bins. Next compute $\underline{e}^* = (e_1^*, e_2^*, \dots, e_r^*)$, the frequency vector obtained using $N(\bar{X}_n, \bar{s}_n^2)$ (i.e. what we expect the frequency vector to be if the sample is perfectly normally distributed). We now wish to compare the sample vector with the perfect one. This is done by computing

$$\chi_0^2 = \sum_{j=1}^r \frac{(v_j^* - e_j^*)^2}{e_j^*}. \quad (5.5)$$

It is then possible to test the hypothesis of normality by choosing a significance level, α , (e.g. 0.01 or 0.05) and determining

$$P(\chi_{r-3}^2 \leq c) = 1 - \alpha \quad (5.6)$$

Where χ_k^2 is a chi-square distribution with $r - 3$ degrees of freedom. We use $k = r - 3$ as we assume both the mean and variance of the distribution to be unknown. The critical value of the χ_k^2 distribution can be found using tables.

5.4.2 Results

In [68] the authors recommended using a symmetric bin procedure (i.e. $x_2 = x_{r-1}, x_3 = x_{r-2}, \dots$). Since the distributions under investigation are assumed to be symmetric this advice was followed. In all cases the number of bins was fixed at 14 with bin 1 = $(-\infty, -3]$ and bin 14 = $(3, \infty]$. The results of the χ^2 -test for the JPEG data are recorded in Table 5.4.

The χ^2 -test value was therefore 9611.5. The critical value for 95% confidence level with 11 degrees of freedom is 19.68. Therefore the hypothesis that the transformed JPEG data is from a Gaussian distribution can not be accepted at a 95% confidence level. This does not necessarily mean the null hypothesis (that the data is drawn from a non-Gaussian distribution) must be accepted.

The χ^2 value for the MPEG video data was 239803 which also exceeds the critical value for a 95% confidence level.

The results for the Ethernet traces are summarised in Tables 5.5 to 5.7.

All the datasets constructed from the Ethernet log files had χ^2 values greater than the critical value for 95% confidence except the W^{10} for pOct.TL. This suggests that the

JPEG data: $\bar{X} = 2.779120e + 04$ $\bar{s}^2 = 6.254219e + 03$				
bin	range	v_j	e_j	$\frac{(v_j - e_j)^2}{e_j}$
1	$-\infty$ to -3.0	5	222.3	212.4
2	-3.0 to -2.5	311	837.9	331.3
3	-2.5 to -2.0	2349	2736.0	54.7
4	-2.0 to -1.5	6290	7524.0	202.3
5	-1.5 to -1.0	16725	15732.0	62.7
6	-1.0 to -0.5	28212	25650.0	255.9
7	-0.5 to 0.0	35388	32832.0	199.0
8	0.0 to 0.5	33419	32832.0	10.5
9	0.5 to 1.0	22920	25650.0	290.6
10	1.0 to 1.5	14072	15732.0	175.2
11	1.5 to 2.0	6200	7524.0	233.0
12	2.0 to 2.5	2353	2736.0	53.6
13	2.5 to 3.0	1613	837.9	717.0
14	3.0 to ∞	1143	222.3	3813.3

Table 5.4: Results of placing the JPEG data into bins to determine the χ^2 -test value.

pAug.TL				
	N	\hat{H}	$c_{95\%}$	χ^2
$W^{0.01}$	314283	0.82	19.68	443600
$W^{0.1}$	31428	0.85	19.68	21830
W^1	3142	0.85	19.68	1178
W^{10}	314	0.83	19.68	65.4

Table 5.5: Results of the χ^2 -test for the Bellcore trace pAug.TL

transformed data is not drawn from a normal distribution. However what is interesting is that the χ^2 values always get smaller as the amount of aggregation (i.e. timeslot size) increases. A smaller χ^2 value implies that the confidence with which we can say the data is drawn from a normal distribution is increased. Therefore we can conclude that aggregating the data makes it appear more normal. This result is in agreement with the central limit theorem (CLT).

pOct.TL				
	N	\hat{H}	$c_{95\%}$	χ^2
$W^{0.01}$	130390	0.84	19.68	161315
$W^{0.1}$	13039	0.84	19.68	928.5
W^1	1303	0.85	19.68	118.0
W^{10}	130	0.85	19.68	18.3

Table 5.6: Results of the χ^2 -test for the Bellcore trace pOct.TL

OctExt.TL				
	N	\hat{H}	$c_{95\%}$	χ^2
W^1	122798	0.83	19.68	307832
W^{10}	12279	0.82	19.68	26204
W^{100}	1227	0.85	19.68	1301
W^{1000}	122	0.80	19.68	174.4

Table 5.7: Results of the χ^2 -test for the Bellcore trace OctExt.TL

5.5 Testing the data with the Shapiro-Wilk test

In Section 5.4 the χ^2 -test for normality was introduced. It was chosen because it is the best known of the goodness of fit tests. However a paper by Shapiro and Wilk concluded that it was the Shapiro-Wilk (SW) test that performed best at determining whether a dataset was drawn from a normal distribution or not [69]. They also observed that the χ^2 -test can behave erratically since the bin intervals are chosen arbitrarily. In addition the SW test is a composite test as opposed to the χ^2 -test which is a hypothetical test (i.e. the SW test is scale and origin invariant, whilst μ and σ^2 must be estimated for the χ^2 -test and the data scaled accordingly).

5.5.1 The Shapiro-Wilk test

The SW test is based on a statistic D which is a ratio of an unbiased estimate of the population standard deviation to the sample standard deviation. If we assume X_1, \dots, X_n is a sample set of size N and if the set is ordered such that $X_{1,N} \leq X_{2,N}, \dots \leq X_{N,N}$, the statistic D is given by

$$D = \frac{T}{N^2 S}, \quad (5.7)$$

where

$$T = \sum_{i=1}^N \left\{ i - \frac{1}{2}(N+1) \right\} X_{i,N}, \quad (5.8)$$

and

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{N}, \quad (5.9)$$

where \bar{X} is the sample mean. The statistic D has $\mu \neq 0$ and $\sigma^2 \neq 1$ but can be normalised to give the statistic Y using

$$Y = \frac{(D - (2\sqrt{\pi})^{-1})\sqrt{N}}{0.02998598}. \quad (5.10)$$

If the set X is not drawn from a normal distribution then $E[Y] \neq 0$. The critical levels for a $(1 - \alpha)$ confidence level can be obtained from tables [70] or by calculation [71], [72].

Some datasets to be tested are of size, $n > 1000$ and tables of the critical values for the statistics D and Y are not available for such large values. Therefore D'Agustino's approximation to the asymptotic distribution of D was employed [71], [72].

For $n > 1000$ the percentiles of the statistic D can be calculated as follows. Assume D_p and Z_p are the $100p$ ($0 \leq p \leq 1$) percentiles of D and $N(0, 1)$ respectively. Then

$$D_p = E(D) + V_p \sqrt{\text{VAR}(D)} \quad (5.11)$$

where

$$V_p = Z_p + \frac{\gamma_1(Z_p^2 - 1)}{6} + \frac{\gamma_2(Z_p^3 - 3Z_p)}{24} - \frac{\gamma_1^2(2Z_p^3 - 5Z_p)}{36} \quad (5.12)$$

$E(D)$, $\text{VAR}(D)$, γ_1 and γ_2 are the mean, variance and third and fourth cumulants of D . If $n > 200$ then the following approximations can be used,

$$E(D) \approx \frac{N-1}{2\sqrt{2N\pi}} \cdot \frac{\Gamma(\frac{1}{2}N - \frac{1}{2})}{\Gamma(\frac{1}{2}N)} \quad (5.13)$$

$$\text{VAR}(D) \approx \frac{0.029986^2}{N} \quad (5.14)$$

$$\gamma_1 \approx -\frac{8.463}{\sqrt{N}} \quad (5.15)$$

$$\gamma_2 \approx \frac{107.9}{\sqrt{N}} \quad (5.16)$$

The percentiles of Y can then be found using (5.10).

5.5.2 Results

The datasets $W^{0.01}$, $W^{0.1}$, W^1 and W^{10} for pAug.TL and pOct.TL were transformed into what is assumed to be fGn as in Section 5.3 and then tested using the adapted SW test and the statistic Y was recorded. The lower and upper critical values for a .95 confidence level were calculated for the required N (c_{low} and c_{high}). Also, the Hurst exponent estimate, \hat{H} , for each of the $X_H[n]$ datasets was calculated using the R/S -statistic method [50]. The trace file OctExt.TL was gathered over a longer time scale and hence the SW test and \hat{H} were calculated for W^1 , W^{10} , W^{100} and W^{1000} . The results and critical values for all the datasets are given in Tables 5.8 - 5.10.

pAug.TL					
	N	\hat{H}	c_{low}	c_{high}	Y
$W^{0.01}$	314283	0.82	-1.967	1.953	-1218.94
$W^{0.1}$	31428	0.85	-1.982	1.937	-149.36
W^1	3142	0.85	-2.030	1.887	-29.27
W^{10}	314	0.83	-2.176	1.724	-7.16

Table 5.8: Results of the adjusted SW test for the Bellcore trace pAug.TL

pOct.TL					
	N	\hat{H}	c_{low}	c_{high}	Y
$W^{0.01}$	130390	0.84	-1.971	1.949	-206.5
$W^{0.1}$	13039	0.84	-1.994	1.925	-4.22
W^1	1303	0.85	-2.060	1.847	-8.24
W^{10}	130	0.85	-2.290	1.585	-10.71

Table 5.9: Results of the adjusted SW test for the Bellcore trace pOct.TL

OctExt.TL					
	N	\hat{H}	c_{low}	c_{high}	Y
W^1	122798	0.83	-1.971	1.948	-1488.39
W^{10}	12279	0.82	-1.996	1.923	-403.53
W^{100}	1227	0.85	-2.072	1.843	-106.43
W^{1000}	122	0.80	-2.298	1.572	-25.77

Table 5.10: Results of the adjusted SW test for the Bellcore trace OctExt.TL

In a similar manner we tested the JPEG and MPEG datasets with the SW test. The results are recorded in Table 5.11.

The results show that, for a 95% confidence level, all the datasets for all the Ethernet traces and the VBR video failed the adjusted Shapiro-Wilk test. Therefore the assumed fGn datasets produced using the transform cannot be true fGn (which would be expected to pass the SW test).

VBR video datasets					
	N	\hat{H}	c_{low}	c_{high}	Y
JPEG	171000	0.80	-1.651	1.639	-73.15
MPEG	171000	0.85	-1.651	1.639	-816.8

Table 5.11: Results of the adjusted SW test for the VBR video datasets.

5.6 Considering LRD in the data

The results in Sections 5.4.2 and 5.5.2 indicate that the datasets in question are not drawn from a Gaussian distribution. However it is possible that the self-similar nature of the data could be affecting the results. Due to LRD, non-parametric measures of self-similar time-series can suffer from convergence problems. In order to investigate this, true fGn traces with varying H and N were constructed and tested using the Shapiro-Wilk test. This experiment was repeated 500 times and the percentage pass rates are recorded in Table 5.12.

N	True fGn trace H					
	0.5	0.6	0.7	0.8	0.9	1.0
1024	80.8	71.8	86.0	85.4	82.6	74.4
2048	83.8	88.0	91.4	89.0	88.0	82.4
4096	92.4	93.2	91.6	91.4	90.2	79.6
8192	96.6	96.6	96.2	95.4	89.8	73.6
16384	98.6	97.4	98.2	96.2	85.5	57.4
32768	97.8	99.0	97.4	96.4	91.2	41.0
65536	99.0	98.8	98.8	98.0	85.6	42.8
131072	98.6	98.0	99.0	98.2	87.0	31.4
262144	99.0	99.4	98.4	98.4	84.2	26.2
524288	98.6	98.6	98.6	97.0	75.8	23.2

Table 5.12: Percentage of the SW tests that passed for true fGn traces.

The bold entries in Table 5.12 indicate where the fGn traces seem to pass the test the expected number of times for a 95% confidence level.

Extreme values of H (i.e. $H \geq 0.9$) and low values of N seem to produce fewer than expected passes and when $H = 1.0$ the pass rate decreases with N . This is due to the fact that a time-series with such a large H possesses very strong correlations over all time scales and any non-parametric measure (such as equations (5.7)-(5.9)) will not converge for any N .

The length and estimated Hurst exponent values for six of the datasets lie in the region in Table 5.12 where the self-similar process passed the tests the expected number of

times. From this we can conclude that it is not the LRD in the data that causes the tests to fail.

5.7 Qualitative study of the data distributions

The results in Sections 5.4.2 and 5.5.2 conclusively show that, even at high levels of aggregation, the data under consideration is not drawn from a normal distribution. In order to determine a possible alternative we want to compare the data with a normal distribution in a more qualitative fashion. Hopefully this visual comparison will allow us to determine how the data distributions differ from the normal.

5.7.1 Procedure

Each of the transformed datasets used in Sections 5.4 and 5.5 were normalised by their sample means and standard deviations. The cumulative density functions (CDFs), $F(x)$, of these datasets were then constructed and compared with the exact CDF for $N(0, 1)$ (obtained from tables).

5.7.2 Results

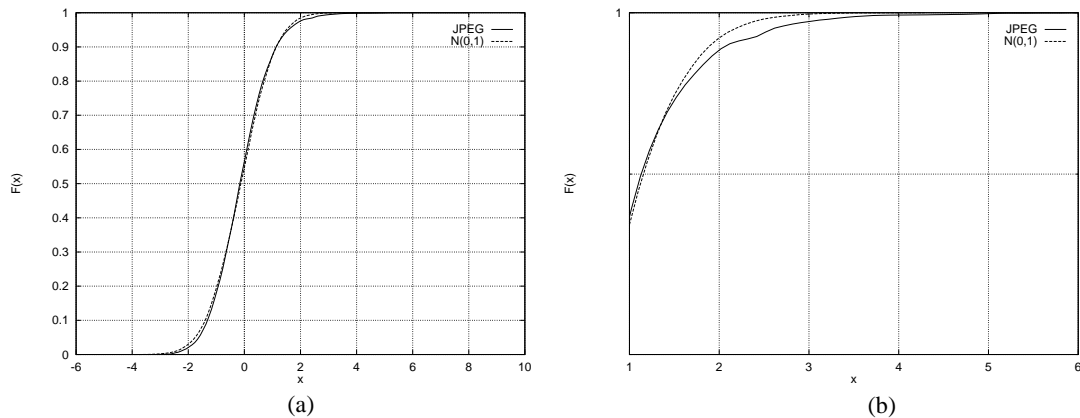


Figure 5.2: (a) The CDF of the transformed JPEG data compared to the $N(0, 1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.

The JPEG data, when normalised is very similar in distribution to the $N(0, 1)$ case (see Figure 5.2(a)). This may explain why the Norros model performs well for VBR video data. There is however a slight difference and this is more clear in Figure 5.2(b). It is

evident that $F(x)$ climbs more slowly for the real data than for $N(0,1)$. This implies that events distant from $x = 0$ are more probable for the real data than for the normal curve i.e. the real data is more impulsive.

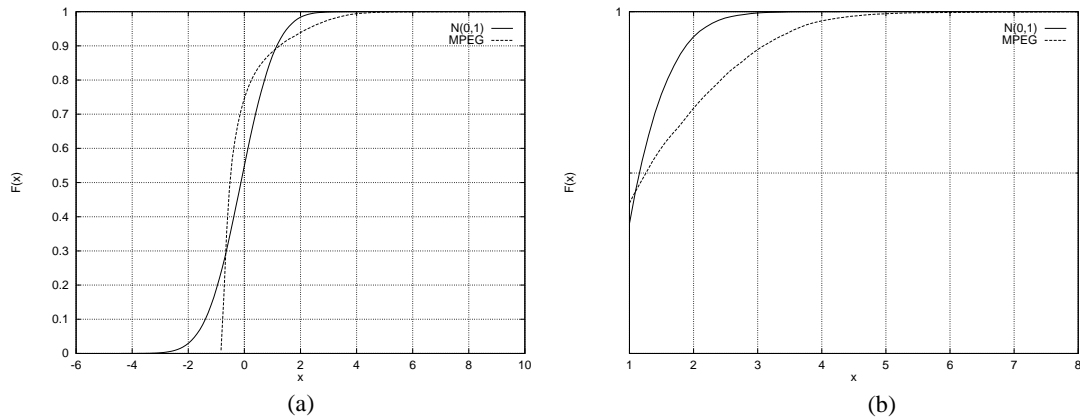


Figure 5.3: (a) The CDF of the transformed MPEG data compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.

The MPEG data does not match the $N(0,1)$ CDF as well as the JPEG data. It is much less symmetrical and a lot more impulsive than the JPEG data. For example $P(x > 3) = 0.006$ for the JPEG data and $P(x > 3) = 0.023$ for the MPEG data.

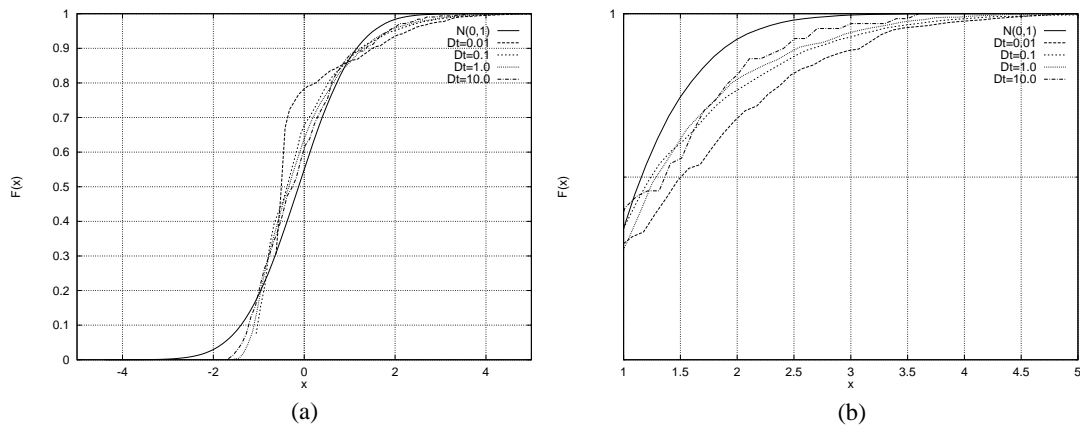


Figure 5.4: (a) The CDF of the transformed Ethernet data *pAug.TL* compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.

In all cases the Ethernet datasets have distribution tails that decay slower than the $N(0,1)$ case. Some of the empirical CDFs suffer from truncation and a lack of data but over the range $1 \leq x \leq 5$ the data is almost always more impulsive than the Gaussian case

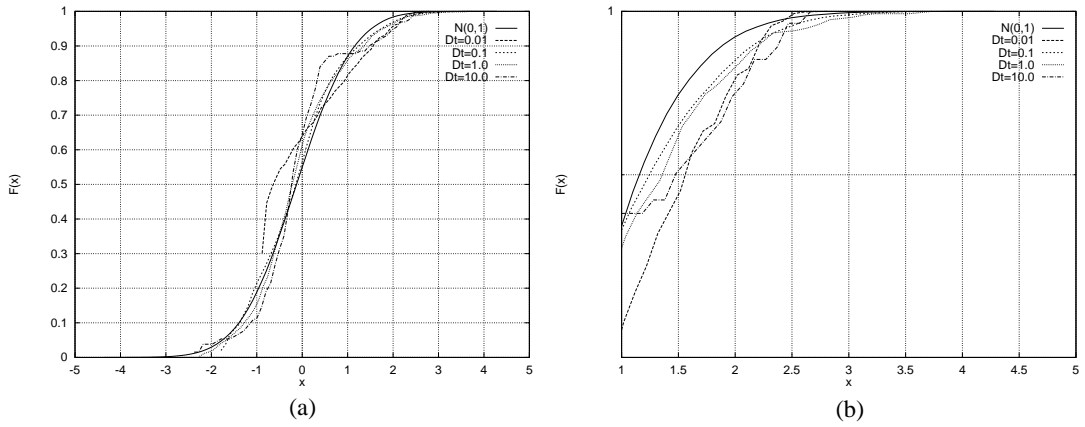


Figure 5.5: (a) The CDF of the transformed Ethernet data $pOct.TL$ compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.

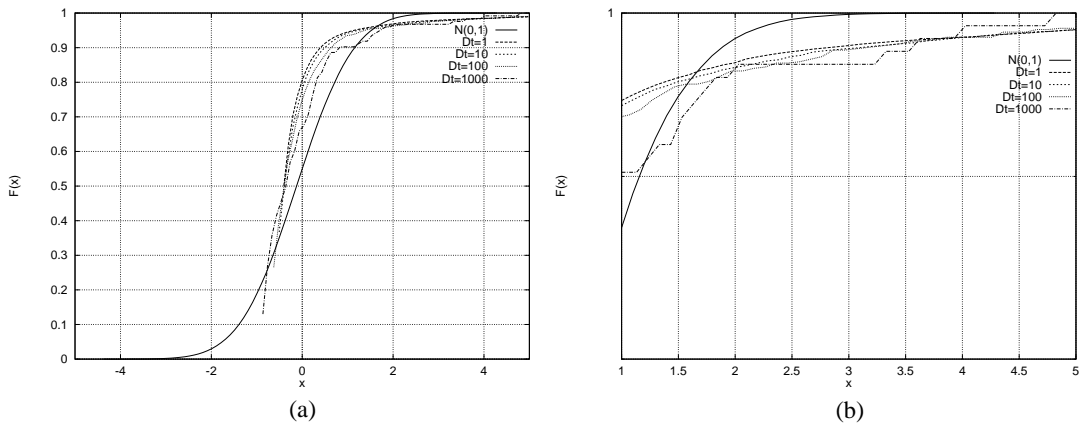


Figure 5.6: (a) The CDF of the transformed Ethernet data $OctExt.TL$ compared to the $N(0,1)$ distribution and (b) the same plot zoomed in on the tails of the distributions.

5.8 Quantitative study of the data distributions

The previous two sections have concluded that the teletraffic data under consideration is (i) not drawn from a Gaussian distribution and (ii) more impulsive than the Gaussian distribution.

The impulsive nature is due to the fact that the tails of the datasets' distributions all decay at a rate slower than the Gaussian case. This rate of decay may be exponential or it may obey some form of power law. If the later is true then the distribution can be classed as heavy tailed and

$$P(X > x) \sim x^{-k}. \quad (5.17)$$

Heavy tailed distributions that have been applied to teletraffic modelling include the

Pareto [52] and the Weibull [20], [61]. There has also been some discussion concerning the link between self-similar processes and heavy-tailed distributions [73], [74]. For example in [74] Resnick and Samorodnitsky showed that self-similar arrival processes form queues with heavy tailed distributions when fed through buffered switches.

We wish to determine whether our data is heavy tailed or not and to do this we can apply the Hill estimator [75]. Let X_1, \dots, X_n be one of the transformed datasets and then let the order statistic be given by $X_{(1)} > X_{(2)} > \dots > X_{(n)}$. Then we define the Hill estimator to be

$$H_{l,n} = \frac{1}{l} \sum_{i=1}^l \log \frac{X_{(i)}}{X_{(l+1)}}, \quad (5.18)$$

where $k < n$ is the number of upper order statistics used. If X_1, \dots, X_n are iid from a distribution F then (5.18) is the MLE of k iff

$$1 - F(x) = x^{-k} L(x). \quad (5.19)$$

Where $L(x)$ is some slowly varying function. If the distribution is heavy tailed then we expect some consistency in the Hill estimate as the number of order statistics is varied.

Figure 5.7 plots the Hill estimate of k for order statistics between 1 and 20000 for the JPEG, MPEG and Ethernet data. The Ethernet trace was a sample from W^1 for pAug.TL.

The Hill estimate for the JPEG data is not that consistent. It gradually decreases from approximately 9 down to 2 before dropping away rapidly as l approaches 20000. The MPEG data is better behaved with $k \approx 1$ over a wide range of order statistics though it too drops away rapidly as l approaches 20000. This rapid drop off is due to the heavy tailed assumption failing for the smaller values of the data. From this we can conclude that the MPEG data appears to be heavy tailed. It is harder to conclude this for the other datasets as the Hill estimate varies with the number of order statistics.

In [73] Resnick commented upon some of the problems encountered with the Hill estimator. These include problems with convergence in l and unstable estimates of k .

Another measure was proposed by DuMouchel [76] which “lets the tails speak for themselves”. By this, he meant that the measure makes no underlying assumption about the distribution but merely examines the upper 10th percentile of the distribution. This

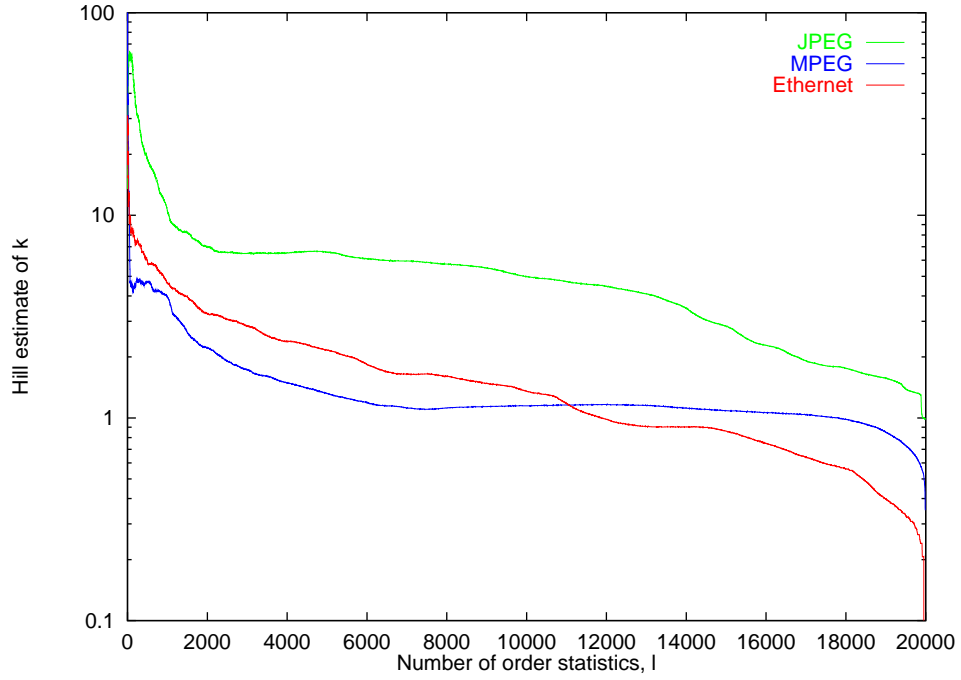


Figure 5.7: *The Hill estimate for JPEG, MPEG and Ethernet data.*

measure is called the generalised Pareto distribution.

$$P(X > x | X > x_0) = \begin{cases} (1 + \gamma(x - x_0)/\sigma)^{\frac{-1}{\gamma}} & \text{for } \gamma > 0, \\ \exp(-(x - x_0)/\sigma) & \text{for } \gamma = 0, \\ (1 + \gamma(x - x_0)/\sigma)^{\frac{-1}{\gamma}} & \text{for } \gamma < 0, x \leq x_0 + \frac{\sigma}{|\gamma|} \\ 0 & \text{for } \gamma < 0, x > x_0 + \frac{\sigma}{|\gamma|}, \end{cases} \quad (5.20)$$

where $\gamma \in (-\infty, \infty)$, $\sigma \in (0, \infty)$ and $x_0 \in (0, \infty)$. This measure can be fitted to the data by determining x_0 as the 0.9 quantile and then minimising

$$I(\gamma, \sigma) = 1000 \log \sigma + (1 + \gamma^{-1}) \sum_{k=0}^{k=999} \log \left(1 + \frac{\gamma z_k}{\sigma} \right) \quad (5.21)$$

with respect to σ and γ . z_k refers to the $(0.90005 + 0.0001k)$ th quantiles of the data distribution. We determined γ for several of the datasets discussed in this chapter and the results are recorded in Table 5.13.

DuMouchel claimed in [76] that $\hat{\gamma} > 0$ implies that the tails decay in accordance with a power law and $\hat{\gamma} \leq 0$ implies the tails decay at a faster rate than a power law. However in [77] McCulloch showed that stable distributions could produce a $\hat{\gamma}$ as low as -0.178. He also showed that stable distribution with α as low as 1.65 could produce a negative $\hat{\gamma}$. Therefore we can conclude that the three of the datasets considered are heavy tailed

Dataset	$\hat{\gamma}$
JPEG	0.029
MPEG	-0.113
pAug.TL ($W^{0.01}$)	-0.157
pAug.TL ($W^{0.1}$)	-0.094
OctExt.TL (W^1)	0.158
OctExt.TL (W^{10})	0.132

Table 5.13: Results for $\hat{\gamma}$ for the generalised Pareto technique.

(JPEG, OctExt.TL (W^1), and OctExt.TL (W^{10})) and we are unable to rule out heavy tails for the other datasets.

5.9 Conclusions

We have presented a considerable amount of evidence to show that the Gaussian distribution assumption may not be an optimal choice for Norros' teletraffic model. Two distribution tests, the χ^2 -test and the Shapiro-Wilk test, were applied to both VBR video and Ethernet data. The results of these were quite conclusive in that almost all datasets failed both tests implying a non-Gaussian distribution in the data.

When the transformed data CDFs were compared with the exact $N(0, 1)$ CDF the data was seen to be more impulsive (i.e. heavier tailed) in almost all cases. We used the Hill estimate and generalised Pareto measure to attempt to determine whether this behaviour was due to an algebraic type decay in the distribution. Our result concluded that this did indeed appear to be the case for some datasets and could not be ruled out for the other datasets.

Chapter 6

Stable distributions & stable parameter estimation

6.1 Introduction

In this chapter we investigate the use of stable distributions for teletraffic modelling. The justification for this is the evidence in Chapter 5 which proves that teletraffic data is more impulsive than the Gaussian case. The reasons we have chosen to investigate stable distributions include:

- They can be chosen to be symmetric or asymmetric.
- They conform to the generalised central limit theorem (GCLT).
- They are almost always heavy tailed and the heaviness of their tails is controlled by a single parameter.
- Although other heavy tailed distributions have been applied to teletraffic few results on the use of stable distributions have been published [78], [79].

This chapter begins with an introduction to stable distributions. In Section 6.2 the properties and parameters of a stable distribution are introduced. We then go on to consider how best to estimate stable parameters for data. Although several estimation techniques have been developed there has never been a direct comparison between them. In this chapter we wish to determine which estimation technique works best and to investigate whether we can increase the accuracy of our estimates by combining results from two or more techniques. By doing this we hope to learn more about how stable parameters estimation techniques should be applied to real data.

The estimation techniques under consideration are introduced in Section 6.4.1. In Section 6.5 we perform a range of experiments to test the estimators under a variety of conditions and we draw conclusions in Section 6.6.

6.2 Stable random variables

The theory of stable processes evolved from the investigation of the characteristic function by Laplace and others in the 18th and 19th centuries. The characteristic function [80], [81] is the Fourier transform of a probability function and for stable distributions is given by

$$\phi(t) = \exp\{jat - |\gamma t|^\alpha [1 + j\beta \text{sign}(t)\omega(t, \alpha)]\}, \quad (6.1)$$

where,

$$\omega(t, \alpha) = \begin{cases} \tan(\alpha\pi/2) & \text{for } \alpha \neq 1, \\ (2/\pi) \log|t| & \text{for } \alpha = 1, \end{cases} \quad (6.2)$$

and,

$$\text{sign}(t) = \begin{cases} 1 & \text{for } t > 0, \\ 0 & \text{for } t = 0, \\ -1 & \text{for } t < 0. \end{cases} \quad (6.3)$$

6.2.1 Stable parameters

The parameters α , β , γ and a describe completely a stable distribution. In this section we introduce these parameters and discuss their effect on the distribution.

The characteristic exponent, α

The characteristic exponent controls the heaviness of the tails of the stable distribution and hence the impulsiveness of a stable process. α can take values in $[0, 2)$; a smaller value implies heavier tails and $\alpha = 2$ is the Gaussian case.

All stable distributions with $\alpha < 2$ possess infinite variance, i.e. the sample variance of a stable process never converges because large events occur often enough to prevent it.

Figure 6.1 demonstrates the effects of α on the tails of a stable distribution. Five symmetric stable distributions are plotted, all with $\beta = 0$, $a = 0$, $\gamma = 1$ but with $\alpha = 0.5, 0.7, 1.1, 1.7$ and 2.0 . The tails of the distributions with small α initially decay at a

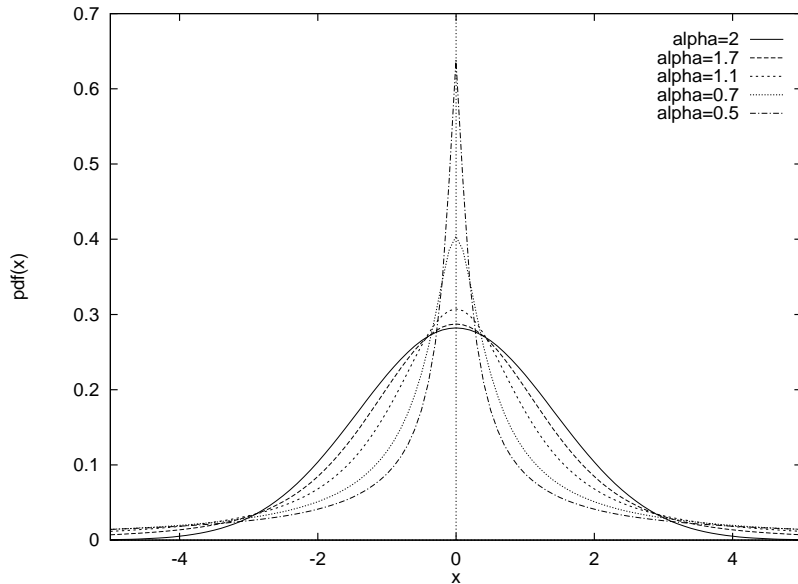


Figure 6.1: *Symmetric stable distributions with varying α . The point to note is that the distribution tails get heavier as $\alpha \rightarrow 0$.*

faster rate but by $|x| = 5$ it is the Gaussian case with the smallest $P(|x| = 5)$. Beyond $|x| = 5$ the Gaussian continues to drop away fastest. We can express the asymptotic rate of tail decay as a function of α ,

$$\lim_{t \rightarrow \infty} t^\alpha P(|x| > t) = \gamma C(\alpha). \quad (6.4)$$

Where $C(\alpha) > 0$ is a constant whose value depends on α . So every stable distribution with $\alpha < 2$ is, asymptotically, a heavy tailed distribution.

The index of skew, β

The index of skew controls the symmetry of the stable distributions and takes values in $(-1, +1)$. $\beta = 0$ implies a distribution is symmetric, $\beta > 0$ implies the distribution is skewed to the right of the location parameter and $\beta < 0$ implies skew to the left. Figure 6.2 plots five stable distributions all with $\alpha = 1.5$, $a = 0$ and $\gamma = 1$ but with $\beta = -1, -0.5, 0, 0.5$ and 1 .

One point of note is that the effect of β decreases as $\alpha \rightarrow 2$. This makes estimating β difficult (and somewhat superfluous) when α is close to 2.

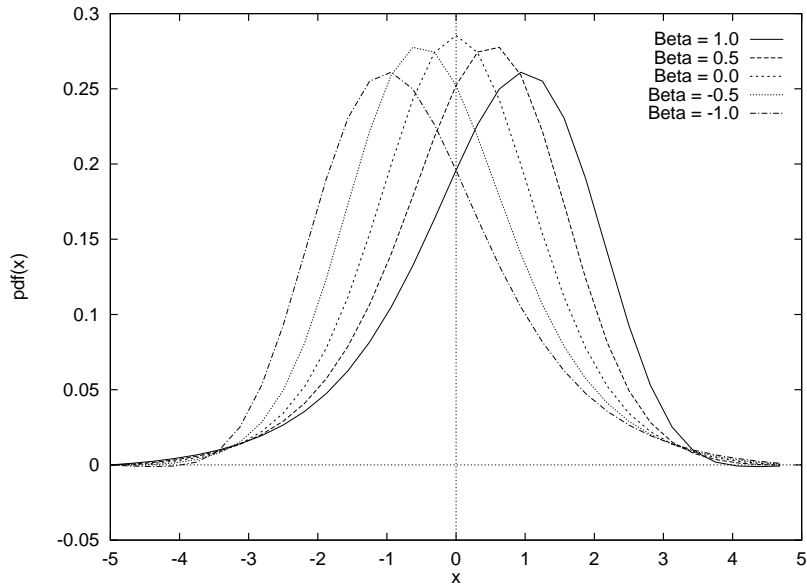


Figure 6.2: *Stable distributions with varying β ($\alpha = 1.5$).*

The dispersion parameter, γ

Since almost all stable distributions have infinite variance the term dispersion is used to refer to the spread of the distribution. γ can take values in $[0, \infty)$. When $\alpha = 2$ the distribution is Gaussian with variance equal to 2γ .

The location parameter, a

This parameter controls the location of the stable distribution and is comparable with the mean of the distribution. However since it is possible that a stable distribution has infinite mean we cannot say that a is equal to the mean. It can take values in $(-\infty, \infty)$.

6.2.2 The generalised central limit theorem

If X is an iid stable random variable then the limit of normalised sums of the form

$$S_n = \frac{(X_1 + \dots + X_n)}{a_n} - b_n \quad (6.5)$$

is also stable. If, in addition, X is a finite variance process then this reduces to the central limit theorem and the limit distribution is Gaussian.

6.2.3 Normalising stable distributions

We define $S_\alpha(\gamma, \beta, a)$ as a stable distribution with characteristic exponent equal to α , an index of skew equal to β , a dispersion parameter equal to γ and a location parameter equal to a . If a random variable X is drawn from $S_\alpha(\gamma, \beta, a)$ then

$$\frac{X - a}{\gamma^{\frac{1}{\alpha}}} \tag{6.6}$$

is drawn from $S_\alpha(1, \beta, 0)$. We use the term c to refer to $\gamma^{\frac{1}{\alpha}}$ and often use c instead of γ when estimating stable parameters.

6.2.4 Fractional lower order moments

In Figure 6.1 we saw how α controls the heaviness of the tails of a stable distribution which suggests that α can control the moments of a stable distribution. This is true and we can write

$$E(|X|^p) < \infty \text{ if } p < \alpha. \tag{6.7}$$

Since p must be less than 2 for all but the Gaussian case, we refer to $E(|X|^p)$ as the fractional lower order moments (FLOMs) of a stable distribution. The FLOM can be estimated empirically or if the distribution is symmetric

$$E(|X|^p) = C(p, \alpha) \gamma^{\frac{p}{\alpha}} \tag{6.8}$$

where

$$C(p, \alpha) = \frac{2^{p+1} \Gamma(\frac{p+1}{2}) \Gamma(\frac{-p}{\alpha})}{\alpha \sqrt{\pi} \Gamma(\frac{-p}{2})} \tag{6.9}$$

This result was first proved by Zolotarev [82] and can be used to produce stable parameter estimation techniques.

6.3 Generating stable random variables

In Section 6.4 we compare the performance of several stable parameter estimation techniques and in Chapter 7 we attempt to model teletraffic traces with self-similar

stable processes. In both cases iid stable random variables are required and these are generated using a method developed by Chambers, Mallows and Stuck [83].

6.4 The estimation of stable parameters

In this section we introduce some of the common techniques that have been developed to estimate the parameters of a stable distribution for empirical data.

6.4.1 Estimation techniques

6.4.1.1 Quantile based techniques

The p th quantile of a set of data is defined as the value x_p that satisfies $f(x_p) = p$. Therefore for a dataset of size N , \hat{x}_f refers to the $(p)(N + 1)$ st order statistic of the set, which we can use as an estimate of x_p . So by estimating certain quantiles from the data it is possible to estimate stable parameters using the tables of quantiles for real stable distributions.

The first quantile estimators were suggested by Fama and Roll [84]. They proposed the following as an estimate for the scale parameter c when the distribution was symmetric and $\alpha \geq 1$

$$\frac{1}{2(0.827)} (\hat{x}_{0.72} - \hat{x}_{0.28}). \quad (6.10)$$

In [85] they showed this estimator to have an asymptotic bias of less than 4% and that it produced estimates that were asymptotically normally distributed.

Fama and Roll also proposed an estimator for α based on the difference between two extreme quantiles, e.g.

$$\hat{Z}_p = \frac{\hat{x}_p - \hat{x}_{1-p}}{2\hat{c}} \text{ and } \hat{\alpha} = G(p, \hat{Z}_p). \quad (6.11)$$

$G(p, \hat{Z}_p)$ is some mapping function and p is chosen to be suitably close to 1.

There are some problems with Fama and Roll estimators. One is that they are restricted to the symmetric case. A second is that α is confined to being ≥ 1 . Even when the data

does possess a distribution which lies within the valid region, the parameter estimates tend to be biased.

McCulloch proposed another quantile based technique [86] which:

1. Relaxes the constraint on β .
2. Permits estimation in the range $\alpha \in (0.5, 2)$.
3. Estimates β and a .
4. Removes the asymptotic bias on the estimate of c .

The four stable parameters are estimated as follows:

$$\hat{v}_\alpha = \frac{\hat{x}_{0.95} - \hat{x}_{0.05}}{\hat{x}_{0.75} - \hat{x}_{0.25}} = \Phi_1(\alpha, \beta), \quad (6.12)$$

$$\hat{v}_\beta = \frac{\hat{x}_{0.95} + \hat{x}_{0.05} - 2\hat{x}_{0.50}}{\hat{x}_{0.95} - \hat{x}_{0.05}} = \Phi_2(\alpha, \beta). \quad (6.13)$$

The Φ functions can be inverted to yield Ψ_1 and Ψ_2 and

$$\hat{\alpha} = \Psi_1(\hat{v}_\alpha, \hat{v}_\beta), \quad (6.14)$$

$$\hat{\beta} = \Psi_2(\hat{v}_\alpha, \hat{v}_\beta). \quad (6.15)$$

McCulloch supplies tables of the values of these functions for determining α and β in [86].

The scale parameter can be estimated by

$$\hat{c} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{\Phi_3(\hat{\alpha}, \hat{\beta})}. \quad (6.16)$$

If we compare this with the estimator previously proposed by Fama and Roll (equation (6.10)) we see that $\frac{1}{2(0.827)}$ has been substituted by $\Phi_3()$.

Finally the location parameter a can be determined using

$$\hat{\zeta} = \hat{x}_{0.50} + \hat{c}\Phi_5(\hat{\alpha}, \hat{\beta}), \quad (6.17)$$

where $\zeta = a + \beta c \tan(\frac{\pi\alpha}{2})$ if $\alpha \neq 1$ and $\zeta = a$ otherwise. The transform from a to ζ is performed to avoid problems caused by the discontinuity of the focus of stability at $\alpha = 1$.

6.4.1.2 Characteristic function based techniques

It is known that the characteristic function (equation (6.1)) uniquely defines a stable distribution. Conversely it should be possible to estimate stable parameters from the empirical characteristic function. In effect, we use empirical data to estimate $\phi(t)$.

$$\hat{\phi}(t) = \frac{1}{N} \sum_{j=0}^{N-1} \exp ix(j)t \quad (6.18)$$

Techniques based on the characteristic function were first developed by Press [87]; others were developed by Paulson, Holcomb and Leitch [88], Koutrouvelis [89], [90] and Kogon [91]. It is Kogon's technique that we concentrate on, since it has been shown to outperform other characteristic function based techniques over much of the solution surface [92]. Kogon's technique is also simpler in computational terms than the other characteristic function techniques making it more attractive with regards to the real-time estimation of stable parameters.

Kogon's method uses linear regression of the empirical estimates of the characteristic function $\hat{\phi}(t)$. By noting that

$$\ln\{-\text{Re}[\ln \hat{\phi}(t)]\} = \alpha \ln |t| + \alpha \ln c, \quad (6.19)$$

it is possible to find $\hat{\alpha}$. In a similar manner $\hat{\beta}$ can be determined by finding the slope of

$$\text{Im}[\ln \hat{\phi}(t)] \text{ vs } |\hat{c}t|^{\hat{\alpha}} \frac{t}{|t|} \tan\left(\frac{\pi\hat{\alpha}}{2}\right). \quad (6.20)$$

To illustrate the performance of this estimation technique we generated 1000 points from $S_{1.5}(0.5, 1, 0)$ and generated the plots suggested above in Figure 6.3. The stable parameters were estimated to be $\hat{\alpha} = 1.543$ and $\hat{\beta} = 0.384$.

One disadvantage of this technique is that the choice of t depends on the values of c and a . This is because the scale and location parameters can render the characteristic function very flat and if t is not chosen correctly then the estimates for α and β can

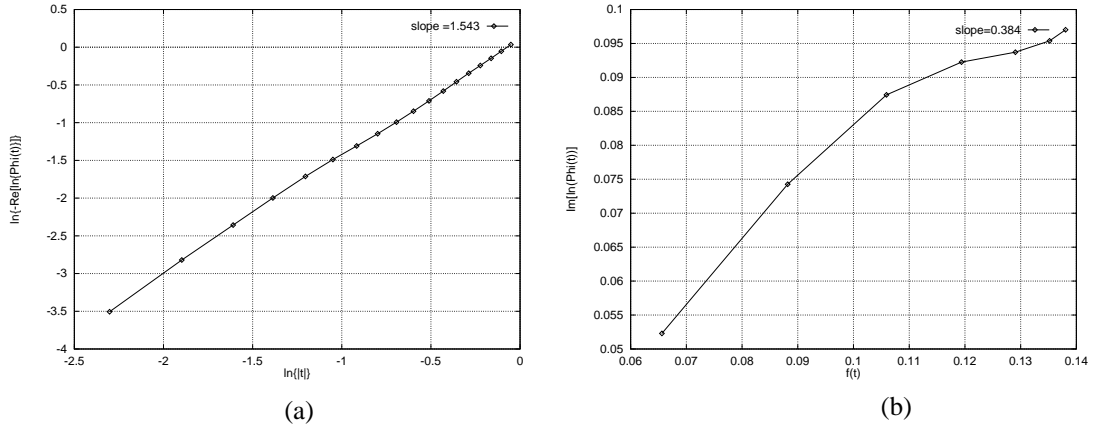


Figure 6.3: The plots for calculating $\hat{\alpha}$ and $\hat{\beta}$ using Kogon's technique.

be very unstable [92]. To overcome this Kogon suggests normalising the data using McCulloch's quantile method to determine \hat{c} and \hat{a} , i.e.

$$x \rightarrow \frac{x - \hat{a}}{\hat{c}}. \quad (6.21)$$

Kogon then recommends using ten equi-spaced points over the interval $t \in [0.1, 1.0]$.

6.4.1.3 Fractional Lower Order Moment based techniques

FLOMS were introduced as a property of stable random variables in Section 6.2.4. Like the characteristic function their values are unique for a given set of stable parameters and hence their empirical approximation can be used for estimation purposes. To recap; the FLOM is defined as $E(|x|^p)$ and is finite for $0 \leq p < \alpha \leq 2$. The FLOM based methods under consideration are the log FLOM method of Ma and Nikias [93] and the extreme order method of Tsihrintzis [94], [95].

The log FLOM method uses the transform $x \rightarrow \log |x|$ because we can then rewrite the FLOM definition as $E(e^{px})$ where $(-1 < p < \alpha)$. The stable parameters α and γ can be estimated by solving

$$E(x) = C_e \left(\frac{1}{\alpha} - 1 \right) + \frac{1}{\alpha} \log \gamma, \quad (6.22)$$

$$\text{VAR}(x) = \frac{\pi^2}{6} \left(\frac{1}{\alpha^2} + \frac{1}{2} \right). \quad (6.23)$$

Where C_e is the Euler constant (0.57721566....).

The second method uses asymptotic extreme value theory (AEVT) to estimate α and a FLOM method to estimate γ . This technique was developed by Tsihrintzis and Nikias [94], [95]. To estimate α the data is divided into L equi-lengthed blocks of length $K = N/L$. In each block the maximum and minimum values are found (\overline{X}_l and \underline{X}_l). Then let $\overline{x}_l = \log \overline{X}_l$ and $\underline{x}_l = \log \underline{X}_l$. If this is done for all L define

$$\overline{s} = \sqrt{\frac{1}{L-1} \sum_{l=1}^L (\overline{x}_l - \overline{x})^2} \quad \overline{x} = \frac{1}{L} \sum_{l=1}^L \overline{x}_l, \quad (6.24)$$

$$\underline{s} = \sqrt{\frac{1}{L-1} \sum_{l=1}^L (\underline{x}_l - \underline{x})^2} \quad \underline{x} = \frac{1}{L} \sum_{l=1}^L \underline{x}_l, \quad (6.25)$$

then

$$\hat{\alpha} = \frac{\pi}{2\sqrt{6}} \left(\frac{1}{\overline{s}} + \frac{1}{\underline{s}} \right). \quad (6.26)$$

The location parameter a is simply estimated as the median of the dataset,

$$a = \text{median}(X_1, X_2, \dots, X_N). \quad (6.27)$$

The dispersion γ is estimated using a FLOM based technique. Tsihrintzis noted that

$$\hat{\gamma} = \left(\frac{\frac{1}{N} \sum_{k=1}^N |X_k - \hat{a}|^p}{C(p, \hat{\alpha})} \right)^{\frac{\hat{\alpha}}{p}}, \quad (6.28)$$

where

$$C(p, \hat{\alpha}) = \frac{1}{\cos(\frac{\pi}{2}p)} \frac{\Gamma(1 - \frac{p}{\hat{\alpha}})}{\Gamma(1 - p)}. \quad (6.29)$$

So by choosing a suitable value of p (ensuring it is less than α) an estimate for γ can be calculated.

6.4.2 Summary

Table 6.1 summarises the permissible range of parameter values for each of the estimation methods. Parameters that can be estimated by a given technique are marked with a *.

Name	α	β	γ	a
McCulloch	$(0.6, 2)^*$	$(-1, 1)^*$	$(0, \infty)^*$	$(-\infty, \infty)^*$
Kogon	$(0, 2)^*$	$(-1, 1)^*$	$(1)^*$	(0)
Ma	$(0, 2)^*$	(0)	$(0, \infty)^*$	(0)
Tsihrintzis	$(0, 2)^*$	(0)	$(0, \infty)^*$	$(-\infty, \infty)^*$

Table 6.1: Permissible ranges of estimation.

6.5 Estimating parameters for stable random variables

In this section the accuracy of the estimation techniques is tested for the case when the data is drawn from a stable distribution with known parameter values. In all the cases in this section the estimation was performed for 1000 independent sets of data each of length 1000. The mean of the 1000 estimates was found, as was the 95% confidence interval of the estimate. We generated the stable random variables using a method based on the work of Chambers, Mallows and Stuck [83] and reproduced in [67].

6.5.1 Procedure & results

6.5.1.1 Experiment 1: Estimating α when $\beta = 0$.

In this experiment sets of symmetric iid stable random variables were generated with $\alpha = 0.1$ (0.1) 2.0¹, $c = 1$ and $a = 0$. The estimates of α are plotted against α for all four techniques in Figure 6.4.

6.5.1.2 Experiment 2: Estimating α when $\beta \neq 0$.

The first part of this experiment was concerned with estimating α and β using McCulloch's and Kogon's techniques. Figures 6.5 - 6.8 plot $\hat{\alpha}$ and $\hat{\beta}$ against α when $\beta = -1, -0.5, 0.5$ and 1 respectively.

In the second part the effect of misapplying Ma's and Tsihrintzis' methods to asymmetric distributions is investigated. We estimated α for skewed stable distributions (i.e. $\beta = -0.5$ and 0.5). $\hat{\alpha}$ is plotted against $\alpha = 0.1$ (0.1) 2.0 and the results are presented in Figure 6.9.

¹a (b) c denotes from a to c in increments of size b.

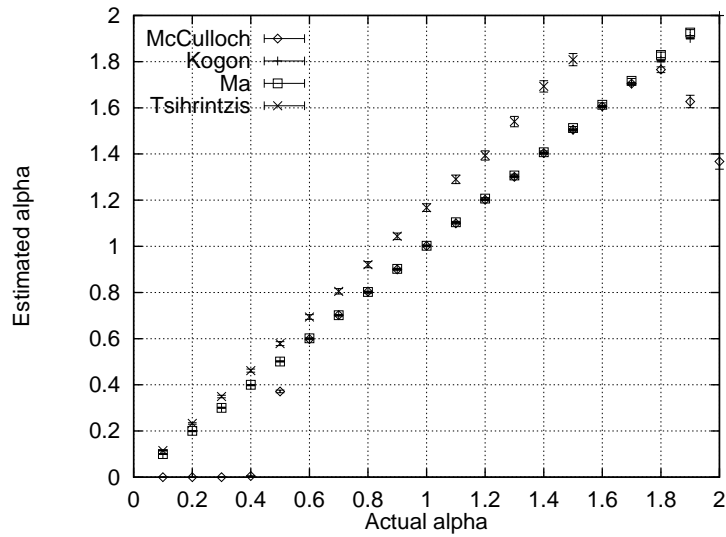


Figure 6.4: Results for $\hat{\alpha}$ when $\beta = 0$.

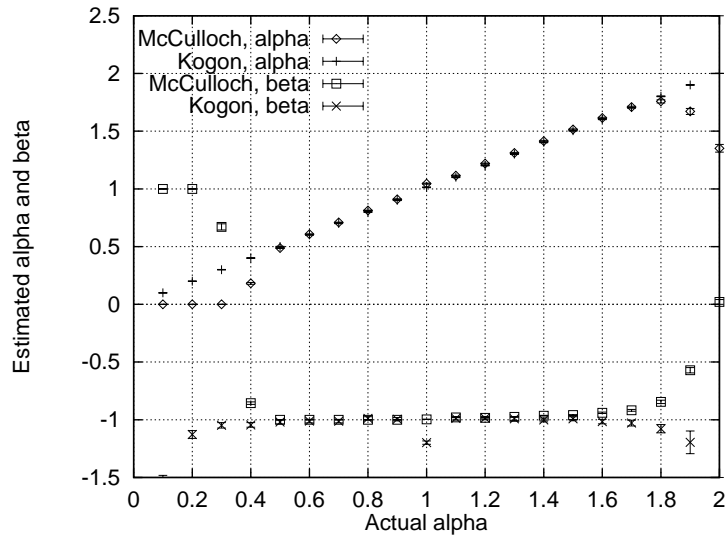


Figure 6.5: Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = -1$.

6.5.1.3 Experiment 3: Estimating c

If data is generated from $S_\alpha(0, 1, 0)$ and scaled with the mapping $X \rightarrow X * k, k > 0$ then the expected value of scale parameter estimate \hat{c} ($c = \gamma^{\frac{1}{\alpha}}$) is k . The cases when $k=0.1$ and $k=10$ are considered and \hat{c} vs α is plotted for all four techniques (Figures 6.10 and 6.11).

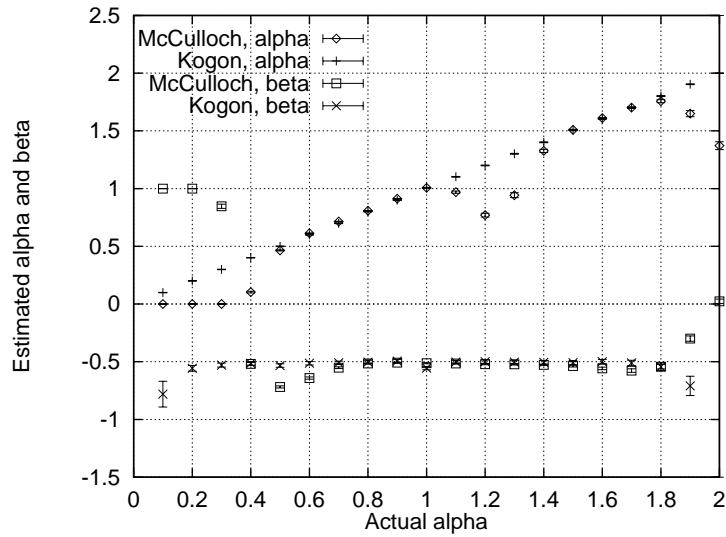


Figure 6.6: Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = -0.5$.

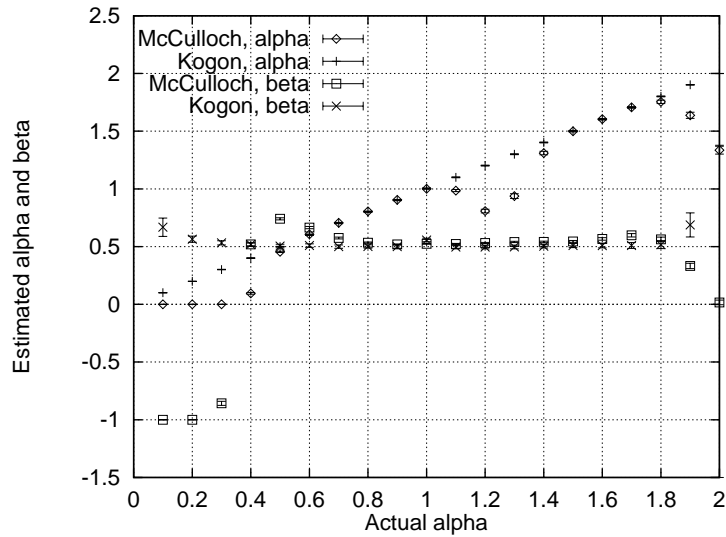


Figure 6.7: Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = 0.5$.

6.5.2 Discussion

In [86] McCulloch points out that his technique is only valid in the range $0.6 < \alpha < 2$. The results in Figures 6.4-6.8, 6.10 and 6.11 would seem to support this. In all the above cases the estimates perform poorly for $\alpha < 0.6$. However the performance of this estimator also seems to degrade for $\alpha > 1.8$. The estimates of α and β are poorer above this level although the estimates for c appear to fare better.

Kogon's estimate of α is consistently good over the range $\alpha \in (0.1, 2)$, both in symmetric and asymmetric cases. The estimates of β are also good although they do tend to degrade when $\alpha < 0.3$ and as $\alpha \rightarrow 2$. However the degradation of $\hat{\beta}$ as $\alpha \rightarrow 2$ is

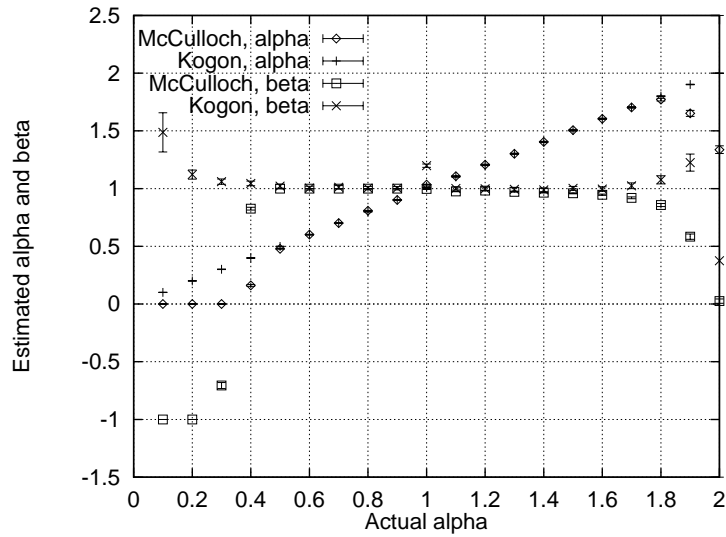


Figure 6.8: Results for $\hat{\alpha}$ and $\hat{\beta}$ when $\beta = 1$.

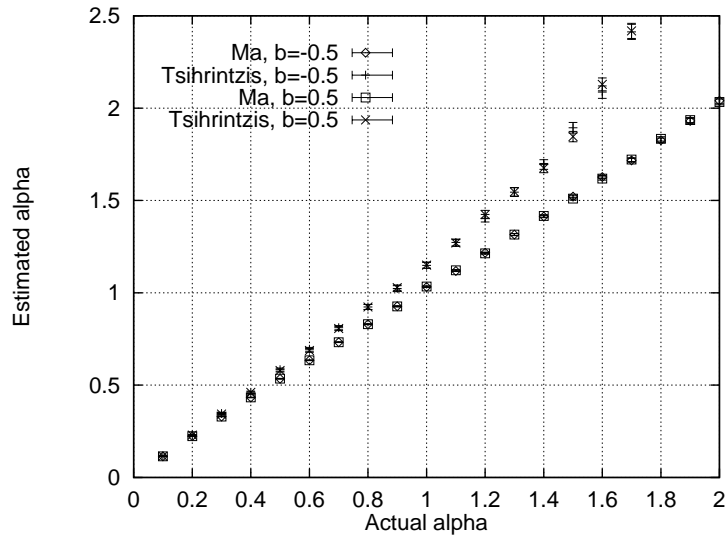


Figure 6.9: Results for $\hat{\alpha}$ when the FLOM based techniques are applied to a skewed distribution.

expected due to the fact that the characteristic function becomes much less sensitive to β . The estimates of c are reasonable when $k = 0.1$ but poor when $k = 10$. This is due to the flattening effect larger c have on the characteristic function. This result would suggest it is sensible to over-scale data, thereby ensuring $c \leq 1$, before using Kogon's technique.

Ma's technique provides very good estimates of α when $\beta = 0$. These estimates are as good as for Kogon's technique over most of the range $\alpha \in (0.1, 2)$ with a very slight degradation as $\alpha \rightarrow 2$. In addition the $\hat{\alpha}$ when $\beta \neq 0$ are just as good. This is a surprising result as theory would suggest that Ma's technique is only applicable to symmetric distributions. In reality it would seem that it is not as sensitive to the value

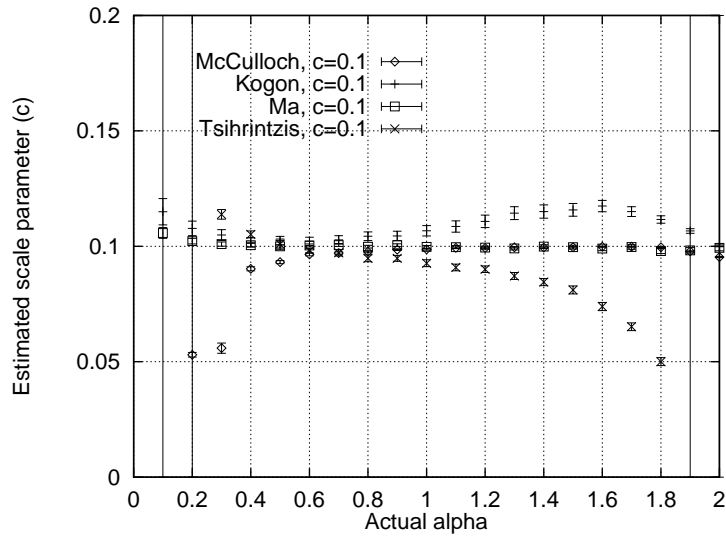


Figure 6.10: Results for \hat{c} when $k=0.1$.

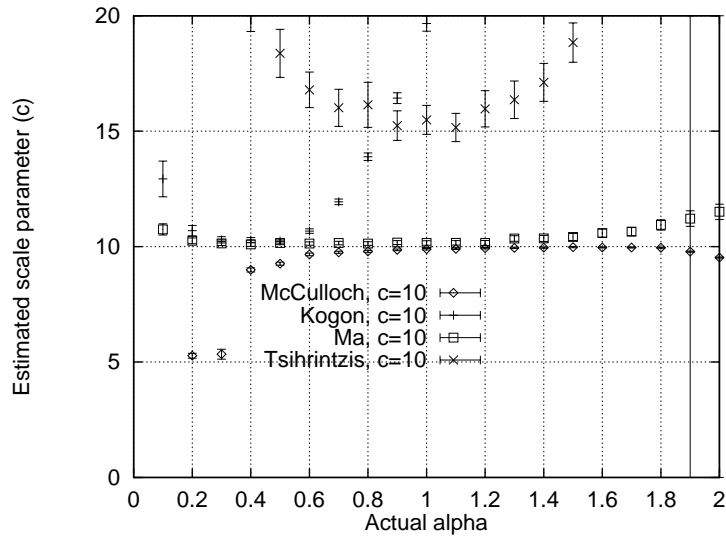


Figure 6.11: Results for \hat{c} when $k=10$.

of β as was assumed. The estimates of c are good for both $k = 0.1$ and $k = 10$. In fact this technique performs better at estimating c than any of the others when $c = 0.1$ and is second best (after McCulloch's) when $c = 10$.

Tsihrintzis' technique also produces good estimates of α when $\beta = 0$, although there does appear to be a slight over-bias that increases with α . The performance degrades dramatically as $\alpha \rightarrow 2$ and does not change significantly when $\beta \neq 0$. However the estimates of c are poor for both $k = 0.1$ and $k = 10$.

With regard to direct comparison between the techniques it would appear there is no overall best estimation technique. Unfortunately the choice of estimator depends on

the actual values of the stable parameters, which are of course unknown. However it is possible to determine that McCulloch's technique is a good all round estimator provided $\alpha \geq 0.6$. Performance falls slightly as $\alpha \rightarrow 2$ so if a high value of $\hat{\alpha}$ is obtained then Kogon's or Ma's technique can be used to improve the estimate. In addition Kogon's technique gives good estimates of α and β as long as $c \leq 1$. This emphasises the point made in [91] that data should be pre-scaled before applying this technique. Ma's estimates of α and c are also very good and $\hat{\alpha}$ is less sensitive to β than expected.

6.6 Conclusions

The focus of this chapter has been the stable distribution which we have chosen as a heavy tailed distribution with potential for teletraffic modelling. The four parameters that characterise a stable distribution were introduced and their effect on the distribution were discussed. In addition the relevant properties of stable distribution distributions were explained.

For stable distributions to be useful for CAC on ATM networks we need a fast, accurate estimation technique that can be applied to teletraffic data. We compared four published techniques that rely on different aspects of a stable distribution. In doing this we hoped to identify which technique was best suited for CAC but in reality we could not find one technique that consistently outperformed all the others.

Whilst performing the comparison between these techniques we were able to draw some conclusions about the estimation methods. These included the robustness of the techniques of Ma and Tsihrintzis to asymmetric distributions and the importance of scaling before applying Kogon's technique.

The analysis and modelling of teletraffic with stable distributions and models

7.1 Introduction

In this chapter we begin to develop a framework for applying stable distributions to the characterisation of teletraffic. If this is to be done successfully then we must be able to estimate stable parameters for real data. In Section 7.2 we apply the four estimation techniques discussed in Chapter 6 to some of the datasets discussed in this thesis. The results of this are discussed in Sections 7.2.3.1 and 7.2.4.1.

Next we apply three self-similar stable models to the same datasets as mentioned above. Two of these models are adaptations of the Norros teletraffic model; the first uses symmetric stable innovations whilst the second uses totally skewed stable distributions. The third model is the marginal distribution mapping model introduced in Section 3.5.3. The model outputs are compared with the original data and we discuss the results in Section 7.3.5.1.

If stable distributions are to be used for CAC schemes then we need some means of linking stable parameters with the performance of data across a network. Such a link would be closed form expressions for the effective bandwidths of the stable models. In Section 7.4 we consider the effective bandwidths of stable distributions and the adapted Norros models.

We finish by presenting some studies of an on-line traffic characterisation scheme that uses stable distributions. This work attempts to correlate the estimate of α for the teletraffic to its performance through a buffered switch.

7.2 Estimating parameters for real data

In the previous chapter we compared four estimation techniques by applying them to traces of stable random variables. Now we turn our attention to estimating stable parameters for real data.

One problem is determining whether the estimated parameters produce a distribution which matches the data well. To test this we perform a goodness of fit test between the density functions of the data and the density functions of the estimated stable distributions. We have already mentioned that no closed form expression for the pdf exists in most cases of stable distribution but we can determine an approximation to it by either calculating the inverse Fourier transform of the characteristic equation (Section 2.3 in [96]) or by generating an empirical approximation from stable random variables.

7.2.1 The teletraffic data

7.2.1.1 VBR video data

The VBR video datasets considered are the JPEG and MPEG encoding of Star Wars as discussed in Section 4.2. We look at both the original MPEG trace and the GOP trace to see how aggregation of the data affects the stable estimates.

7.2.1.2 The Ethernet data

The Ethernet data is the W^1 dataset constructed from the pAug.TL file and first discussed in Section 5.2. We also consider the differenced version of this dataset to see if this improved the consistency of the estimates. The statistics of the differenced dataset were

$$\begin{aligned} N &= 3143 & \sigma &= 79018 & \nabla &= -377870 \\ \mu &= 24.3 & \Delta &= 395970 \end{aligned}$$

7.2.2 Data transform

The transform used in Section 7.2.3 is identical to that in Section 5.3.

7.2.3 The estimation of the assumed Gaussian innovations

In this experiment the mean was removed from the data and it was scaled by the estimate of the standard deviation. As was noted in Section 5.3 the resultant data should conform to the $N(0,1)$ distribution (i.e. a normal distribution with zero mean and unit variance) if the Gaussian assumption holds.

The next step was to estimate stable parameters for these transformed datasets. In Section 6.5.2 several important points about the performance of the estimation techniques were made. One point is that Kogon's technique gives good estimates of α and β but only when $c \leq 1$. McCulloch's technique was applied first and then the data was scaled accordingly. Since evidence was found to suggest that Ma's and Tsihrintzis' techniques work well even when $\beta \neq 0$ they were applied to the data. Finally Kogon's technique was applied and all the estimated values were recorded in Tables 7.1 - 7.5.

Finally the cumulative density functions (CDFs) of the datasets and the estimated stable densities were compared. This was done by generating 100 000 stable random variables with the estimated stable parameters and generating an approximate CDF from that. Although not an exact match to the true stable CDF it was sufficient for the comparative purposes of this experiment. The results are given in Figures 7.1 - 7.5.

7.2.3.1 Results & discussion

For the JPEG estimates, the $\hat{\alpha}$ was close to 2 for every technique (except Tsihrintzis) which suggests that the $\hat{\beta}$ are very unstable. This explains the large (and impossible) value for $\hat{\beta}$ from Kogon's estimator. The NaN (Not a Number) results for the Tsihrintzis method result from the fact that the method must take the square-root of a value which in this case was negative. Since no similar result occurred when testing with stable random variables this may suggest the data in question is not truly stable.

Figure 7.1 compares the CDFs and shows that only the McCulloch estimate produces a good match with the real data whereas the Kogon and Ma estimates are slightly

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.898	1.000	0.601	0.010
Ma	1.985	-	0.462	-
Tsihrintzis	NaN	-	NaN	-0.053
Kogon	1.944	4.561	1.018	-

Table 7.1: Results for the JPEG video data.

inferior.

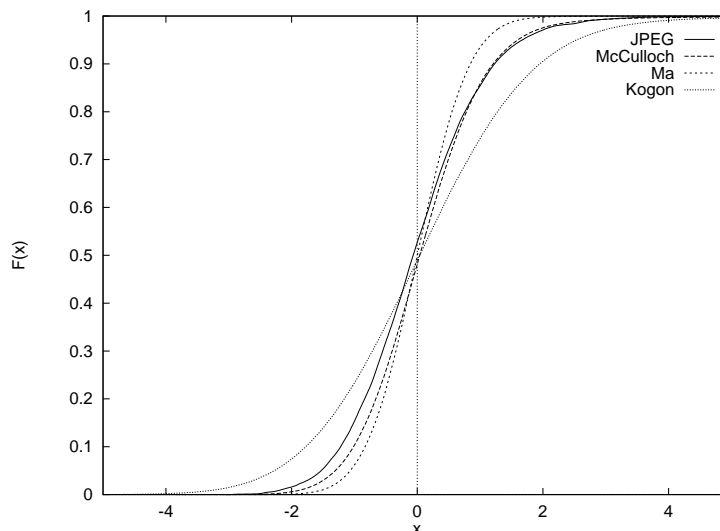


Figure 7.1: The CDF of the transformed JPEG data compared to the CDF of stable random variables generated using the estimates from Table 7.1.

For the original MPEG data only McCulloch's technique produced estimates inside the valid parameter space. Both Ma's and Tsihrintzis' technique estimated $\alpha > 2$ and Kogon's technique estimated $\beta > 1$

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.121	1.000	0.267	0.845
Ma	3.984	-	2.782	-
Tsihrintzis	13.74	-	0.000	-0.414
Kogon	1.784	-3.453	0.553	-

Table 7.2: Results for the MPEG video data.

The CDF estimated from the MPEG data did not match the data as well as for the JPEG data. It is worth noting that the CDF for the MPEG data is very long tailed and asymmetric.

The MPEG GOP estimates are similar to those for the MPEG data in that Ma's and Tsihrintzis' technique estimate $\alpha > 2$ and Kogon's technique estimate $\beta < 1$. Again it is only McCulloch's technique that produces a valid set of estimates.

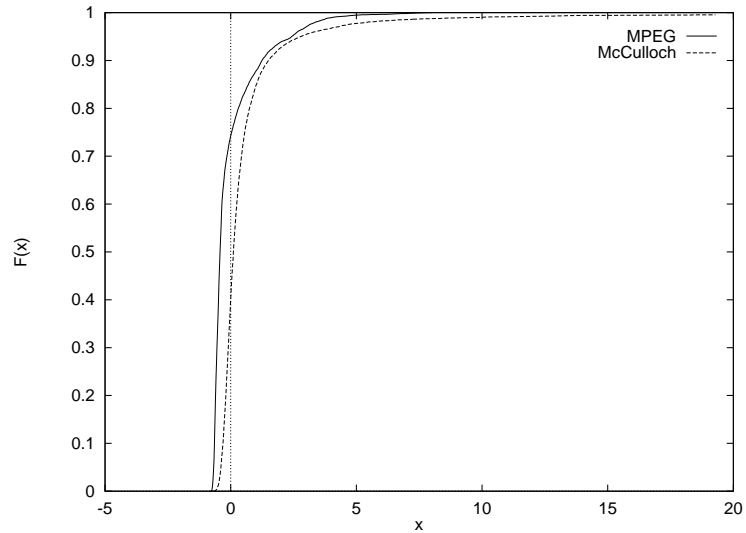


Figure 7.2: *The CDF of the transformed MPEG data compared to the CDF of stable random variables generated using the estimates from Table 7.2.*

An interesting point is that one test for a stable distribution is to estimate α for aggregated version of a dataset [97], [98]. This is because the stability property states that the aggregated data should have the same α as the original. However for the MPEG data the estimates were 1.121 for the original data and 1.958 for the for the GOP data. This would seem to suggest that the data is not stable. However McCulloch notes that the stability property only applies to iid stable random variables [77]. He also cites several studies that suggest that aggregating correlated stable data raises the value of $\hat{\alpha}$, [99], [100].

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.958	1.000	0.655	-0.216
Ma	2.279	-	0.4815	-
Tsihrintzis	3.362	-	0.605	-0.242
Kogon	1.925	-5.297	0.651	-

Table 7.3: Results for the MPEG GOP video data.

Figure 7.3 plots the CDFs of the data and the stable estimate. The real data has tails that decay more slowly than the McCulloch estimate (which has a $\hat{\alpha}$ very close to 2).

In the case of the original Ethernet data even the McCulloch estimate is not sensible. Although the estimate is within the valid range, the estimate of α is less than 1. This implies the Ethernet data has an infinite mean.

The CDFs are plotted for the Ma estimates and the Kogon estimates (although we use $\hat{\beta}$ from the McCulloch estimates). The match with the real data is not good.

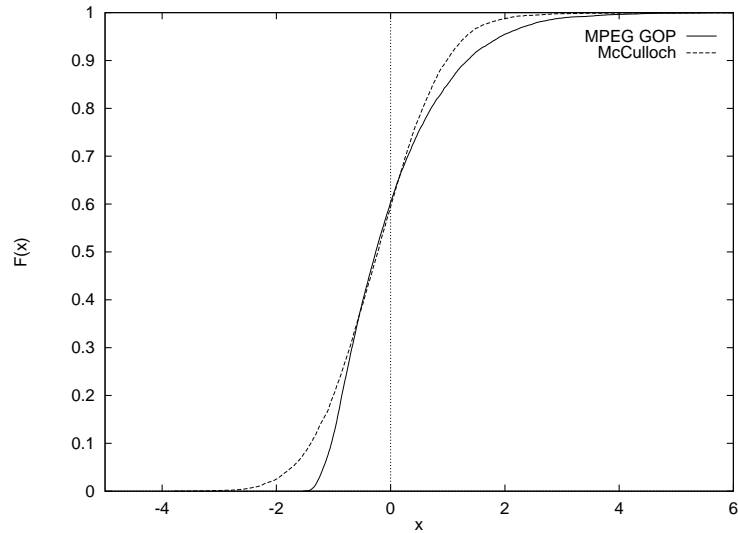


Figure 7.3: The CDF of the transformed MPEG GOP data compared to the CDF of stable random variables generated using the estimates from Table 7.3.

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	0.805	0.328	0.505	-0.067
Ma	1.279	-	0.410	-
Tsihrintzis	NaN	-	NaN	0.531
Kogon	1.946	30.27	0.626	-

Table 7.4: Results for the original Ethernet activity data.

It is interesting to note that differencing the Ethernet data had little effect on the McCulloch estimates. However the other techniques' estimates did change, though not necessarily for the better. This would suggest that there is little advantage in differencing the data.

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	0.825	0.328	0.539	-0.870
Ma	1.979	-	0.489	-
Tsihrintzis	5.397	-	-	-0.232
Kogon	1.769	-1.765	1.095	-

Table 7.5: Results for the differenced Ethernet activity data.

The CDFs are compared in Figure 7.5 and it is possible to see that although the slope of the Ma estimate matches that of the data, there is a considerable difference in location.

7.2.4 The estimation of the dataset pdfs

In the previous section the stable parameters for the innovation distribution of the Norros model were estimated. In this section a direct estimate is made on the datasets

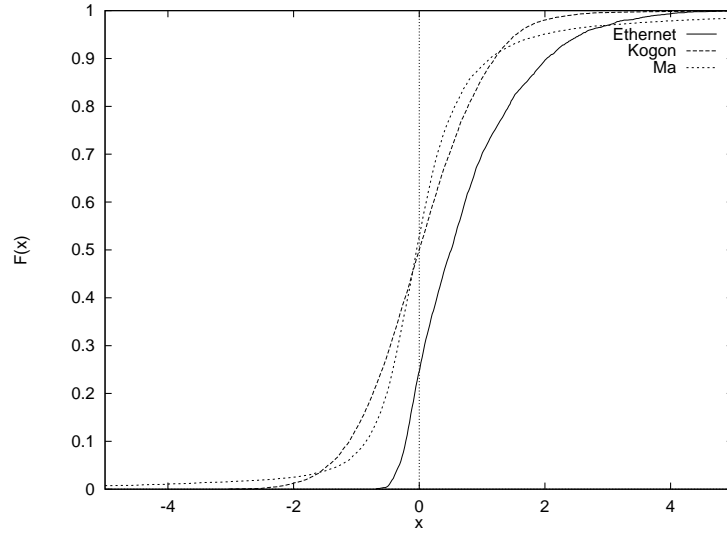


Figure 7.4: *The CDF of the transformed Ethernet data compared to the CDF of stable random variables generated using the estimates from Table 7.4.*

themselves. This is done to permit marginal distribution mapping modelling (Section 3.5.3) where an arrival process $A[i]$ is generated from fGn using a mapping between $N(0, 1)$ and a stable CDF ($F_{S_\alpha(\beta, \gamma, a)}(X)$), matched to the data,

$$A[i] = F_{S_\alpha(\beta, \gamma, a)}^{-1}(F_{N(0,1)}(X_H[i])). \quad (7.1)$$

The estimation procedure is identical to that in Section 7.2.3 but no data transform takes place. The stable parameter estimates are recorded in Tables 7.6 - 7.10 and the real and estimated pdfs are compared in Figures 7.6 - 7.9. The stable pdfs were generated from the estimates using the same technique as the previous section.

7.2.4.1 Results & discussion

The stable estimates for the four techniques for the JPEG data are recorded in Table 7.6. Again Tsihrintzis' estimates were unobtainable and it is worth noting that McCulloch's estimates of α and β are unchanged by scaling.

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.898	1.000	4130	27856
Ma	1.985	-	1.052	-
Tsihrintzis	NaN	-	NaN	-0.008
Kogon	1.944	4.338	1.018	-

Table 7.6: Results for JPEG video data.

The JPEG estimates behave in a similar manner to those in Section 7.2.3.1. In this

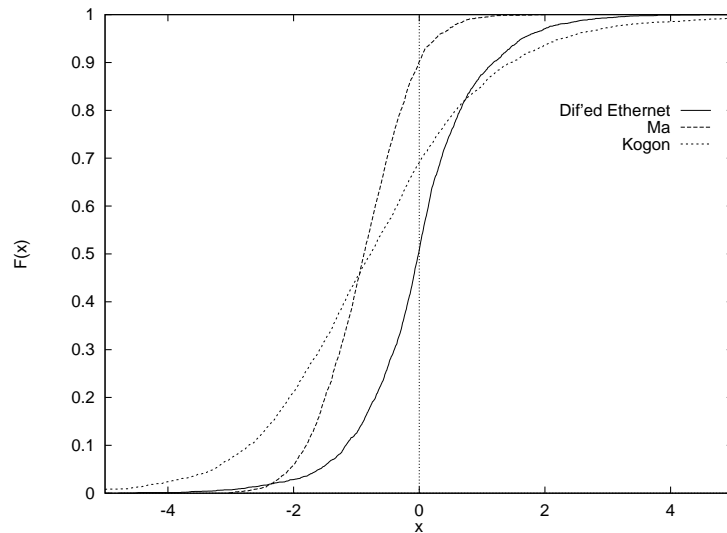


Figure 7.5: *The CDF of the transformed differenced Ethernet data compared to the CDF of stable random variables generated using the estimates from Table 7.5.*

case Ma's, Kogon's and McCulloch's estimates all provide a good match for the pdf of the real data (see Figure 7.6).

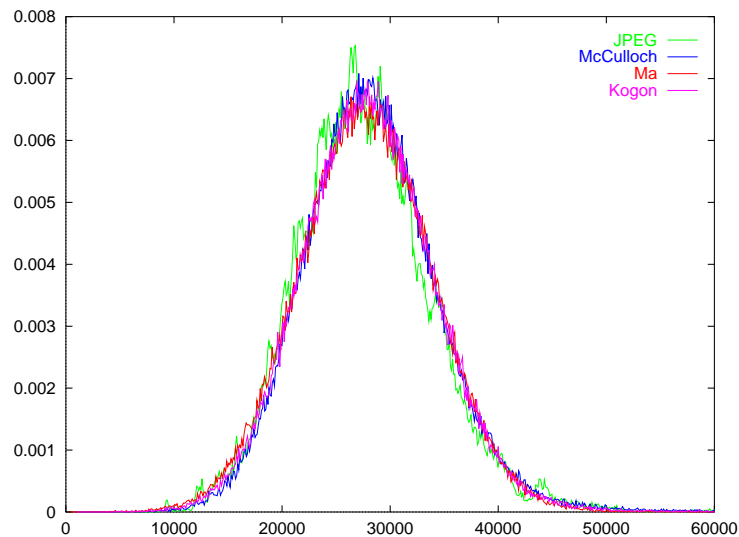


Figure 7.6: *The real and estimated stable pdfs for the JPEG data.*

The estimates for the MPEG video data did not match the data as well as in the JPEG case. Again it is only McCulloch's technique that produces valid parameter estimates but when these are used to generate a pdf we see there are differences in shape and location when compared to the real data (Figure 7.7).

Only one technique produced valid parameters for the MPEG GOP data. Both Ma's and Tsihrintzis' techniques failed to produce any estimates at all, suggesting that the data may be removed from a true stable distribution.

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.121	1.000	4862.6	31017.8
Ma	NaN	-	NaN	-
Tsihrintzis	26.35	-	NaN	-4.712
Kogon	0.913	5.993	0.752	-

Table 7.7: Results for the MPEG video data.

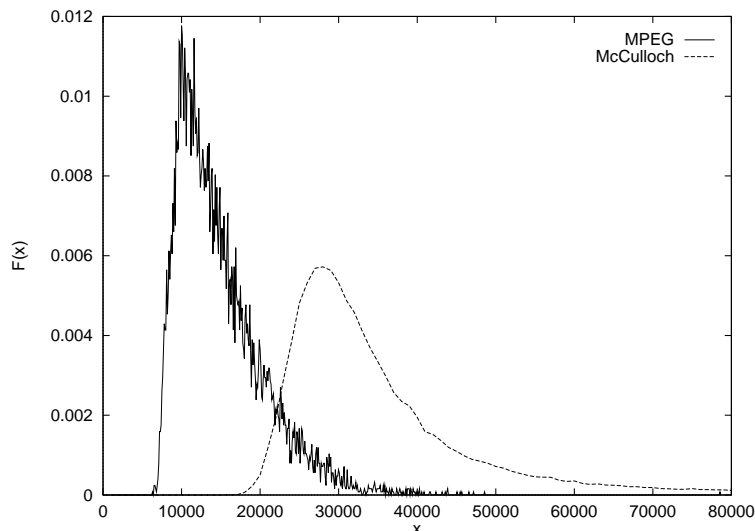


Figure 7.7: The real and estimated stable pdfs for the MPEG data.

The estimated pdf for the MPEG GOP matched the real data better than the MPEG data estimate did. This might mean that aggregation improves the consistency of the estimates.

All the Ethernet data estimates were unrealistic since both Ma's and Tsihrintzis' techniques failed, Kogon's estimated $\beta > 1$ and McCulloch's estimated $\alpha < 1$. The last of these is a valid value but implies that the Ethernet data has an infinite mean.

The estimates for the differenced Ethernet data seem a bit more consistent than the original Ethernet data. Aside from Tsihrintzis' estimate, the other $\hat{\alpha}$ are in reasonably close agreement. The $\hat{\beta}$ are not so consistent but three of the techniques give valid estimates. This would suggest that differencing the data may improve the consistency of the estimates.

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.958	1.000	3611.73	13769.2
Ma	NaN	-	NaN	-
Tsihrintzis	NaN	-	NaN	13627.0
Kogon	1.846	4.038	0.937	-

Table 7.8: Results for the MPEG GOP video data.

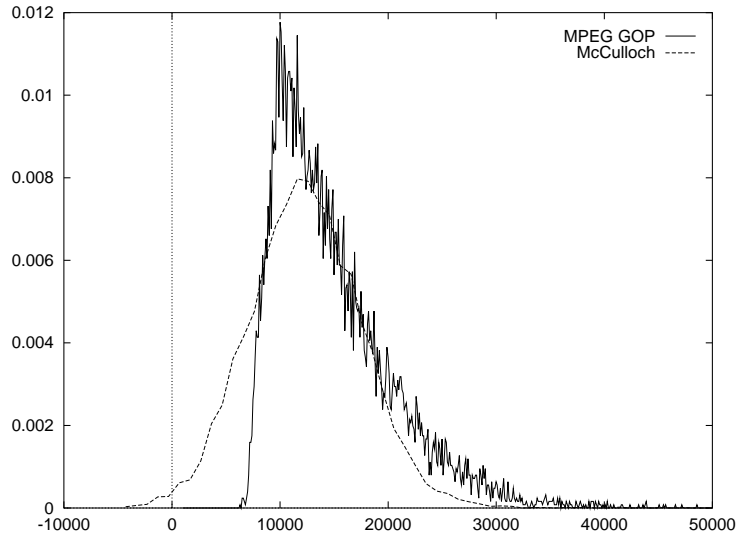


Figure 7.8: *The real and estimated stable pdfs for the MPEG GOP data.*

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	0.805	0.328	46097	63738.3
Ma	NaN	-	NaN	-
Tsihrintzis	NaN	-	NaN	118314
Kogon	1.772	-1.788	1.098	-

Table 7.9: Results for the Ethernet activity data.

Even though the distribution was again symmetric in appearance (see Figure 7.9), the $\hat{\alpha}$ were significantly less than 2 in all cases. This suggests that the distribution is more impulsive than for the other cases. It also means that $\hat{\beta}$ are more stable and have more influence over the shape of the distribution. An interesting point is that it is Ma's technique estimate of α that achieves the best match, even though Kogon's and McCulloch's techniques estimate $|\beta| > 0$. This supports the theory that Ma's technique is less sensitive to the value of β than theory suggests.

Technique	$\hat{\alpha}$	$\hat{\beta}$	\hat{c}	\hat{a}
McCulloch	1.511	0.046	41924	1577
Ma	1.426	-	0.878	-
Tsihrintzis	4.349	-	0.000	-0.033
Kogon	1.788	0.696	1.123	-

Table 7.10: Results for the differenced Ethernet activity data.

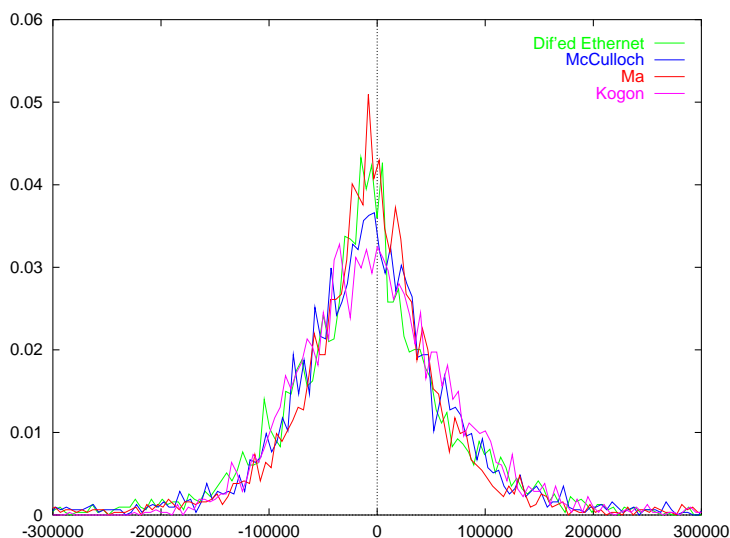


Figure 7.9: *The real and estimated stable pdfs for the differenced Ethernet data.*

7.3 Modelling data with stable distributions

7.3.1 Introduction

The later half of this thesis has dealt with two main arguments. In Chapter 5 the case for using stable distributions in teletraffic analysis was argued. In Chapter 6 and the first part of this chapter the estimation of stable parameters for teletraffic was investigated. Now we wish to carry this work though to the obvious next stage: Can we construct models based on stable distributions and do they capture the behaviour of real data?

In this section of the thesis we consider some models based on stable distributions and generate traffic which we compare with the original datasets.

The first stable model is an adaptation of the Norros model (discussed in Chapters 3, 4 and 5) but with stable innovations as opposed to Gaussian ones. The second one is based on the marginal distribution mapping model discussed in Section 3.5.3.

7.3.2 Self-similar stable processes

Chapter 7 in [67] provides a comprehensive introduction both to Brownian motion and self-similar stable processes. In the introduction the authors observe that stable H -sssi (stable, self-similar with stationary increments and Hurst exponent, H) processes have

the potential to capture two observable properties of real teletraffic data (e.g. see [50], [54] and [55]). Both these effects are named after famous Biblical characters. The Joseph effect refers to the seven years of famine followed by seven years of plenty experienced by Egypt. This is compared to the long term oscillations (or $\frac{1}{7}$ -noise) observed in self-similar traffic. The Noah effect refers to the forty days and forty nights of rains and is compared to the impulsive (infinite variance) type behaviour of some types of teletraffic. The fact that stable H -sssi processes can capture these effects is motivation for considering their application to teletraffic models.

There are many stable H -sssi processes but the one considered here is fractional Lévy stable noise (fLsn) [101], [102] which is formulated by the following integral,

$$L_{\alpha,\beta,H}(t) = \int_{-\infty}^{\infty} [(t-x-1)^{H-\frac{1}{\alpha}} - (t-x)^{H-\frac{1}{\alpha}}] M_{\alpha,\beta}(dx). \quad (7.2)$$

Here $M_{\alpha,\beta}(dx)$ is an iid stable process. In [101] it is noted that $H > \frac{1}{\alpha}$ implies positive LRD, $H < \frac{1}{\alpha}$ implies negative LRD and $H = \frac{1}{\alpha}$ implies an independent process.

7.3.3 The Norros model with symmetric stable innovations

In this section the fractional Brownian motion (fBm) process used in the Norros model is replaced by a symmetric self-similar stable process. The new Norros model generates traffic in accordance with

$$A[i] = m + \sqrt{am}(L_{\alpha,0,H}[i]) \quad (7.3)$$

and is therefore completely determined by four parameters (m , a , H and α). As for the Norros model m is the mean term and a is the scaling term. However since fLsn is an infinite variance process (and infinite mean if $\alpha < 1$) the interpretation of m and a should be treated with care. In addition since fLsn is more impulsive than fBm the probability of $A[i] < 0$ is greater for any given a and m .

7.3.3.1 Simulation results & discussion

The adapted Norros model was applied to the datasets discussed in 7.2.1. The values of α were found from the best match of the estimation techniques in 7.2.3. H was found using the R/S statistic technique, m was taken to be the mean of the arrival process, whilst a was chosen to match the second order statistic of the data. The model parameters for the datasets are recorded in Table 7.11.

Dataset	α	H	m	a
JPEG	1.90	0.80	27791.2	3880
MPEG GOP	1.28	0.82	14954.4	318
Ethernet	1.95	0.85	138177	85615.3

Table 7.11: The model parameters for the adapted Norros model with symmetric innovations.

Samples of the original data are given in Figure 7.10 and samples of the model outputs are given in Figure 7.11.

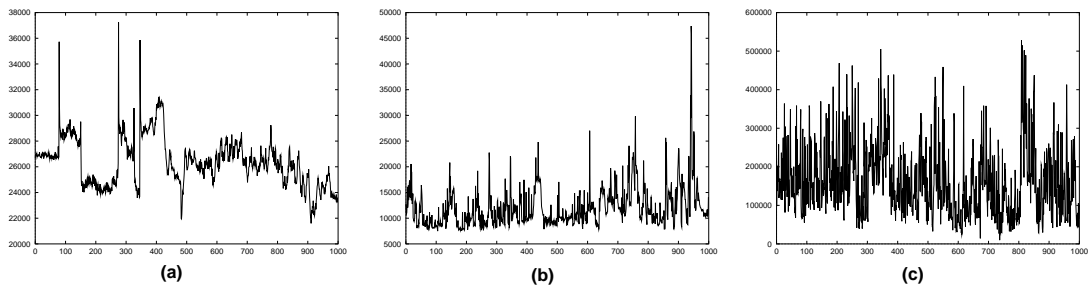


Figure 7.10: Samples from the original (a) JPEG, (b) MPEG GOP and (c) Ethernet teletraffic datasets (the y axis represents bits per frame or packet and the x axis represents frame or packet number).

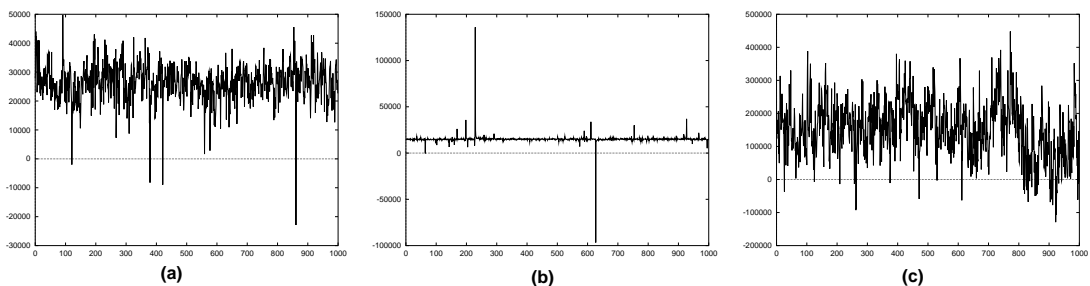


Figure 7.11: Samples from the adapted Norros model (with symmetric stable innovations) trained with (a) JPEG, (b) MPEG GOP and (c) Ethernet teletraffic datasets (the y axis represents bits per frame or packet and the x axis represents frame or packet number).

In the case of the JPEG data the model does manage to capture the marginal distribution well but all SRCS is lost. This is a problem inherent to the Norros model and has

nothing to do with the choice of innovation process. The MPEG GOP data is quite one sided in distribution but this cannot be captured with the symmetric assumption of the model. In addition the symmetry results in some negative values which must be dealt with (usually by truncation). The lower value of α compared to the other two models means that very large innovations are more likely; two of these are evident in Figure 7.11(b). The Ethernet data and model output match quite well although, again, there are some negative values.

7.3.4 The Norros model with asymmetric stable innovations

The flexibility of the model in the previous section can be increased by replacing the symmetric stable random variables with asymmetric ones. However this increases the number of parameters that must be estimated for the model. Now the arrival process is modelled by

$$A[i] = m + \sqrt{am}(L_{\alpha,\beta,H}[i]) \quad (7.4)$$

7.3.4.1 Simulation results & discussion

The three traces in Section 7.3.4 were modelled using the parameter in Table 7.12. We choose $\beta = 1$ in all cases as this skews the distribution towards positive innovations, reducing the need for truncation of $A[i] < 0$.

Dataset	α	β	H	m	a
JPEG	1.90	1.00	0.80	27791.2	3790.4
MPEG GOP	1.28	1.00	0.82	14954.4	728.5
Ethernet	1.95	1.00	0.85	138177	55594.3

Table 7.12: The model parameters for the adapted Norros model with asymmetric innovations.

Since α for the Ethernet data is so close to 2 the fact that $\beta = 1$ has little effect on the distribution. However in the other two cases the effect of the skewed innovations is more marked. In both case there is no need for truncation as all values are larger than zero. The JPEG model data matches the real data quite well apart from 2 large arrivals near the beginning of the trace. The MPEG GOP model data is also affected by larger arrivals than seen for the real data.

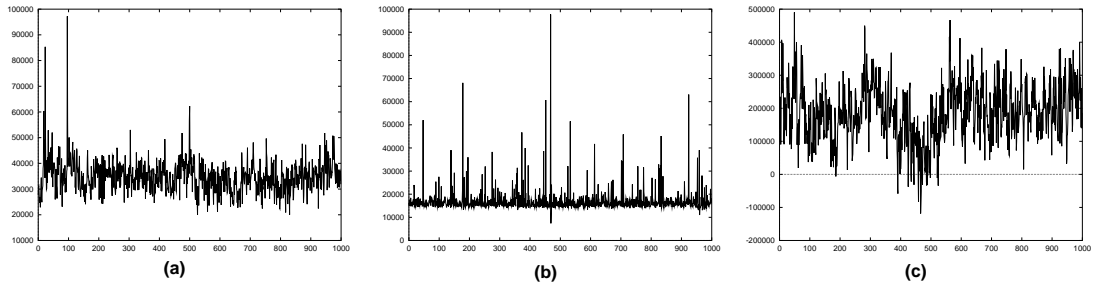


Figure 7.12: *Samples from the adapted Norros model (with asymmetric stable innovations) trained with (a) JPEG, (b) GOP MPEG and (c) Ethernet teletraffic datasets (the y axis represents bits per frame or packet and the x axis represents frame or packet number).*

7.3.5 The marginal distribution mapping model

The pdfs of the real datasets were matched to stable distributions in Section 7.2.4. Some of the results matched the data well but in other cases it was poor. However in this section we wish to demonstrate how the marginal distribution mapping model can be used to generate self-similar data from these estimates.

We use the JPEG data and the MPEG GOP data as the estimated pdf seemed to match the true pdf reasonably well. The stable parameters are given in Tables 7.6 and 7.8 and the CDF can be calculated from these using the IFT technique. The Hurst exponent for the datasets was estimated using the R/S -statistic technique.

7.3.5.1 Results & discussion

The teletraffic traces obtained for the marginal distribution mapping model are given in Figure 7.13.

The model output for the JPEG parameters performed well, although the SRCS of the original data was lost. The range of the output data is similar to that of the original, which is not surprising since the stable pdf matched the JPEG data pdf well (Figure 7.6). The MPEG GOP model output suffers from some negative values and this is unsurprising when Figure 7.8 is examined. The stable pdf is greater than zero for negative values so we would expect negative values in the marginal distribution mapping model output.

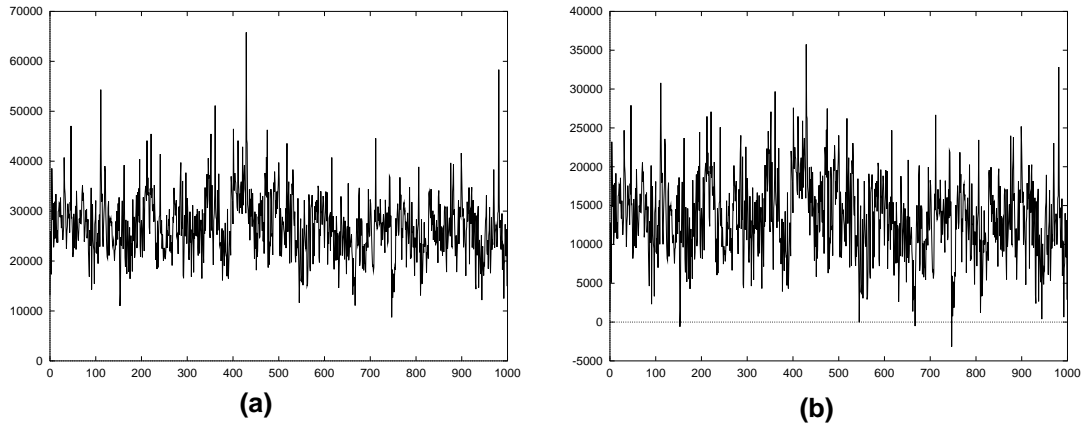


Figure 7.13: *Samples from the marginal distribution mapping model for (a) JPEG and (b) GOP MPEG data (the y axis represents bits per frame and the x axis represents frame number).*

7.4 The effective bandwidth of stable distributions

We first introduced the concept of effective bandwidths in Section 2.3.2. They have been proposed as a method for characterising teletraffic. This is because the effective bandwidth of an arrival process is directly related to the behaviour of that process as it passes through a single server queue. In addition, the effective bandwidth of independent sources are independent and summable [11].

In this section we consider the effective bandwidths of stable distributions and stable models. The first result is disheartening as it proves that all stable distributions (apart from the Gaussian) have an infinite effective bandwidth for all positive loss coefficients. However we then go on to discuss the practical implications of this result and suggest that for any real-world scenario this result will not occur. We argue that truncated stable distributions are more realistic and we investigate the effective bandwidths of these.

7.4.1 The infinite moment generating function

Recall from Section 2.3.2 that the effective bandwidth of an arrival process, $A[i]$, can be determined from

$$\Lambda(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E[\exp(\theta(A[1] + \dots + A[n]))]. \quad (7.5)$$

But $E[\exp(\theta(A[1] + \dots + A[n]))]$ is merely an expression for the moment generating function of the arrival process which can also be expressed as

$$M(\theta) = \int_{-\infty}^{\infty} f(x).e^{\theta x} dx. \quad (7.6)$$

$f(x)$ is the pdf of the arrival process. If this is stable then we know that $f(x)$ decays asymptotically in accordance with some power law,

$$\lim_{|x| \rightarrow \infty} f(x; \alpha, \beta) = C(\alpha, \beta)|x|^{-\alpha}. \quad (7.7)$$

Therefore the integral in (7.6) consists of a term that grows exponentially in x (for all $\theta > 0$) and one that decays in accordance with a power law. If the asymptotic tail behaviour in (7.7) comes into effect over the ranges $x \leq l$ and $x \geq u$ ($l \leq u$) then

$$\begin{aligned} M(\theta; \alpha, \beta) = & \left[\int_{-\infty}^l C(\alpha, \beta)|x|^{-\alpha}.e^{\theta x} dx \right. \\ & \left. + \int_u^{\infty} C(\alpha, \beta)|x|^{-\alpha}.e^{\theta x} dx + \int_l^u f(x).e^{\theta x} dx \right]. \end{aligned} \quad (7.8)$$

It is possible to show that the second term on the RHS of (7.8) is infinite for $\theta > 0$. We can therefore conclude that the effective bandwidth for all realistic values of the loss coefficient will be infinite.

7.4.2 The adapted moment generating function

At first sight the result in the previous section is disheartening since it suggests that if we attempt to estimate the effective bandwidth of any stable process then the result will be infinite. However if we consider a real world scenario we know that it is possible to model teletraffic with a stable model. In fact the result in the previous section is due to the fact that the moment generating function considers the entire probability space. In reality some form of truncation will occur because (i) negative arrivals are never permitted and (ii) the network will have some upper bound on its maximum transfer rate.

In order to investigate the growth of the moment generating function we need to rewrite

(7.6) in a slightly different form

$$M(\theta, T; \alpha, \beta) = \int_{-T}^T e^{\theta x} f(x; \alpha, \beta) dx. \quad (7.9)$$

The moment generating function has been limited by the value $T > 0$ (hence equation (7.6) can not be considered a true moment generating function, we will use the term adapted moment generating function to distinguish it from (7.6)) and we have included the stable parameters α and β . So by calculating the pdf of a stable distribution with characteristic exponent α and skew β we can find $M(\theta, T; \alpha, \beta)$ over a range of T .

In fact it is possible to obtain an exact expression for the adapted moment generating function in the cases where an expression for the pdf exists. In the case when $\alpha = 2$ an expression for the pdf is known and we can write

$$M(\theta, T, 2, 0) = e^{\frac{\theta^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-T-\theta}^{T-\theta} e^{\frac{x^2}{2}} dx. \quad (7.10)$$

Using the fact that $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$ and $\Phi(-z) = 1 - \Phi(z)$ we can write

$$M(\theta, T, 2, 0) = e^{\frac{\theta^2}{2}} (2\Phi(T - \theta) - 1). \quad (7.11)$$

We use the above expression in the next two subsections to compare with the approximate techniques for calculating the adapted moment generating function.

7.4.2.1 Fourier transform approximation to the pdf

The most obvious way to obtain an approximation of the pdf of a stable distribution is to take the inverse Fourier transform (IFT) of a sample of the characteristic function [96]. By suitably selecting the IFT size and the distance between the IFT points it is possible to generate a pdf approximation over any range of x . In Figure 7.14 the approximate pdf and adapted moment generating function are plotted for three values of α and $0 < T \leq 10$.

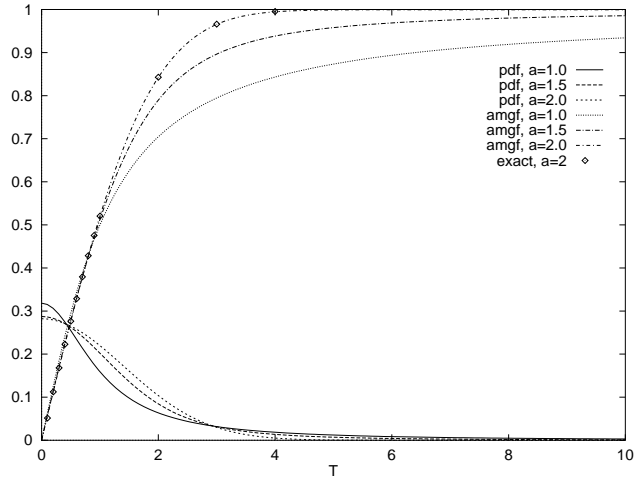


Figure 7.14: *The approximate pdf and adapted moment generating function for $\theta = 0.0001$ and $\alpha=1.0, 1.5$ and 2.0 using the IFT. The exact values for the $\alpha = 2$ case are also given.*

7.4.2.2 Asymptotic expansion approximation to the pdf

Another pdf approximation technique is that of asymptotic expansion of the power series [96]. A SaS pdf can be approximated by

$$f(x, \alpha, 0) = \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^k}{K!} \Gamma\left(\frac{2k+1}{\alpha}\right) x^{2k} \quad (7.12)$$

for $1 \leq \alpha \leq 2$. Several extensions to this technique can be employed to increase the accuracy and reduce rounding errors. We approximated the pdf and adapted moment generating function for the same θ , α and T as for the technique in Section 7.4.2.1. A comparison between these results and those in Section 7.4.2.1 are in very good agreement (to within 1%). It would seem reasonable to suggest that the differences in the results are due to inaccuracies in the approximation technique and that the estimates for the adjusted moment generating function are approximately correct.

7.4.3 Practical implications

Now consider the Norros model with the Gaussian innovations replaced by iid symmetric stable innovations,

$$A[i] = m + \sqrt{am} L_{\alpha,0,0.5}[i]. \quad (7.13)$$

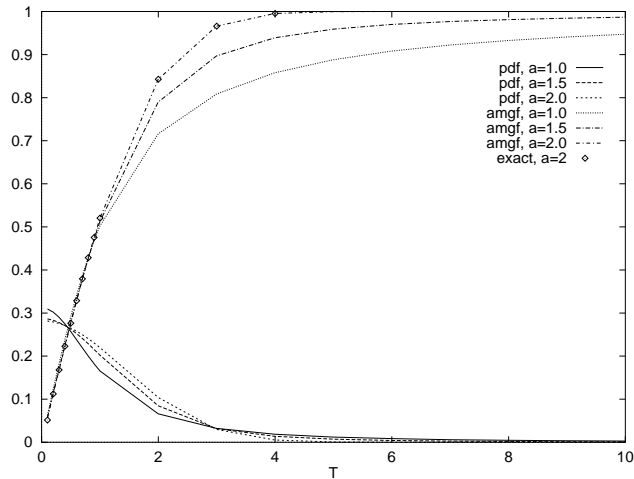


Figure 7.15: *The approximate pdf and adapted moment generating function for $\theta = 0.0001$ and $\alpha = 1.0, 1.5$ and 2.0 using asymptotic expansion. The exact values for the $\alpha = 2$ case is also given.*

If the $L_{\alpha,0,0.5}[i]$ process is truncated to $T = 10$ and $m = 10$, $a = 1$ then a semi-definite positive arrival process with $A[i] \in (0, 20)$ is produced. We estimate $\Lambda(\theta)$ for this process using the technique in [15], [16]. The result for $\theta = 0.1$ was 0.1018377, which suggests that the effective bandwidth is 10.18377. When we serviced the source at this rate we obtained the buffer occupancy probability curve in Figure 7.16.

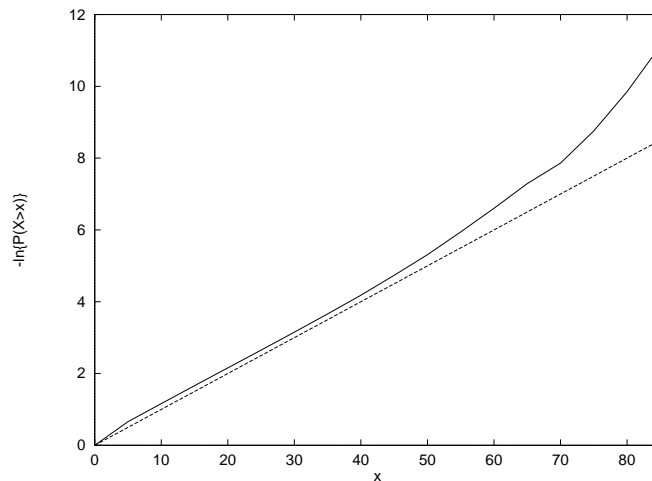


Figure 7.16: *The buffer occupancy curve for the truncated iid stable Norros model serviced at 10.1838 bits/s with the expected value from effective bandwidth theory (dashed line).*

This is a positive result in that it suggests that the on-line estimation techniques developed by Crosby *et al.* can still be applied to data with a stable innovation process. Now we wish to determine whether we can use (7.9) to estimate the effective bandwidth of the model. We applied the techniques described in Sections 7.4.2.1 and 7.4.2.2 to estimate the adjusted moment generating function with the parameters $\alpha = 1.5$, $\beta = 0$,

$T = 10$ and $\theta = 0.1$. The results were 1.005169 and 1.003800 respectively. The effective bandwidth can be estimated for the model using the following,

$$\delta(\theta) = 10 + \frac{1}{\theta} \log M(\theta, T, \alpha, \beta). \quad (7.14)$$

So the effective bandwidth estimates are 10.0518 and 10.0379. These may seem close to the 10.18377 value but in fact the system is very sensitive to $\delta(\theta)$. This is obvious when we plot the buffer occupancy plot for a service rate of 10.0518 (Figure 7.17).

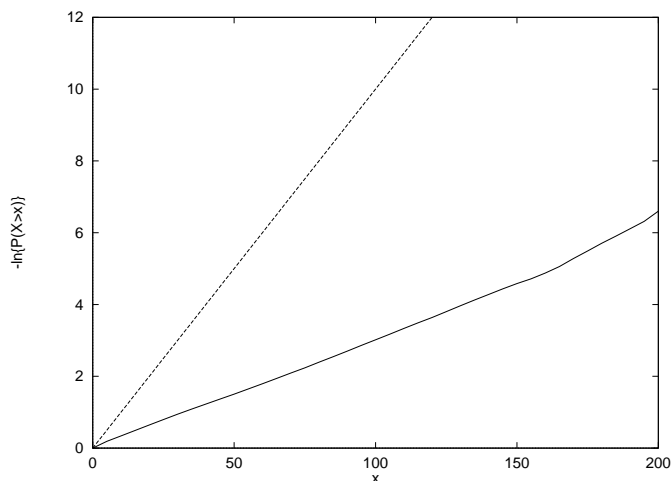


Figure 7.17: *The buffer occupancy curve for the truncated iid stable Norros model serviced at 10.0518 bits/s with the expected value from effective bandwidth theory (dashed line).*

So the adjusted moment generating function estimates of the effective bandwidth are less than the true value for the given loss figure. This is due to the fact that the adjusted moment generating function approximation to the complete pdf is not as accurate as was hoped.

It may be possible to bias the adjusted moment generating function estimates to allow for the approximation error, however this work was beyond the scope of this thesis. Instead we must conclude that approximations to the effective bandwidths of truncated stable models remain elusive for the present.

7.5 A CAC scheme based on the online estimation of α

The characteristic component α controls the tails of the stable distribution and we know that the tails of the marginal distribution of a traffic source can impact on its behaviour through a queue [54], [74]. Therefore it is possible that α could be employed

as a measure for estimating a traffic source's impact on a single server queue.

In this section we carry out several simulation studies to determine whether the α of the adapted Norros model affects its performance through a buffered switch.

7.5.1 The effect of α on the queueing behaviour of the adapted Norros model

Consider the Norros model with the fractional Gaussian noise replaced by symmetric stable noise. We wish to observe how the value of α affects the behaviour of the model traffic through a switch.

In this experiment we are more interested in investigating how the marginal distribution affects the buffer occupancy than the correlation characteristics. Therefore we begin by using iid stable random variables generated using the method in [67] from distributions of the form $S_\alpha(0, 1, 0)$.

We plotted the buffer occupancy curves for arrival processes generated by the adapted Norros model. The load in each case was 0.99, $m = 1000$ and $a = 10$. The results for $1.1 \leq \alpha \leq 1.5$ are given in Figure 7.18(a) and for $1.6 \leq \alpha \leq 2$ in Figure 7.18(b). In each case the simulation was run for 10^7 arrivals.

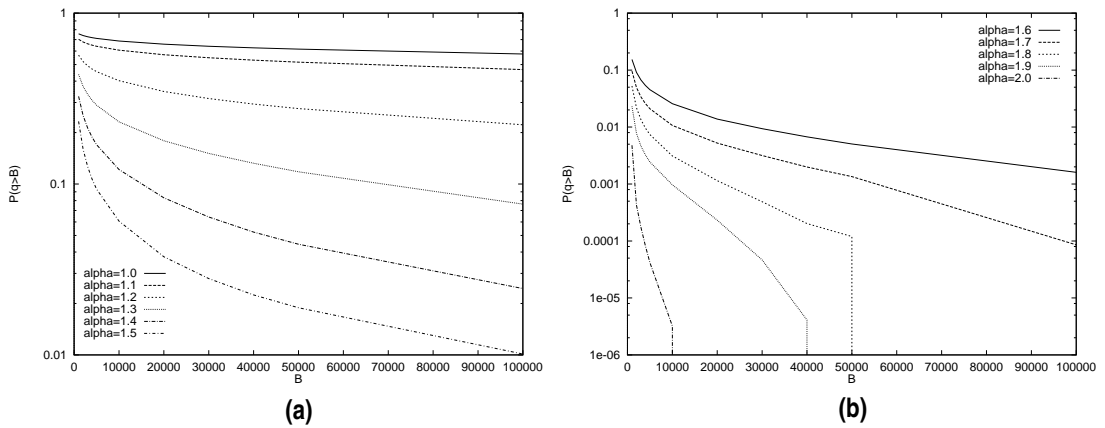


Figure 7.18: *The probability of buffer overflow for (a) $1.1 \leq \alpha \leq 1.5$ and (b) $1.6 \leq \alpha \leq 2$.*

It is obvious that the results for $\alpha = 1.9$ and $\alpha = 2.0$ suffer from the finite nature of the simulation but it is still possible to see how the buffer overflow probabilities vary with α . This experiment was repeated five times and the gradient of the linear parts of each of the slopes above were estimated. The results are plotted in Figure 7.19.

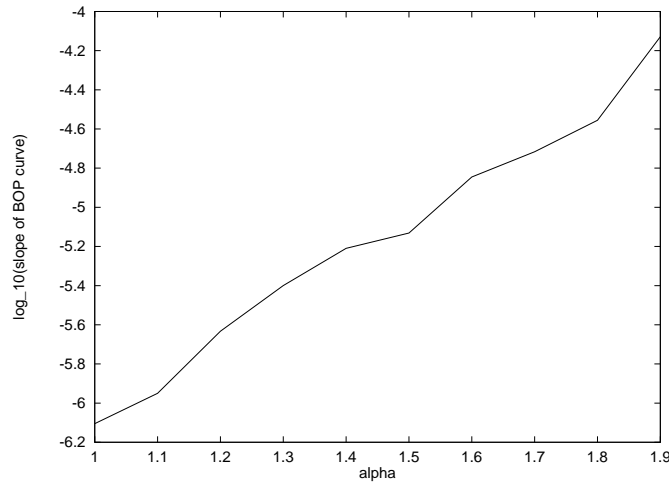


Figure 7.19: *The estimated slope of the buffer overflow probability curves vs α .*

The fact that this curve is approximately linear with a slope of 2.0529 and an intercept of -8.1441 suggests the following relationship.

$$\frac{d(\log_{10}(P(x > B)))}{dB} = 10^{(2.0529\alpha - 8.1441)}. \quad (7.15)$$

The regression coefficient for the best fit curve on this slope was 0.993, which suggests the linear fit is reasonable. Therefore we can conclude that a log-linear relationship exists between the slope of the buffer occupancy probability curve and α .

An extension to this work would be to investigate the effect of α when $H > 0.5$. However this requires the generation of long ($\geq 10^7$ arrivals) of stable self-similar noise which is a huge computational burden. Until faster algorithms for the generation of fractional Lévy stable noise exist this work remains impractical.

7.6 Conclusions

When the stable parameter estimation techniques are applied to real data the problems with accuracy are apparent. Although several results did provide a good stable estimate to the data there were occasions when the estimates provided a poor match. Therefore there is still work to be done in obtaining a quick and consistent estimation method for real data.

One observation that should be made is that in Section 6.4 of Chapter 6 the estimates for α were always similar for all four techniques (to within 10%) and these estimates

were almost always within the valid region of estimation. In Section 7.2 the estimates were often very different (up to 600%) and in many cases were outside valid parameter ranges. Although not rigorously investigated, these results would tend to suggest that large differences in the $\hat{\alpha}$, between the four techniques, implies the data is not drawn from a truly stable distribution. If this is the case then non-robust methods like Tsihrintzis' appear to be the first to fail. McCulloch's technique appears to be more robust, since it is based on sample quantiles of the data which will exist whether it is stable or not.

When we applied stable models to teletraffic data we achieved some success, both for the Norros model adaptations and the marginal model. However due to the problems with the estimation process the model matches were not always as accurate as they could be. In [79] Karasaridis and Hatzinakos used a mean absolute error technique to fit the skewed Norros model to Ethernet data. Although this technique produced more accurate results, such a technique cannot be applied in real-time and is therefore inappropriate for CAC techniques.

The effective bandwidth of all stable distributions with $\alpha < 2$ was shown to be infinite so we investigated the truncated stable distribution. This was because such distributions are more realistic in real world scenarios. However the effective bandwidth estimates based on this technique underestimated the required service rate. This is because of a mismatch between the adapted moment generating function and the true pdf of the truncated stable distribution. It may be possible to bias the adapted moment generating function to correct for this mismatch but this remains an area of future work.

Finally we looked at how α affects the queueing behaviour of the Norros model with symmetric stable innovations. As expected, lower α tended to result in longer queues in the buffer of a single server queue. This relationship implies that α may be a significant measure for CAC.

Conclusions & discussion

In this thesis we have investigated the characterisation of teletraffic with a view to developing CAC schemes for ATM networks. In this final chapter we draw some conclusions from the work and identify both achievements and limitations. In Section 8.1 we discuss areas of future work arising for the research reported on in this thesis.

In Chapter 2 we applied the structured Markovian model to ATM and derived expressions for the model parameters in terms of measurable quantities of the real traffic such as sustainable (average) cell rate and peak cell rate. This model was shown to be able to capture a wide range of traffic behaviour simply by selecting different points on a solution surface. However, although this model could have been developed further we turned our attention to self-similar processes.

Chapters 4 and 5 concentrated on investigating the two properties of teletraffic that affect its performance through a queueing environment; the correlation structure and the marginal distribution. In Chapter 4 we derived two novel non-parametric measures in an attempt to determine whether VBR video is non-stationary or self-similar. Although this problem has been looked at in a qualitative fashion [17] we wished to place some quantitative figure on the degree to which VBR video data could be considered stationary. The first measure, the wide sense stationary quotient, was derived from standard equality of mean and variance tests and was able to differentiate between iid data, data with an SRCS only and data which either had a LRCS or was non-stationary. It was unable to differentiate between data which either had a LRCS or was non-stationary because the requirement of independence between blocks was violated by the self-similar data. The second measure considered the relationship between the Hurst exponent estimate and blocksize. It was also unable to differentiate between self-similar data and the VBR data. Therefore we are unable to determine whether the VBR datasets considered are non-stationary or not but we can conclude that self-similar models best captured the behaviour of the real data.

In Chapter 5 we questioned the choice of Gaussian innovations used by several popular self-similar models. The Gaussian assumption does have the advantage that it is the limiting distribution of the aggregation of independent finite variance processes and can therefore be considered as the distribution obtained by multiplexing many independent traffic sources together. However in several cases it has been noticed that models with a Gaussian distribution are unable to capture the behaviour of unaggregated streams. We tested several teletraffic datasets using two tests for normality which conclusively showed that the distribution of the real data did not conform to the Gaussian distribution. A more quantitative analysis of the data led to the conclusion that several of the datasets had heavy tailed distributions and the others had distributions that could not be discounted as being heavy tailed.

The results in Chapter 5 suggested than any alternative to the Gaussian should be impulsive and heavy tailed. However in high levels of aggregation the Gaussian model is known to produce good results. The stable distribution can be either Gaussian or heavy tailed so we suggested it as an alternative distribution for self-similar teletraffic models.

In Chapter 6 several fast, non-MLE stable parameter estimation techniques were compared for accuracy. From this work we were able to make some suggestions about the use of these estimation techniques. It is worth pointing out that these results are not limited to teletraffic analysis and are of importance to any researcher interested in estimating stable parameters from real data.

In Chapter 7 we used the estimation framework developed in Chapter 6 to estimate stable parameters for the teletraffic data. The results suggest that further work in developing a consistent estimation technique for real data is required because in several cases the comparison between the estimated density function and the true distribution was not good. Part of the problem may be due to the fact that the teletraffic distributions are not truly stable and the estimation techniques are not robust enough to allow for this.

The Norros model was adapted to use stable self-similar innovations, both with $\beta = 0$ and $\beta = 1$. In both cases the model was able to capture aspects of the real teletraffic although problems with negative and extreme values occurred. The marginal distribution mapping model merely performed as well as the match between the real and estimated pdf. From this we can conclude that the models have potential provided the estimation procedure can be improved.

The effective bandwidth of all stable distributions with $\alpha < 2$ was shown to be infinite for realistic values of the loss coefficient. This is not as bad a result as it initially appears since a truncated stable distribution is more realistic than a true stable distribution. However when we tried to use the adjusted moment generating function to estimate the effective bandwidth of a truncated stable process we found the estimate underestimated the required service capacity.

Finally we considered the effect of the characteristic exponent on the queue behaviour of the adapted Norros model. A relationship between the two was established for an iid symmetric innovation process although limitations on computational resources prevented an investigation of the model with self-similar innovations.

In Figure 1.1 the system that is the long term goal of this work was presented. The following points summarise the progress that has been made towards developing this system.

- Chapter 4 identified that self-similar models (and hence the Hurst exponent) have potential as members of the parameter vector θ .
- Chapter 5 proved that the Gaussian self-similar models are not impulsive enough and we determined that heavy tailed distributions are more appropriate.
- Chapter 6 compared several fast stable parameter estimation techniques. If these parameters are found to be useful for CAC then this work will assist in the determination of an online estimation procedure.
- Chapters 6 and 7 present results which show that α , β , γ and a may have a role to play in a CAC scheme based on stable distributions.
- Section 7.5 developed a relationship between α and the queue behaviour of the Norros model. This may assist in the formulation of the decision function $f()$.

8.1 Future work

The long-term goal of this work is a complete CAC system and several steps have been made towards achieving this. However several areas where further work is required are discussed here.

One limitation of this work is that the stable parameter estimates do not always match real teletraffic data. A more consistent estimation technique needs to be developed and the robustness of the estimator needs to be increased. The estimation techniques discussed work when the datasets are drawn from a truly stable distribution but not when we tried to approximate a stable distribution from real data. The most robust technique was that developed by McCulloch. It generated valid estimates when the other estimation methods failed and accuracy of the estimates may be improved by pre-treating the data (e.g. scaling or differencing). Although the MLE estimation technique is computationally intensive [103] it may become practical to implement it for online estimation in the future. For this reason the robustness of this technique should be investigated by applying it to real data.

The stable self-similar Norros model has been shown to be capable of modelling real teletraffic with reasonable accuracy [79]. This fact suggests that it can be employed for parametric CAC if a link between the model's parameters and its queuing behaviour can be found. The effect of parameters m and a is well understood and in Section 7.5 we established a relationship with α but the cases when $|\beta| > 0$ and $H > 0.5$ still need to be investigated.

If the adjusted moment generating function can be biased to correct for the difference between it and the moment generating function of a truncated stable pdf then it may be possible to find closed form solutions for the effective bandwidth of a truncated stable distribution. This result could be used to develop a parametric CAC scheme based on the Norros model with stable self-similar innovations.

Finally we would like to comment on the structured Markovian model that was applied to ATM traffic in Section 2.4. The investigations into this model were concluded at an early stage as we wished to address the issues of self-similarity and heavy tails in teletraffic. However the structured Markov model still has potential for parametric CAC schemes and other Markov models have been applied with success [24]-[27]. The same success could be achieved with the structured Markov model if a relationship between the model's parameters and its performance through a buffered switch can be found.

References

- [1] J. M. Griffiths, "ATM: customer needs in the transition to B-ISDN," *Electronics & Communication Engineering Journal*, vol. 8, no. 5, pp. 234–240, 1996.
- [2] S. Bates, "The asynchronous transfer mode: Is it a waste of space?," *Electronics & Communication Engineering Journal*, vol. 8, no. 5, pp. 225–233, 1996.
- [3] M. Fontaine, "Bandwith allocation and connection admission control in ATM networks," *Electronics & Communication Engineering Journal*, vol. 8, no. 4, pp. 156–164, 1996.
- [4] E. Gelenbe and G. Pujolle, *Introduction to Queueing Networks*. John Wiley & Sons, Inc., 1987.
- [5] A. Alles, "ATM internetworking." Available at <http://cell-relay.indiana.edu/cell-realy/docs/cisco.html>, 1995.
- [6] M. H. Lyons, K. O. Jensen, and I. Hawker, "Traffic scenarios for the 21st century," *British Telecom Technology Journal*, vol. 11, pp. 73–83, October 1993.
- [7] J. Hui, "Resource allocation for broadband networks," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 1598–1608, 1988.
- [8] G. de Veciana, C. Courcoubetis, and J. Walrand, "Decoupling bandwidths for networks: A decomposition approach to resource managment," *IEEE/ACM Transactions on Networking*, 1993. submitted.
- [9] F. Kelly, "Effective bandwidths at multi-class queues," *Queueing Systems*, vol. 9, pp. 5–16, 1991.
- [10] F. Kelly, *Stochastic Networks: Theory and Applications*, ch. Notes on effective bandwidths, pp. 141–168. Oxford University Press, 1996.
- [11] G. de Veciana and J. Walrand, "Effective bandwidths: Call admission, traffic policing & filtering for ATM networks," *Queueing Systems: Theory and Applications*, no. 20, pp. 37–59, 1995.

- [12] I. Hsu and J. Walrand, "Admission control for ATM networks," in *Proceedings of the IMA Workshop on Stochastic Networks*, 1994.
- [13] I. Hsu and J. Walrand, "Admission control for multi-class ATM traffic with overflow constraints." submitted to *Computer Networks and ISDN Systems*, 1994.
- [14] I. Hsu and J. Walrand, "Dynamic bandwidth allocation for ATM switches." preprint, February 1995.
- [15] S. Crosby, I. Leslie, J. Lewis, N. O'Connell, R. Russell, and F. Toomey, "Bypassing modelling: An investigation of entropy as a traffic descriptor in the Fairisle ATM network," in *Proceedings of the 12th UK Teletraffic Symposium*, March 1995. Available from <http://stp01.stp.dias.ie/DAPG/pubs.html>.
- [16] S. Crosby, I. Leslie, M. Huggard, J. Lewis, B. McGurk, and R. Russell, "Predicting bandwidth requirements of ATM and Ethernet traffic," in *Proceedings of the 13th UK Teletraffic Symposium*, pp. 15/1 – 15/10, 1996. Available from <http://stp01.stp.dias.ie/DAPG/pubs.html>.
- [17] N. G. Duffield, J. Lewis, N. O'Connell, R. Russell, and F. Toomey, "Statistical issues raised by the Bellcore data," in *Proceedings of the 11th UK Teletraffic Symposium*, 1994. Available from <http://stp01.stp.dias.ie/DAPG/pubs.html>.
- [18] D. E. Duffy, A. A. McIntosh, M. Rosenstein, and W. Willinger, "Statistical analysis of CCSN/SS7 traffic data from working subnetworks," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 3, pp. 544–551, 1994.
- [19] N. G. Duffield, J. T. Lewis, R. Russell, and F. Toomey, "Entropy of ATM traffic streams: A tool for estimating QoS parameters," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 981–990, August 1995.
- [20] N. G. Duffield and N. O'Connell, "Large deviation and overflow probabilities for the general single server queue with applications," *Mathematical proceedings of the Cambridge Philosophical Society*, vol. 118, no. 1, 1993.
- [21] R. J. Gibbens and F. P. Kelly, "Measurement-based connection admission control," in *Proceedings of the 15th International Teletraffic Symposium*, pp. 879–888, 1997.
- [22] A. J. Ganesh and S. Bates, "Effective bandwidths of a VBR video source: Implications for call admission control in ATM networks." Unpublished, July 1995.

- [23] D. M. Lucantoni, M. F. Neuts, and A. R. Reibman, "Methods for performance evaluation of VBR video traffic models," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 176–180, April 1994.
- [24] G. Kesidis, J. Walrand, and C. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 424–428, August 1993.
- [25] G. Kesidis, "Modelling to obtain the effective bandwidth of a traffic source in an ATM network," in *Proceedings of MASCOTS*, IEEE, 1994.
- [26] Z. Fan and P. Mars, "Effective bandwidth approach to connection admission control for multimedia traffic in ATM networks," *Electronic Letters*, vol. 32, pp. 1438–1439, August 1996.
- [27] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 329–343, June 1993.
- [28] B. Melamed, B. S. D. Raychaudhuri, and J. Zdepski, "TES-based video traffic modelling for performance evaluation of integrated networks," in *Proceedings of INFOCOMM*, pp. 75–84, IEEE, 1992.
- [29] B. Melamed, B. S. D. Raychaudhuri, and J. Zdepski, "TES-based video source modelling for performance evaluation of integrated networks," *IEEE/ACM Transactions on Networking*, vol. 42, pp. 2773–2777, October 1994.
- [30] M. E. Anagnostou, J. Sanchez-P., and S. Venieris, "A multiservice structured Markovian traffic source model," in *Proceedings of ICC*, pp. 1008–1013, IEEE, 1994.
- [31] D. H. Evans, *Probability and its Applications for Engineers*. Marcel Dekker, Inc., 1992.
- [32] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation and Estimation*. John Wiley & Sons, Inc., 1990.
- [33] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–14, February 1994.
- [34] B. B. Mandelbrot, "How long is the coast of Britian? Statistical self-similarity and fractal dimension," *Science*, vol. 155, pp. 636–638, 1967.

- [35] B. B. Mandelbrot and J. W. van Ness, “Brownian motion, fractional noises and applications,” *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.
- [36] B. B. Mandelbrot and J. R. Wallis, “Some long-run properties of geophysical records,” *Water Resources and Reservoirs*, vol. 5, pp. 321–340, 1969.
- [37] B. B. Mandelbrot, *Fractals: Form, chance and dimension*. W. H. Freeman, 1977.
- [38] B. B. Mandelbrot, *The fractal geometry of nature*. W. H. Freeman, 1982.
- [39] J. Feder, *Fractals*. Plenum Press, 1988.
- [40] H. E. Hurst, “Long-term storage capacity of reservoirs,” *Transactions of the Society of Civil Engineers*, vol. 116, pp. 770–808, 1951.
- [41] H. E. Hurst, “Methods of using long-term storage in reservoirs,” *Proceedings of the Institution of Civil Engineers, Part 1*, pp. 519–577, 1955.
- [42] B. B. Mandelbrot and J. R. Wallis, “Noah, Joseph and operational hydrology,” *Water Resources and Reservoirs*, vol. 4, pp. 909–918, 1968.
- [43] B. B. Mandelbrot, “Forecasts of futures proces, unbiased markets and Martingale models,” *Journal of Business of the University of Chicago*, vol. 39, pp. 242–255, 1966.
- [44] J. McCulloch, *Statistics in Finance*, ch. Financial Applications of Stable Distributions, pp. 393–425. Elsevier, 1996.
- [45] J. Beran, *Statistics for Long-Memory Processes*. Chapman & Hall, 1994.
- [46] B. Stuck and B. Kleiner, “A statistical analysis of telephone noise,” *Bell Systems Technology Journal*, vol. 53, no. 7, pp. 1263–1320, 1974.
- [47] D. Wolf, “Noise in physical systems,” in *Proceedings of the 5th International Conference on Noise*, 1978.
- [48] V. Paxson and S. Floyd, “Wide area traffic: The failure of Poission modelling,” *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, 1995.
- [49] K. Meier-Hellstern, P. E. Wirth, Y. L. Yan, and D. A. Hoeflin, “Traffic models for ISDN data users: Office automation application,” in *Proceedings of the 13th International Teletraffic Congress* (A. Jensen and V. B. Ibersen, eds.), pp. 167–172, 1991.

- [50] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Transactions on Communications*, vol. 43, pp. 1566–1579, April 1995.
- [51] H. J. Fowler and W. E. Leland, "Local area network traffic characterisation with implications for broadband network congestion management," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp. 1139–1149, 1991.
- [52] M. W. Garrett and W. Willinger, "Analysis, modelling and generation of self-similar VBR video traffic," in *Proceedings of SIGCOMM*, pp. 269–280, 1994.
- [53] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 9, no. 3, pp. 387–396, 1994.
- [54] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 209–223, April 1996.
- [55] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high variability: Statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71–86, February 1997.
- [56] B. K. Ryu and A. Elwalid, "The importance of long range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," *Computer Communications Review*, vol. 26, pp. 3–14, October 1996.
- [57] M. Grossglauser and J. C. Bolot, "On the relevance of long range dependence in network traffic," *Computer Communications Review*, vol. 26, pp. 15–24, October 1996.
- [58] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat.*, vol. 2, 423–434 1953.
- [59] K. A. Kolmogorov, "Weinersche spiralen undeinigo andere interessante kurven in Hilbertschen raum," *Arcad. Sci.*, vol. 26, pp. 115–118, 1940. USSR.
- [60] F. Chen, J. Mellor, and P. Mars, "Comparisons of simulation algorithms for self-similar traffic models," in *Proceedings of the 13th UK Teletraffic Symposium*, pp. 8/1–8/11, 1996.
- [61] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 953–962, August 1995.

- [62] B. K. Ryu, *Fractal Network Traffic: From Understanding to Implications*. PhD thesis, Columbia University, 1996. Available at <http://www.ctr.columbia.edu/ryu>.
- [63] S. Bates and S. McLaughlin, "Is VBR video non-stationary or self-similar? Implications for ATM traffic characterisation," in *Proceedings of the European Simulation Multiconference*, June 1996.
- [64] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Prentice-Hall, Inc., 1988.
- [65] H. Akaike, "A new look at statistical model identification," *IEEE Transactions on Automation and Control*, vol. 19, pp. 716–723, 1974.
- [66] R. Shiavi, *Introduction to Applied Statistical Signal Analysis*. Aksen Associates, 1991.
- [67] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models and Infinite Variance*. Chapman & Hall, 1994.
- [68] P. E. Greenwood and M. S. Nikulin, *A guide to chi-squared testing*. Wiley, 1996.
- [69] S. S. Shapiro, M. B. Wilk, and H. J. Chen, "A comparative study of various tests for normality," *Journal of the American Statistical Society*, vol. 63, pp. 1343–72, 1968.
- [70] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3, pp. 591–611, 1965.
- [71] R. D'Agostino, "An omnibus test of normality for moderate and large size samples," *Biometrika*, vol. 58, no. 2, pp. 341–348, 1971.
- [72] R. B. D'Agostino, "Small sample probability points for the D test of normality," *Biometrika*, vol. 59, pp. 219–221, 1972.
- [73] S. Resnick, "Heavy tailed modeling and teletraffic data." Preprint, School of ORIE, Cornell University, 1995.
- [74] S. Resnick and G. Samorodnitsky, "Performance decay in a single server exponential queueing model with long range dependence," *Operations Research*, vol. 45, pp. 235–243, 1997.
- [75] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

- [76] W. H. DuMouchel, “Estimating the stable index α in order to measure tail thickness: A critique,” *The Annals of Statistics*, vol. 11, no. 4, pp. 1019–1031, 1983.
- [77] J. H. McCulloch, “Measuring tail thickness to estimate the stable index α : A critique,” *Journal of Business & Economic Statistics*, vol. 15, pp. 74–81, January 1997.
- [78] S. Bates and S. McLaughlin, “An investigation of the impulsive nature of ethernet data using stable distributions,” in *Proceedings of the 12th UK Performance Engineering Workshop* (J. Hillston and R. Pooley, eds.), pp. 17–32, 1996.
- [79] A. Karasaridis and D. Hatzinakos, “On the modeling of network traffic and fast simulation of rare events using alpha-stable self-similar processes,” in *Proceedings of the IEEE Higher Order Statistics Workshop*, pp. 268–272, July 1997.
- [80] M. Shao and C. L. Nikias, “Signal processing with fractional lower order moments: Stable processes and their applications,” *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.
- [81] A. Janicki and A. Weron, *Simulation and Chaotic Behaviour of Alpha-Stable Stochastic Processes*. Marcel Dekker, Inc., 1994.
- [82] V. M. Zolotarev, “Mellin-stieltjes transforms in probability theory,” *Theory of Applied Probability*, vol. 12, no. 4, pp. 433–460, 1957.
- [83] J. M. Chambers, C. L. Mallows, and B. W. Stuck, “A method for simulating stable random variables,” *Journal of the American Statistical Association*, vol. 71, pp. 340–344, June 1976.
- [84] E. F. Fama and R. Roll, “Parameter estimates for symmetric stable distributions,” *Journal of the American Statistical Association*, vol. 66, pp. 331–338, June 1971.
- [85] E. F. Fama and R. Roll, “Some properties of symmetric stable distributions,” *Journal of the American Statistical Association*, vol. 63, pp. 817–836, 1968.
- [86] J. H. McCulloch, “Simple consistent estimators of stable distribution parameters,” *Communications on Statistics - Simulation*, vol. 15, no. 4, pp. 1109–1136, 1986.
- [87] S. J. Press, “Estimation of univariate and multivariate stable distributions,” *Journal of the American Statistical Association*, vol. 67, no. 340, pp. 842–846, 1972.

- [88] A. S. Paulson, E. W. Holcomb, and R. A. Leitch, "The estimation of the parameters of the stable laws," *Biometrika*, vol. 62, pp. 163–170, 1975.
- [89] I. A. Koutrouvelis, "Regression-type estimation of the parameters of stable laws," *Journal of the American Statistical Association*, vol. 75, pp. 918–928, December 1980.
- [90] I. A. Koutrouvelis, "An iterative procedure for the estimation of the parameters of stable laws," *Communications in Statistics - Simulation and Computation*, vol. 10, no. 1, pp. 17–28, 1981.
- [91] S. Kogon and D. Williams, "On the characterization of impulsive noise with α -stable distributions using fourier techniques," in *Proceedings of the 29th Asilomar Conference of Signals, Systems and Computing*, 1995.
- [92] S. M. Kogon and D. B. Williams, *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*, ch. Characteristic Function Based Estimation of Stable Distribution Parameters. 1997.
- [93] X. Ma and C. L. Nikias, "Parameter estimation and blind channel identification in impulsive signal environments," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2884–2897, December 1995.
- [94] G. A. Tsihrintzis and C. L. Nikias, "Fast estimation of the parameters of alpha-stable impulsive interference using asymptotic extreme value theory," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 1840–1843, 1995.
- [95] G. A. Tsihrintzis and C. L. Nikias, "Fast estimation of the parameters of alpha-stable impulsive interference," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1492–1503, June 1996.
- [96] C. L. Nikias and M. Shao, *Signal processing with Alpha-Stable Distributions and Applications*. John Wiley & sons, 1996.
- [97] R. C. Blattberg and N. J. Gonedes, "A comparison of the stable and Student distributions as statistical models for stock prices," *Journal of Business*, vol. 47, pp. 244–280, 1974.
- [98] V. Akgiray and G. G. Booth, "The stable-law model of stock returns," *Journal of Business & Economic Statistics*, vol. 6, pp. 51–57, 1988.

- [99] M. Gibbons and P. Hess, “Day of the week effects and assets returns,” *Journal of Business*, vol. 54, pp. 579–596, 1981.
- [100] H. S. Lau and A. H. Lau, “The reliability of the stability-under addition test for the stable-paretian hypothesis,” *Journal of Statistical Computation and Simulation*, vol. 48, pp. 67–80, 1993.
- [101] S. M. Kogon and D. G. Manolakis, “Linear parametric models for signals with long-range dependence and infinite variance,” in *Proceedings of ICASSP*, pp. 1597 – 1600, 1995.
- [102] S. Kogon and D. Manolakis, “Signal processing with self-similar α -stable processes: The fractional Lévy motion model,” *IEEE Transactions on Signal Processing*, vol. 44, pp. 1006–1010, April 1996.
- [103] W. H. DuMouchel, *Stable distributions in statistical inference*. PhD thesis, Yale University, 1971.

Stable distribution estimation and modelling software

This appendix details the stable distribution software on the disk included with this thesis.

A.1 Overview

The disk included with this thesis contains the program *alpha-stable* which can:

- Generate stable random variables.
- Estimate stable parameters from input data.
- Estimate a stable pdf using the inverse Fourier transform technique.

A.2 Installation

The source files for *alpha-stable* are zipped and tarred into a file called *stable-code.tar* on the disk accompanying this thesis. To install the software onto a Unix based computer simply follow the instructions below.

- Copy the file *stable-code.tar* into your working directory space.
- Type “tar -xvf stable-code.tar” to untar the source files.
- Type “gunzip *.gz” to decompress the source files.
- Type “make” to compile the code.

- Type “./alpha-stable” to execute the program.

Note: The makefile has been written for Solaris and the C code was compiled using gcc. It may be necessary to alter the makefile according to the configuration of the target system.

A.3 Usage

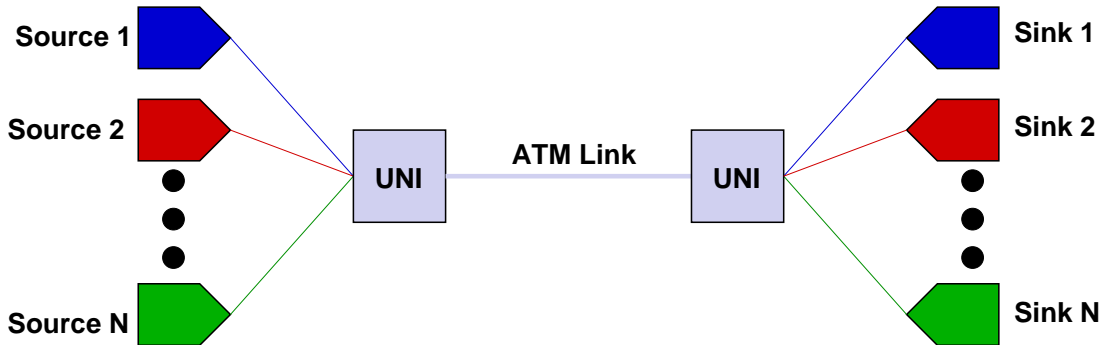
The executable *alpha-stable* is a command line prompted program and usage it reasonably obvious. Simply type “./alpha-stable” at the shell prompt and follow the on screen instructions.

A.4 Disclaimer

Although I have made every effort to ensure the accuracy of this code I can take no responsibility for any errors. This is a “use at own risk” piece of code and as such all source code is included and may be altered and reused as desired.

An experiment to illustrate statistical gain

In order to demonstrate statistical gain we can construct a simple simulation of a network link and observe how it is utilised. Our simple network consists of N sources transferring information across a single link to N sinks at the receiving end.



Each source generates a cell during each timeslot with probability p independent of other timeslots and other sources. In each case we set the link capacity equal to the maximum possible output level (i.e. N cells per timeslot) and then plot utilisation against timeslot where

$$\text{utilisation} = \frac{\text{number of cells in timeslot}}{N} \quad (\text{B.1})$$

This is the most simple of VBR models but from the results it is easy to see that link utilisation tends to p as $N \rightarrow \infty$. In the simulation results below $p = 0.5$ and the simulation length was 1000 timeslots.

So even when $N = 10$ we can see that utilisation only reaches 1 on two occasions. Of course these levels of utilisation can be calculated directly by noting that the system conforms to a binomial distribution. So if we want to determine what capacity we require to ensure that the probability of overflow (CLR) is about $1e - 8$ (a reasonable

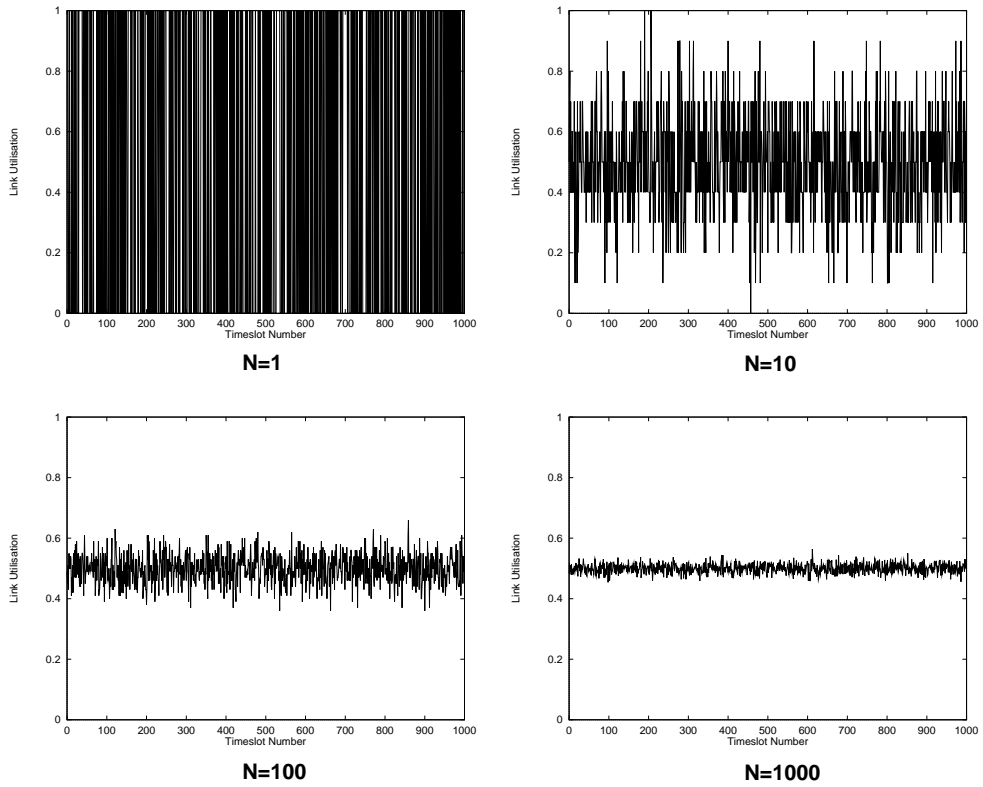


figure for ATM), we can calculate it. For example when $N = 100$ we can use a link capacity of 78 cells per timeslot and the overflow probability will be

$$\begin{aligned}
 P_{\text{overflow}} &= \sum_{i=78}^N \binom{N}{i} p^i (1-p)^{N-i}. \\
 &= 7.953e-9
 \end{aligned}
 \tag{B.2}$$

So we can reduce the link capacity to 78 cells/timeslot and maintain an acceptable CLR; this saving is the statistical gain.

The problem gets more intractable when we apply more complicated source models. This is the motivation for research into the development of suitable models and characterisation techniques for ATM networks.

The JPEG and MPEG coding schemes

C.1 Introduction

The joint photographic experts group (JPEG) and motion pictures experts group (MPEG) compression/decompression schemes have been proposed for the VBR transmission of digital image and video data on ATM networks. In this appendix we give a brief explanation of how these techniques achieve compression.

C.2 JPEG

JPEG is a lossy compression technique that was originally developed for “real-life” type images such as scanned photographs and computer rendered landscapes. By lossy we mean that JPEG discards information in accordance with some quality factor i.e. the compressed image does not contain all the information in the original.

JPEG utilises the fact that the human eye has limits in spatial resolution and colour discrimination and can achieve compression ratios of up to 50:1 without noticeable loss in picture quality. It does not work as well for two tone, greyscale and simple computer images due to the coding technique employed.

JPEG achieves compression using the following technique.

1. Transform the image into a suitable colour space (such as luminance/chrominance (Y, C_r, C_b)). This is because the eye is very sensitive to luminance but not chrominance. Therefore we can lose a lot of the chrominance information but not much of the luminance information.

2. Average the chrominance information by a factor of 2:1 horizontally and 2:1 vertically (this stage is optional).
3. Perform a discrete cosine transform on blocks of the images of size 8x8 pixels. The resultant contains information about the frequency components of each block. High frequency information can not be perceived by the eye and can be discarded.
4. The remaining frequency coefficients are rounded into integers and encoded using some lossless coding scheme.
5. A JPEG header containing the encoding parameters is added

Several extensions to the baseline JPEG standard exist, for further information refer to the website

<http://www.uni-mannheim.de/ftp/info/ftp/compression-faq/part2>.

C.3 MPEG

MPEG has been developed for motion pictures and works best with “real life” type data. It is a lossy compression technique and employs limits in the eye’s spatial and temporal resolution to increase compression. In addition MPEG employs the fact that many frames in a video sequence are similar to their nearest neighbours. By merely sending information on pixels that have changed between adjacent frames, large increases in the compression factor can be achieved.

MPEG uses 3 modes for encoding frames. These are called intraframe (I), predictive (P) and interpolative (B). The I frame is a JPEG encoding of a single frame (hence makes no use of temporal redundancy). It is the least efficient of the 3 modes but can correct any running errors produced by the B and P frames. Normally every 12th frame is an I frame. i.e. the coding sequence ...IBBPBBPBBPBBIBBPBBP... is used. The P frames are generated from the nearest I or P frame if the match between the frames is close enough. The P frame is encoded into 8x8 block DCT coefficients and if these values are close to the previous coefficients for that block then the differential information is sent. Therefore the P frame only improves in the compression of an I frame if there is temporal correlation between adjacent I and P frames. The B frames are always generated using differential information and use forward and backward correlations.

The encoder compares the previous and future I and P frames and use the ones that are closest to the B frame.

If the original video has little motion or scene changes then MPEG can achieve extremely high compression ratios. This has major implications for applications such as video telephony. On the other hand if the video sequence has rapid panning or scene changes then the compression ratio will not be as high and “artifacts’ may be introduced due to errors in the B and P frames. For further information see the web site

<http://www.bok.net/tristan/MPEG/>.

Appendix D

Publications

The following publications have either been published or submitted to journals or conferences. The publications marked with a † have been included in this appendix.

- S. Bates and S. McLaughlin**, “The investigation of a structured Markovian model for ATM networks,” *Proceedings of the Second Communication Network Symposium*, Manchester, pp 154-157, 1995.
- S. Bates and S. McLaughlin**, “The Investigation of a structured Markovian traffic source model,” *Proceedings of the IFIP ATM Workshop*, Ilkley, pp 37/1-37/10, 1995.
- †**S. Bates and S. McLaughlin**, “Is VBR video non-stationary or self-similar? Implications for ATM traffic characterisation,” *Proceedings of the European Simulation Multiconference*, Budapest, pp 531-534, 1996.
- S. Bates and S. McLaughlin**, “An investigation of the impulsive nature of Ethernet data using stable distributions ,” *Proceedings of the 12th UK Performance Engineering Workshop*, Edinburgh, pp 17-32, 1996.
- †**S. Bates**, “The Asynchronous Transfer Mode: Is it a waste of space?,” *Electronics and Communication Engineering Journal*, pp 225-233, 1996.
- †**S. Bates and S. McLaughlin**, “The estimation of stable parameters,” *Proceedings of the IEEE Higher Order Statistics Workshop*, Banff, pp 390-394, 1997.
- S. Bates and S. McLaughlin**, “Testing the Gaussian assumption for self-similar teletraffic model,” *Proceedings of the IEEE Higher Order Statistics Workshop*, Banff, pp 444-447, 1997.

Submitted:

S. Bates and S. McLaughlin, “The estimation of stable parameters (extended version),” *IEEE Transactions on Signal Processing*, August 1997.