INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Towards a New Method of Traffic Characterization
based on the Moment Generating Function
with Application to Queueing

by

Tarkan Taralp, B.A.Sc.
University of Ottawa

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master of Engineering

Ottawa-Carleton Institute for Electrical Engineering
Department of Systems and Computer Engineering

Carleton University
Ottawa, Ontario
29 August 1997
© 1997 Tarkan Taralp

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-27006-8

Canada

The undersigned recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis

**Towards a New Method of Traffic Characterization based on the Moment
Generating Function with Application to Queueing**

submitted by Tarkan Taralp. B.A.Sc.
in partial fulfillment of the requirements for
the degree of Master of Engineering

Thesis Supervisor

Thesis Co-supervisor

Chair, Department of Systems and
Computer Engineering

Carleton University

15 October 1997
date

Abstract

A new approach to traffic characterization is developed which is motivated by the traffic-queue analysis techniques of effective bandwidth and the use of log moment generating functions. A traffic stream exhibiting non-Gaussian marginal and long range dependence is modeled by four dissimilar stochastic models: a Markovian TES, an autoregressive ARM, a renewal SRP, and a non-Gaussian self-similar model. Generated traces of the model traffic are fed into a $G/D/1/\infty$ queue system. Experimentation demonstrates that the marginal-autocorrelation pair of traffic descriptors have shortcomings such that different input models with the same apparent characterizations can produce vastly different queueing results. The ability of log moment generating functions to characterize traffic is demonstrated to be robust with regard to queueing effects, in comparison to the marginal-autocorrelation methodology. Secondary objectives met during thesis research include: development and analysis of an Extended-TES model, an experimental MPEG video model, and an efficient FGN generation technique.

Acknowledgements

I wish to thank my thesis supervisor Professor Mihail Devetsikiotis. I also wish to thank Professor Ioannis Lambadaris, group leader Professor Roger Kaye, and my co-supervisor Professor Amit Bose from the Department of Mathematics and Statistics. Their helpful suggestions, direction and encouragement has been deeply appreciated. I gratefully acknowledge the funding and equipment provided by the Telecommunications Research Institute of Ontario and by the Natural Sciences and Engineering Research Council in support of this research.

Table of Contents

Acceptance Sheet	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Acronyms	xi
List of Symbols	xii
1 Introduction	1
1.1 Stochastic Processes	2
1.2 Stochastic Traffic	3
1.3 Traffic Characterization	7
1.4 Thesis Objective – The Characterization of Traffic	9
1.5 Thesis Organization	14
2 Background	16
2.1 Survey of Traffic Models	16
2.1.1 Poisson Models	16
2.1.2 Markovian Models	17
2.1.3 Renewal Models	19
2.1.4 Self-Similar Models	20
2.1.5 Other Modeling Approaches	20
2.2 Network Traffic Issues	22
3 Experimental Design	24
3.1 Traffic Specification	24
3.2 Choice of Models	27
3.3 Simulation Experiment	28
4 The TES Model	32
4.1 Review of TES ⁺ Processes	32

4.2	Specification of Experimental Study Model	36
5	The E-TES Model	38
5.1	Description	38
5.2	Autocorrelation Co-efficient Function	40
5.2.1	General Algorithm	42
5.3	Implementation	45
5.3.1	Extended TES Generator Program	47
5.4	Specification of Experimental Study Model	47
5.5	Verification	48
5.6	Conclusion	51
6	The NG-LRD Model	52
6.1	Description	53
6.2	Specification of Experimental Study Model	54
6.3	Conclusion	55
7	The SRP Model	56
7.1	Description	57
7.2	Specialized SRP Model	58
7.2.1	Specification of Experimental Study Model	61
7.2.2	Implementation	62
7.2.3	Verification	63
7.3	Conclusion	64
8	Characterization Based on Moment Generating Functions	65
8.1	Literature	66
8.2	The Case for Adopting the Moment Generating Function	69
8.3	The $\mathcal{B}(\theta, T)$ Function and its Interpretation	70
8.3.1	Implementation	72
8.4	Conclusion	73
9	Simulation Results and Measurements	74
9.1	Simulation Runs	74
9.2	The E-TES, SRP and NG-LRD Traffic Runs	75
9.3	The TES Traffic Run	80
9.4	Characterization Using the $\mathcal{B}(\theta, T)$ Function	82
9.5	Further Evidence with the SRPMOD Traffic	86
10	Conclusion	90
10.1	General Discussion	90
10.2	Achievements	94

10.3 A New Modeling Methodology?	95
References	97
A Simulation Using MODSIM	102
A.1 Introduction	102
A.2 Class Types	102
A.2.1 GenTypeObj Objects	104
A.2.2 SourceTypeObj Objects	104
A.2.3 QueTypeObj Objects	105
A.2.4 RecordToolObj Object	107
A.3 Data and Performance Measures	107
A.4 Design of the Simulation Program	107
B An MPEG Generation Scheme	110
B.1 Generation of I, P and B Frames	112
B.1.1 Experimental Results	114
B.2 MPEG Scene Changes	117
B.2.1 Experimental Results	118
B.3 Implementation	118
B.4 Discussion	120
C An FGN Generator Using SRP	121
C.1 Description	121
C.2 Results	122
C.3 Discussion	124

List of Tables

3.1	Parameters Considered in Main Experiment	30
4.1	Background Innovation Process	35
4.2	TES Model Input Parameters	37
5.1	Background Innovation Process for Extended TES	40
5.2	Pascal's Triangle and Mapping	43
5.3	Analytical Calculation of Autocorrelation	44
5.4	E-TES Model Input Parameters	48
6.1	NG-LRD Model Input Parameters	54
7.1	SRP Model Input Parameters	61
7.2	Pseudo-Code for Routine <code>genSrp.c</code>	63
9.1	Comparison of $B_X^\pi - B_Y^\pi$ Mean Differences	85
9.2	Comparison of $B_X^\pi - B_Y^\pi$ RMS Differences	86
9.3	Comparison of $B_{huang}^\pi - B_{srpmod}^\pi$ Differences	87
A.1	Pseudo-code for <code>Mlindleysim.mod</code>	109

List of Figures

3.1	Steady-State Marginal Distribution Specification for Traffic	25
3.2	Autocorrelation Specification for Traffic	25
3.3	Autocorrelation of PseudoEmpirical Trace	26
3.4	Illustration of Slotted Queueing System	29
3.5	Block Diagram of Experiment	29
4.1	Example of TES Sample Path	33
4.2	TES Model Innovation PDF and Stitching Functions	36
4.3	Autocorrelation Spec versus TES Analytical	37
5.1	Example of Moving-Average Filter	39
5.2	Autocorrelation Spec versus E-TES Analytical	49
5.3	Example of ETES Sample Path	49
5.4	Histograms of TES and E-TES	50
6.1	Example of NG-LRD Sample Path	52
7.1	Example of SRP Sample Path	57
7.2	Autocorrelation $\rho(t)$ - Early Test	59
7.3	Sample Path - Early Test	60
7.4	Autocorrelation Specification versus SRP Analytical	62
8.1	Idealized Fluid-Flow Traffic & Queue	68
8.2	Empirically Measured $\mathcal{B}(\theta, k)$ Function of Binomial Series	72
9.1	Buffer Occupancy $\Pr[Q > y]$ for E-TES, SRP, NG-LRD Traffic	76
9.2	Marginal Distribution of E-TES, SRP, NG-LRD Traffic vs. Spec. . .	77
9.3	Autocorrelation of E-TES Traffic vs. Specification	78
9.4	Autocorrelation of SRP Traffic vs. Specification	78
9.5	Autocorrelation of NG-LRD Traffic vs. Specification	79
9.6	Buffer Occupancy $\Pr[Q > y]$ for TES Traffic	80
9.7	Marginal Distribution of TES Traffic vs. Specification	81
9.8	Autocorrelation of TES Traffic vs. Specification	81

9.9	$\mathcal{B}(\theta, t)$ Characterization of TES Traffic	83
9.10	$\mathcal{B}(\theta, t)$ Characterization of E-TES Traffic	83
9.11	$\mathcal{B}(\theta, t)$ Characterization of SRP Traffic	84
9.12	$\mathcal{B}(\theta, t)$ Characterization of NG-LRD Traffic	84
9.13	Buffer Occupancy $\Pr[Q > y]$ with SRPMOD Traffic	87
9.14	Marginal Distribution of SRPMOD Traffic vs. Specification	87
9.15	Autocorrelation of SRPMOD Traffic vs. Specification	88
9.16	$\mathcal{B}(\theta, t)$ Characterization of SRPMOD Traffic	88
A.1	Class Hierarchy of Objects	103
A.2	Test run of DiscreteCellObj and LindleySrcObj Objects	105
A.3	Test run of Continuous-Time Source Objects	106
A.4	Test run of Lindley Source Objects	106
A.5	Component Context Diagram for Mlindleysim.mod	108
B.1	Autocorrelation of I-Frames	111
B.2	Block Diagram of MPEG-SRP Generator	113
B.3	Sample Path Empirical IBP and Synthetic IBP	115
B.4	Histogram of IBP Empirical and Synthetic	116
B.5	Autocorrelation of IBP Empirical and Synthetic	116
B.6	Autocorrelation of IBP Empirical and Synthetic	117
B.7	Autocorrelation Functions with MPEG Scene Changes	119
C.1	Analytical Autocorrelation Curve with $H=0.75$	123
C.2	Renewal CDF Function with $H=0.75$	123
C.3	Empirically measured Marginal Distribution of Gaussian10 process	124
C.4	Empirical Vs Analytical Autocorrelation of Gaussian10 process	125
C.5	Variance-Time Plot of IID process vs Gaussian10 process	125

List of Acronyms

PDF	Probability Density Function
PMF	Probability Mass Function
CDF	Cumulative Distribution Function
IID	Independent Identically Distributed
IAT	Interarrival Time
SRD	Short Range Dependent(ence)
LRD	Long Range Dependent(ence)
MMRP	Markov Modulated Rate Process
MMPP	Markov Modulated Poisson Process
MAP	Markov Arrival Process
FBM	Fractional Brownian Motion
FGN	Fractional Gaussian Noise
ARMA	Autoregressive Moving Average
F-ARIMA	Fractional Autoregressive Integrated Moving Average
TES	Transform Expand Sample
ARM	Autoregressive Modular Process
E-TES	Extended-TES
SRP	Spatial Renewal Process
NG-LRD	Second-Order Self-Similar model developed by Huang
LAN	Local Area Network
ATM	Asynchronous Transfer Mode
MPEG	Motion Picture Experts Group
CBR	Constant Bit Rate
VBR	Variable Bit Rate

List of Symbols

X_n	Discrete Time Stochastic Process
X_t	Continuous Time Stochastic Process
$f_X(x)$	Steady-state Marginal Distribution of X
$\rho_X(k)$	Autocorrelation of the k^{th} time lag of X_n
$\mathcal{E}[X]$	Expected Value of random variable X
$M_X(\theta)$	Moment Generating Function of X
A_n^T	Partial Sum Process of X_n aggregated over T samples
$A^*(\theta, T)$	Envelope Process of X_n
$a^*(\theta)$	Effective Bandwidth Function of X_n
$\mathcal{B}_X(\theta, T)$	Normalized Log Moment Generating Function of X_n
$U_n^\pm$	TES Background Process
$X_n^\pm$	TES Foreground Process
$\langle x \rangle$	Modulo-1 operator
$[x]^+$	Absolute value of x if $x \geq 0$ otherwise 0
$\mathcal{F}$	Fourier transform operator
$\mathcal{L}$	Laplace transform operator
$\tilde{f}$	Laplace transform of function f
$\Re$	Real part operator

Chapter 1

Introduction

The aim of *traffic modeling* is to describe the random flow of traffic associated with such sources as VBR video and LAN/WAN data in terms of stochastic models. The stochastic traffic models drive queueing system models to simulate the operation of a network or switch design. The resulting estimates of the system's performance provide the necessary information for buffer sizing, capacity allocation and other QoS provisioning design tasks. While systems can be modeled by analytical queueing techniques, they are more amenable to event simulation on computer, the former option being more theoretical in its approach while the latter more practically applicable to realistically complex designs.

In this thesis an example of long range dependent traffic will be characterized by its marginal distribution function, its autocorrelation function, and a measurement related to effective bandwidth. Evidence provided by a computer simulation comparing the queueing effects of five models of traffic (spanning four types) will show that the first two characterizations are not always adequate, and that the third method is a promising new candidate. The traffic modeling area is heavily indebted to the mathematics of *stochastic processes* because the quantity of traffic flow is a phenomenon of time and is random in nature.

1.1 Stochastic Processes

In order to better understand the random processes that are observed throughout science and nature, random variables and random events play a key role in conceptual models of such phenomena. The random event in logic parallels the random variable in mathematics in a very intuitive sense since they both serve the purpose of quantifying or acting as proxy for unknown or uncertain information comprising the state of a model. By definition a random variable X assumes a value from a set of possible values, or state-space. This state space can be infinite or finite and it can be discrete or continuous. The likelihood of a given value or event occurring relative to any other is determined by the Cumulative Distribution Function (CDF) or one of its derivatives, the Probability Density Function (PDF) for the case of continuous space, or the Probability Mass Function (PMF) for discrete space. The CDF function is defined in Equation 1.1 and followed by the definitions of the PDF and PMF functions in Equations 1.2 and 1.3 respectively.

$$F_X(x) = \Pr[X \leq x] \quad (1.1)$$

$$f_X(x) = \frac{d}{dx} \Pr[X \leq x] \quad (1.2)$$

$$f_X(x) = \Pr[X = x] \quad (1.3)$$

When attempting to model the almost limitless random events found in the real world it is natural to consider random events occurring not only in terms of their state-space but also in terms of time. A sequence of random variables $\{X_k\}$ defines a discrete-time stochastic process, indexed by time parameter k . It is a countable family of random variables. The term independent and identically distributed (IID) refers to a discrete-time process whose random variable elements are all statistically independent of one other. This special case can be fully characterized by amalgamating the individual element CDFs into a multi-dimensional CDF by the simple operation of multiplication. The definition of stochastic process can be further extended to

include continuous-time processes $\{X_t\}$ which are defined over a continuous range of t . But in contrast to the discrete case the individual random variable elements are indexed infinitesimally densely, resulting in an uncountably infinite family of random variables.

When statistical dependence exists among sets $(X_{t_1}, X_{t_2}, \dots)$ of random variable elements indexed at different times, the process is referred to as having *temporal structure* or *time dependence*. To adequately describe a process with time dependence all the multi-dimensional joint distributions having form (1.4) must be known.

$$\Pr[X_{t_1} \leq x_1, X_{t_2} \leq x_2, X_{t_3} \leq x_3, \dots] \quad (1.4)$$

With the introduction of time dependence, analysis and manipulation of the process becomes much more complex as compared to IID processes, even after the application of some simplifications and assumptions. Nevertheless stochastic processes prove very useful when describing random events associated with natural and man-made phenomena. An example: the selection, production, transmission and utilization of digital information and resources in the context of a large network or system are for all intents and purposes random processes indexed by time. A thorough treatment of stochastic processes from an electrical engineering perspective is provided in the textbook by Garcia [1]. It has become clear recently that stochastic processes describing such realistic phenomena exhibit heavy temporal structure, so current attention has focused more on specific issues of time dependence such as the importance of long range dependence. When time dependence is not sufficiently represented, the results of a statistical experiment such as in a network performance simulation may be grossly in error [2, 3].

1.2 Stochastic Traffic

A stochastic process may be interpreted in the traffic modeling context as the flow rate function of the data (in some arbitrary unit e.g. bits/sec) with respect to time.

Data is conceptualized as a randomly fluctuating fluid flow. The discrete-time process $\{X_k\}$ represents the quantity of data produced in one time unit, while the continuous time version $\{X_t\}$ describes the instantaneous rate of flow at time t . The translation of flow rate into discrete packets or data cells is considered to be a subsequent but separate issue. This viewpoint is justifiable for traffic whose data unit size is relatively small in comparison to the data flow rate, for example the quasi-fluid flow of ATM cells.

In some papers, a discrete-time process A_i is used to directly represent the interarrival (IAT) times between the i^{th} and $i + 1^{th}$ discrete data unit, and as such does not model flow rates but rather the simulation level objects. Such a view of traffic is perhaps more applicable to large-packet based traffic, however it will not be considered in this thesis. The first convention, that traffic is a fluid flow measured by a rate function, will be used to measure and model traffic.

There is no shortage of literature on the topic of network traffic models. The ultimate goal is the capability to analyze and evaluate network architecture designs while on the drawing board, with the use of simulation and realistically adequate synthetic traffic. A survey of literature over the past few years recounts many individual approaches to modeling network traffic, and while some methods are quite creative there is no consistent general method or principle that is being followed. Rather, specific modeling scenarios are described and analyzed with various ad-hoc techniques in order to serve a specific purpose. One is resigned to accept that there exists at yet no general theory of stochastic traffic and queueing systems as there is for example a well established understanding of circuit theory based on linear mathematics. This is not to imply that the exploratory work being done today is not useful, indeed industry is actively recruiting the expertise of researchers for the performance simulation of systems with video, packet data and even ATM traffic. But it is an open question as to how well simulation and analysis can be scaled up indefinitely without a general theory and whether it will even be valid to do so.

One widely made assumption in the traffic modeling area is that real network traffic can be adequately described by *time homogeneous* processes. This implies a process whose underlying stochastic nature does not change over time [1] or a process which is indifferent to translation in the time axis. A necessary but insufficient condition for time homogeneity is given by Wide Sense Stationarity (WSS). A process is said to be Wide Sense Stationary if its one- and two-dimensional joint distributions are invariant of absolute time as defined by conditions 1.5 and 1.6 below.

$$\Pr[X_{t_1} \leq x] = \Pr[X_{t_2} \leq x] \quad \forall t_1, t_2 \quad (1.5)$$

$$\Pr[X_{t_1} \leq x_1, X_{t_1+\tau} \leq x_2] = \Pr[X_{t_2} \leq x_1, X_{t_2+\tau} \leq x_2] \quad \forall t_1, t_2, \tau \quad (1.6)$$

The WSS stationarity assumption simplifies the task of characterizing a stochastic process because it implies a constant mean $\mathcal{E}[X_t] = m$, and a covariance statistic $\mathcal{E}[X_{t_1} X_{t_2}]$ dependent only on a single $\tau = t_2 - t_1$ parameter. Ultimately most academic papers presume the WSS stationarity property in their modeling exercises without further justification, likewise in this thesis traffic will be considered a WSS stationary phenomenon.

Another notion discussed in the traffic modeling field is that of *time scale* which refers to the change or immunity to change of the process structure upon scaling of the time axis [4]. One can define a process $\{X_t\}$ to be scaling invariant if for some $\alpha \in [a, b]$ the process as described by (1.4) is equal in distribution to its scaled $\{X_{\alpha t}\}$ version. If a traffic is not scale invariant an examination of it from small to large time scales should show a reduction in relative magnitudes and lengths of the random fluctuations and bursts, degenerating towards white noise. Traditional traffic models such as the Poisson process exemplify this kind of behavior, which can be described as lacking burstiness. But typical network traffic is characterized by random fluctuations, cycles and bursts of various magnitude and length, exhibiting patterns and bursts that co-exist over many magnitudes of time scale from minutes

to hours to days, implying time scale invariance [3]. Good synthetic models should capture this time scale characteristic whether it is explicitly accounted for or implicitly carried by other characterizations.

To more directly address the issue of time scaling one can take the “self-similar” traffic modeling approach. In recent years *self-similarity* has gained attention as more complete empirical studies of real network traffic have become available. The self-similarity characteristic of traffic is predominantly defined with regard to the second-order statistical measures of the traffic sequence. The term long range dependence (LRD) is used to refer to this type of self-similarity. Let $\{X_k : k = 0, 1, 2, \dots\}$ be a WSS stationary discrete-time stochastic process. The sequence $\{X_k^{(m)} : k = 0, 1, 2, \dots\}$ defined in Equation 1.7 begins with the original process and aggregates it in time by factor m where m is integer.

$$X_k^{(m)} = \frac{1}{m} \sum_{i=k-m+1}^{km} X_i \quad (1.7)$$

The second-order statistic used to measure the aggregate process $\{X^{(m)}\}$ is:

$$\rho^{[m]}(j) = \frac{\mathcal{E}[X_k^{(m)} X_{k+j}^{(m)}] - \mathcal{E}[X^{(m)}]^2}{\mathcal{E}[(X^{(m)})^2] - \mathcal{E}[X^{(m)}]^2} \quad m \geq 2 \quad (1.8)$$

A WSS stochastic process is defined to be exactly second-order self-similar if it satisfies condition 1.9 or asymptotically self-similar if it satisfies 1.10 [5, 6].

$$\rho^{(m)}(j) = \rho^{(1)}(j) \quad j = 0, 1, 2, \dots \quad (1.9)$$

$$\lim_{m \rightarrow \infty} \rho^{(m)}(j) = \rho^{(1)}(j) \quad j = 0, 1, 2, \dots \quad (1.10)$$

In other words the second-order statistic $\rho^{(m)}(k)$ of the aggregated sequences $X_k^{(m)}$ is similar to that of the original process $\rho^{(1)}(k)$. This condition implies that scaling the process in time does not degrade the time dependence structure proportionally as would be expected with traditional short range dependent (SRD) processes. Another feature of self-similarity is that the power spectrum density function of the process

$S_X(f)$ approaches infinity as f approaches zero. Similarly such traffic has a sub-exponentially decaying second-order statistic and a summation $\sum |\rho(k)| \rightarrow \infty$ that is infinite [2, 7]. Stochastic processes not exhibiting second-order self-similarity or LRD time dependence are termed Short Range Dependent (SRD).

Evidence from empirical studies of network traffic is starting to show that packet data and compressed video are self-similar or LRD [3, 8, 9, 7, 10]. Typically the traffic rate or packet interarrival distribution is extremely burstly, having heavy traffic interspersed among long quiet periods. This is not to imply that all traffic is LRD as [11] found that videoconferencing traffic (as opposed to MPEG traffic) was very satisfactorily modeled with exponentially decaying Markov Processes. A more philosophical debate questions the validity of assuming self-similarity a priori and analyzing the empirical traffic as such, when in fact the traffic might be better understood in terms of time homogeneous SRD processes modulated by some external phenomenon [12, 2]. But ultimately modelers of LRD traffic are interested in providing a reasonable model of some given traffic that can be easily specified with as few parameters as possible, so taking the self-similar view might facilitate that goal.

1.3 Traffic Characterization

A stochastic process poses an interesting problem in that its random nature by definition precludes an exact deterministic description of itself. So a process must be described by statistically based measurements of one or more of the process's characteristics. Amongst the methods used in traffic modeling to describe and identify stochastic traffic it would be fair to indicate the use of the marginal distribution and autocorrelation function pair as one of the more established methods [13, 14].

The primary characteristic of a process $\{X_t\}$ is the steady-state first-order distribution of X . The one-dimensional CDF (or its equivalent PDF function) captures this statistic, and is commonly referred to as the *marginal distribution* of the

traffic. It will be presumed that the process is WSS stationary and consequently that the definition of the marginal distribution and the PDF function (Equation 1.11) is invariant of time.

$$f_X(x) = f_{X_t}(x) = \frac{d}{dx} \Pr[X_t \leq x] \quad (1.11)$$

If one ignores time dependence and creates a mathematical model based solely on the PDF, an IID process results. These “first generation” models have been successfully used in telephony and early networks, the most ubiquitous model being the Poisson Arrival Process and its associated Exponential interarrival time sequence. For the most part it is no longer accepted that these traditional models are satisfactory to model network traffic [15, 16, 2].

In order to describe the time dependence structure of a stochastic process a second-order measurement is used to supplement the marginal distribution function. Usually referred to as the *autocorrelation function* its general definition is given in Equation 1.12. However under WSS stationarity the autocorrelation function is invariant with respect to time because it need consider only the difference τ between two relative time points. Under such an assumption, the autocorrelation as defined by Equation 1.13 applies.

$$\rho(t_1, t_2) = \frac{\mathcal{E}[X_{t_1} X_{t_2}] - \mathcal{E}[X_{t_1}] \mathcal{E}[X_{t_2}]}{\sqrt{\mathcal{E}[X_{t_1}^2] \mathcal{E}[X_{t_2}^2] - \mathcal{E}[X_{t_1}] \mathcal{E}[X_{t_2}]}} \quad (1.12)$$

$$\rho(\tau) = \frac{R_X(\tau) - \mu_X^2}{\sigma_X^2} \quad \text{where} \quad \begin{aligned} R_X(\tau) &= \mathcal{E}[X_t X_{t+\tau}] \\ \mu_X &= \mathcal{E}[X_t] \\ \sigma_X^2 &= \mathcal{E}[(X_t - \mu_X)^2] \end{aligned} \quad (1.13)$$

The autocorrelation $\rho(\tau)$ measures statistical relationship of the process $\{X_t\}$ between two points in time X_{t_1} and X_{t_2} . The relationship is expressed as a linear correlation value, and as such does not truly capture all the possible second-order interaction.

There is an exact relationship between the correlation function $R_X(\tau)$ of $\{X_t\}$ and its power spectral density function $S_X(\nu)$ as illustrated in Equation 1.14.

$$R_X(\tau) = \mathcal{F}^{-1}\{S_X(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) e^{\sqrt{-1} \cdot \tau \omega} d\omega \quad (1.14)$$

It appears that the two statistical measures of marginal and autocorrelation used as a pair have become well established, judging by a survey of the topical literature. The aim of most traffic models is to reproduce a specified marginal distribution and autocorrelation structure in the traffic, although most models by their nature cannot match both exactly and need to optimize one at the expense of the other. One can informally refer to the development of synthetic models based on matching a given set of marginal and autocorrelation characteristics as a “marginal-autocorrelation” modeling methodology.

1.4 Thesis Objective – The Characterization of Traffic

The simplest traffic model, the Poisson process, is fully described by a single parameter μ which specifies the mean rate of traffic flow per unit time. With only one degree of freedom the ability of the Poisson process to emulate the multitude of realistic traffic types is very limited. The sequence of interval batch counts $\{X_k\}$ produced by the Poisson process is Poisson in distribution, constraining the marginal distribution characteristic of the traffic to the same. The autocorrelation function $\rho(k)$ of $\{X_k\}$ is zero for all values $k \neq 0$ because it is an IID sequence, implying a total absence of time dependence structure.

The lack of time dependence was addressed by the introduction of Markovian processes. Markovian processes $\{X_t\}$ have a greater freedom to specify the marginal distribution, which can be anything from a Gaussian to uniform to arbitrary distributions depending on the level of sophistication. They introduce exponentially decaying

time dependence as measured by the $\rho(t)$ autocorrelation statistic. The autocorrelation function of these processes is a linear combination of complex exponential functions, which decays to 0 as time lag $t \rightarrow \infty$. The time dependence associated with the autocorrelation is of the SRD type since the Markovian process cannot produce second-order self-similarity. The Autoregressive Moving Average family of processes also produce complex exponential autocorrelation functions with good control, but are typically constrained to Gaussian form marginals [14].

Renewal processes can exhibit autocorrelation structure that is exponential, polynomial or sub-exponential. The advantage renewal models enjoy over Markovian ones is that they are not constrained to exponential decay, and can produce LRD self-similar time dependence. A wide range of marginal distributions can be accommodated from the Gaussian, as in the case of aggregated ON-OFF with Pareto renewals, to arbitrary ones as demonstrated by the SRP model. Lacking the analytic tractability of Markovian processes, renewal models are more difficult to analyze and fit by comparison.

The latest sophisticated processes to appear have been motivated by the goal of producing second-order self-similar LRD traffic. Processes such as FGN and FARIMA result in Gaussian marginal distributions although enhanced versions can specify arbitrary distribution, and the traffic shows very good LRD time dependence as measured by the autocorrelation. The parameters are typically easy to specify.

The evolution of traffic models has been motivated by the increasing complexity seen in the marginal and autocorrelation characteristics observed of empirical traffic. The marginal distribution might be highly skewed or asymmetrical in shape as opposed to a nominal Gaussian distribution. The autocorrelation, which has been relied upon extensively to measure the time dependence structure of empirical processes, has shown sub-exponential decay rates. While the models have improved their ability to emulate such complex processes this improvement has largely been measured against the same criteria, namely marginal and autocorrelation. The majority

of researchers have implemented a wide range of traffic models based on the steady-state marginal distribution and the autocorrelation function as the benchmark of modeling accuracy. There is a lack of research to validate and qualify the use of these two criteria in traffic characterization, and the marginal-autocorrelation modeling methodology is left unsupported as a result.

The question that has not been satisfactorily addressed is the appropriateness of modeling based on marginal-autocorrelation especially with respect to queueing phenomena. In queueing studies it is hesitantly assumed that two traffic streams with identical marginal and autocorrelation will produce the same statistics such as buffer occupancy and waiting time distribution in the queue. As a statistical measurement the autocorrelation function is appropriate for use in signal theory but is not necessarily so for the non-linear operation of a queueing system. The autocorrelation function does not capture so-called third-order and higher statistics [17, 14]. Furthermore the autocorrelation measurement only captures (in an information sense) the power spectrum of a given time series but not the phase spectrum, so clearly it is vital to question the validity of autocorrelation.

A growing body of research which has appeared in the last few years approaches the traffic-queue modeling problem by avoiding the use of autocorrelation. Of the references to be cited one [18] stands out due to its attempt to characterize traffic with the notion of entropy. Other work based on the low-band power spectrum, although derived from autocorrelation, uses the frequency domain viewpoint which is more appropriate for modeling traffic with long range dependence. Perhaps more appropriate still is the effective bandwidth function as a characterization of a traffic because it estimates *workload arrival* as opposed to signal composition. The log moment generating functions (which includes effective bandwidth) will be studied in this thesis as an alternate methodology for characterizing traffic. Because much of the academic work in the effective bandwidth area is of an analytical nature there is a need for experimentation and empirical study.

A large body of work by S. Q. Li et al. has presented evidence that traffic should be characterized by the marginal distribution and power spectrum function [19, 14, 20]. The power spectrum is equivalent to autocorrelation in a mathematical sense, but the approach towards time dependence from the frequency point of view is quite original. The essential point is that the total power measured in the low-frequency band of the spectrum characterizes the traffic and accurately predicts the queueing effects. The queue however is not ignored and its behavior particularly the service rate is included as part of the characterization method. Third and higher order characteristics not captured by power spectrum appeared to have negligible effects on the queue in this study [14]. The power spectrum draws attention to traffic with heavy time dependence (LRD in the extreme case) by an increasingly narrow concentration of power in the low frequency band about the origin. These observations have led to a practical traffic modeling tool SMAQ, which matches an empirical traffic's marginal and power spectrum and designs a Circulant Markov Modulated Poisson Process (CMPP) to reproduce the low band power spectrum to an arbitrary accuracy [21]. Another potential use for the power spectrum has been demonstrated for Call Admission Control [22, 23, 24].

In a complete departure away from the traditional characterization of stochastic processes as a signal, a more queue-oriented approach denoted “effective bandwidth” has in recent years attracted interest [25, 26, 27, 28]. Based on Large Deviation Theory, a process is approximated as a fluid source feeding a queue of infinite capacity. The notion of workload arrival to the queue is emphasized over the marginal distribution and time dependence structure of the traffic. The term effective bandwidth and the topic itself is still new but the approach is exciting because it conceptually recognizes the queue buffer, as opposed to the autocorrelation function which is a measurement of time varying signals with general application.

The use of autocorrelation, or rather the dual measures of marginal and autocorrelation, is not without its rationale. Any stochastic time series constrained to

a Gaussian (Normal) marginal distribution can be uniquely characterized by its autocorrelation function, due to the jointly-Gaussian covariance property [29]. If high speed aggregated traffic network links conform to these conditions then they would be amenable to traffic modeling based solely on the autocorrelation characteristic, and queueing results should be reliable. But it is risky and limiting to rely on this assumption in view that much compressed video and LAN traffic appear to exhibit non-Gaussian marginal distributions in addition to LRD time dependence structure [10, 30]. For this reason our research has concentrated on the queueing effects of a single traffic stream which has very non-Gaussian marginal distribution and LRD time dependence. Thus the simplifying assumptions of aggregation and SRD dependence are being shunned in favor of a harsh experiment, designed to expose the weaknesses of methods used in characterizing traffic. The objectives are simple:

- To show by example that the marginal distribution and autocorrelation function method is inadequate under harsh modeling conditions to characterize stochastic traffic for the purposes of queueing performance, and
- To give evidence that a measurement technique, with ideas borrowed from effective bandwidth and large deviation theory, is a better discriminator of traffic than marginal distribution and autocorrelation, and is empirically measurable.

The tasks of developing the simulation experiment and a set of synthetic models involved the following secondary objectives:

- To propose and implement an extension to the TES model which can produce hyperexponential autocorrelation, and to develop an analytical solution for its autocorrelation function,
- To implement an SRP model which can produce long autocorrelation with a few simple parameters, and secondarily to investigate and implement an MPEG generator and a Fractional Gaussian Noise generator based on the SRP model,

- To implement an object-oriented simulation program with custom built traffic and queue objects using a specialized commercial software language.

A trace of empirical traffic will be individually modeled by four different models: TES, ETES, SRP and NG-LRD. A computer simulation of a $G/D/1/\infty$ queueing system will be performed with the synthetic models, and the steady-state queue buffer size statistic will be measured as an output result. The output results together with the input traffic characterizations will provide evidence for discussion.

1.5 Thesis Organization

Chapter 2 surveys the traffic models used in current research from Poisson to self-similar, and discusses some general issues in traffic modeling. In Chapter 3 a hypothetical traffic specification is created that closely emulates the characteristics of compressed video traffic. The queue simulation experiment and the key measurements of interest are summarized with the aid of a block diagram and a table of the experimental variables.

The next four chapters concentrate on the synthetic models that are used for experiment. The quantity of report and analysis presented corresponds roughly to the amount of original research and development involved for each case. Melamed's TES model is reviewed in Chapter 4. In Chapter 5 the ARM model is reviewed and a specific implementation of an ARM model called Extended-TES is proposed. An analytical solution for the autocorrelation function of E-TES is developed and implemented in software. In Chapter 6 the self-similar NG-LRD model and generator is reviewed. Chapter 7 reviews the SRP model, and a version with simplified autocorrelation parameters is proposed to model LRD traffic. For each of the above four models a realization of the model, with parameters optimized to fit the reference traffic specification, is created for use in the simulation experiment.

Chapter 8 argues that in addition to the marginal distribution and autocorrelation functions which are used to describe traffic, characterization based on log moment generating class functions such as effective bandwidth has a sound basis. The formulae used to empirically measure the log moment generating functions of traffic are described.

Chapter 9 presents the various measurement results from the simulation experiment: marginal distribution and autocorrelation of the synthetic traffic, buffer occupancy probability of the queue, and the characterization of traffic using log moment generating functions. Chapter 10 summarizes the development of the thesis, rounds out the discussion of results and analysis, and hints at the potential for a new traffic modeling methodology.

MODSIM is an object-oriented simulation language used to implement the traffic-queue performance simulation experiments. Appendix A describes the objects that were created to incorporate the various traffic generators and queue types. Two secondary explorations in Appendices B and C make use of the SRP traffic model by proposing and implementing an MPEG IBP traffic modulator and an efficient FGN traffic generator.

Chapter 2

Background

2.1 Survey of Traffic Models

2.1.1 Poisson Models

The Poisson process $\{S_t\}$ is a step function which starts at $S_0 = 0$ and counts the number of arrivals over the time interval $(0, t]$. The arrival event is interpreted as the flow of a given data unit such as a packet or cell. Consider the segmentation of the time axis t into consecutive intervals of Δ length. The sequence of interval packet counts $X_n = S_{(n+1)\Delta} - S_{n\Delta}$ describes the traffic flow process, and by construction it has a Poisson marginal distribution.

Furthermore, the packet count in any one interval is guaranteed to be statistically independent of any other disjoint interval, regardless of interval length or rate parameter. Thus the $\{X_n\}$ sequence is IID. Equivalently, the interarrival times between successive packets have an exponential distribution which forms a complementary sequence $\{A_n\}$ which is also IID. Thus by nature the Poisson traffic model cannot and does not incorporate any form of time dependence structure into traffic flow. Arguably Poisson traffic can be used to model packet traffic in limited situations. Generally the Poisson distribution and similar distributions such as the Erlang have been very successful in many modeling areas such as telephony, manufacturing,

etc. but are not as appropriate for modeling flows of traffic with complex temporal structure [2].

2.1.2 Markovian Models

The family of Markovian processes encompasses a wide range of models, from simple ON-OFF sources to Markov Modulated Poisson Processes to Markov Arrival Processes. The latter is the generalized model and can in theory capture any process but is in practice difficult to design. The simplest, the ON-OFF source, can be defined with just three degrees of freedom corresponding to the mean rate, peak rate and autocorrelation decay rate. Markov Processes can be defined in continuous-time $\{X_t\}$ such as the exponential ON-OFF source, or in discrete-time $\{X_n\}$ an example being the Markov Chain. The state-space or range of values produced may be discrete as with a finite-state Markov Process, or continuous such as the TES model.

All Markov Processes share an exponential time dependence structure, that is, they all have autocorrelation function structures that are described by a linear combination of complex exponential functions decaying at various rates. The dominant decay rate shows itself in the tail part of the autocorrelation curve at long time lags. While this tail curve can be very long or hyperexponential, which implies long time scale dependencies, the corresponding process is still short range dependent (SRD) with finite autocorrelation energy. The statistical dependency between two separate points in time essentially vanishes given a time separation large enough, but this is not to say that the autocorrelation produced is inadequate to model real-life traffic. Ultimately, the extensive mathematical understanding of Markov Chains and their solvability by matrix techniques has made them attractive to many researchers.

Markovian ON-OFF sources [31, 32, 19] are used because their parameters are easy to capture, and for a given mean rate they tend to conservatively overestimate queueing effects. They function as an on-off gate which controls a simple traffic source

such as Poisson Arrival, Constant Rate Arrival or Fluid Rate models. Often independent ON-OFF sources are aggregated together to produce an output traffic with a Gaussian-like distribution that approximates multiplexed traffic. The Markov Modulated Rate Process (MMRP) and the slightly more sophisticated Markov Modulated Poisson Process (MMPP) are very popular. The output rate is state-dependent, as directed by an underlying finite-state Markov Chain. The Markov Chain is amenable to analysis because the eigenvalues of the transition matrix have a close relationship to the autocorrelation function of the process [21]. Papers in which Markov Modulated processes are used include [33, 34, 27, 23]. Of special interest to this thesis is the TES model and design methodology developed by Melamed [13] and employed in works by [35] and others. The TES model and its variants can exactly match the steady-state marginal distribution, and produce autocorrelation decay at least moderately hyperexponential in shape.

Markov Arrival (MAP) Processes and Batch Markov Arrival (BMAP) Processes [36, 37] have a more complicated probability kernel than the previous categories. The MAP Process is defined with reference to the past and current Markov state rather than only with respect to the past state, creating a more flexible but complicated Markovian process [28].

One characteristic of Markov Processes is that they produce sample paths which do not tend to exhibit time scale invariance, while many empirical traffics do. Some authors have attempted to create a more chaotically natural process based on Markov Chains but designed to generate output with a burstier and hence “more” scaling invariant nature. The approaches taken by [38, 39, 40] attempt to capture different aspects, such as short time scale behavior versus long term behavior separately, and integrate them back into one Markov Chain with multiple dimensions or hierarchical groupings of states. While this in theory appears to be an elegant solution, it is unclear how an empirical trace can be analyzed in terms of this complex Markov Chain structure in order to extract and fit the key parameters for the model.

2.1.3 Renewal Models

Renewal processes are more flexible than Markov processes because they are not constrained to generating sojourn times (renewal periods) that are exponentially distributed. By specifying a holding time distribution which is non-exponential, time dependence effects can be produced that are unattainable using their roughly equivalent Markov counterparts. Renewal processes can produce autocorrelation functions that decay exponentially to zero as do Markovian models or can abruptly become zero after some finite lag, resulting in SRD models. They can also be made Long Range Dependent (LRD) by a choice of holding time that is distributed with infinite variance such as the Pareto distribution.

A Renewal ON-OFF model changes state based on a transition matrix, but unlike a Markov Process one or both of its sojourn times is non-exponentially distributed. These models have gathered interest for their simplicity and for their ability to create traffic with heavy (even LRD) autocorrelation [41, 3, 42]. Similar to the Markov Modulated models are the semi-Markov models which change state according to an embedded Markov Chain but have non-exponential state holding times [43]. The Spatial Renewal Process developed by Jelenkovic et al. [44] is examined further in this thesis. It is capable of reproducing any steady-state marginal distribution and a wide range of monotonically decaying autocorrelation functions both SRD and potentially LRD as well. Other semi-Markovian models make use of holding times that are Erlang, hyper-exponential, and shifted-geometric [45, 46, 47]. The Pareto distribution (a polynomial function decaying slower than exponential) produces random variates with finite mean but infinite variance. MPEG traffic is modeled in [11] with scene durations described by a Pareto renewal process. Multiplexed point arrival processes having Pareto distribution are used in [30, 2] to model packet data traffic where it was observed that the arrival counts per slot generated by the models displayed LRD time correlation.

Renewal process models are likely to become as useful as Markovian models particularly in places where self-similar or LRD-like traffic is desired. Markovian and autoregressive models by definition are not LRD but appropriately designed Renewal models can be. They are more convenient to implement than Markovian models which are constrained to producing hyper-exponential autocorrelation in order to approximate the autocorrelation in LRD traffic.

2.1.4 Self-Similar Models

In addition to the LRD Renewal processes that have been examined, more sophisticated models such as Fractional Brownian Motion (FBM), Fractional Gaussian Noise (FGN) and Fractional Autoregressive Integrated Moving Average (F-ARIMA) have become popular for generating LRD traffic [9, 48, 49, 7]. These models require only a few parameters to be specified such as mean, variance and a Hurst value and they seem to produce natural looking traffic, but they are computationally demanding as compared to Renewal models [50, 30].

Self-similar or LRD traffic is important with respect to queueing models because of the effect on the buffer occupancy probability $\Pr[Q > x]$ especially in the asymptotic range for low probability values. While SRD traffic produces exponentially decaying tails, LRD traffic is expected to produce sub-exponential decay with higher overflow probabilities. Queueing model studies done with LRD traffic show as predicted that the queue occupancy probability has a considerably heavier tail compared to using an SRD traffic source, with implications for buffer sizing and other design issues [49, 12, 19, 2, 30].

2.1.5 Other Modeling Approaches

The Autoregressive Moving Average (ARMA) family of stochastic processes can be appreciated in terms of signal processing theory. An IID process is run through a

digital filter that imparts a linear autocorrelation structure onto the process determined completely by the filter's co-efficients [51]. Depending on the construction of the filter an Autoregressive (AR) or Moving Average (MA) process results [1]. Within the limits of the filter chosen, any autocorrelation function over a finite range of lags is easy to specify, but the marginal distribution which is not independently controllable tends to degenerate into a Gaussian shape [52, 14]. To overcome this drawback, a recent paper introduces the Autoregressive-To-Anything model which generates a foreground process with arbitrary marginal distribution by transforming an AR background process with a nonlinear function. The foreground and background processes do not share the same autocorrelation function due to the transformation, so it is necessary to numerically or analytically solve for the background process such that the foreground acquires the desired autocorrelation [53]. ARMA processes can produce arbitrarily long time dependence but at the cost of increasing filter size. They are however SRD and limited to producing asymptotically decaying autocorrelation like Markovian models, which is expected given that an AR model can be transformed into an equivalent Markov Modulated Process as noted in [28].

There are other approaches that are interesting or novel but have not been as popular. An ON-OFF Renewal model in [17] is controlled by a chaotic map that is designed to produce non-exponential renewal periods. The resulting traffic is LRD and approximates FBM when aggregated. In [18] traffic is characterized by its entropy which is estimated using a compression algorithm on the traffic series. The traffic is run through queueing simulations and the entropy is empirically measured both entering and leaving the queue and compared. Another method is diffusion which analytically or numerically approximates a discrete state traffic-queue process with continuous differential equations and boundary conditions [54, 47].

2.2 Network Traffic Issues

This thesis examines the appropriateness of traffic characterization methods, and does so in the context of a single-traffic single-queue system. But the characterization and synthetic reproduction of traffic is only one open issue in the area of traffic modeling. Researchers have also begun looking beyond single-queue traffic models to the network scale with flows of multiple traffic streams through queues connected in tandem and network fashion. There is a perceived need to simplify analysis of large scale network models, ideally by inventing a “linearized” mathematical system that would facilitate understanding and analysis, analogous to circuit theory. The works based on power spectrum characterization have been carried further to tandem queueing systems. There it was observed that a queue acts to filter out high frequency power above a critical frequency from a traffic stream passing through it, dependent upon the buffer size, service rate and traffic characteristics. The traffic in the low frequency band was observed to pass through the queue unattenuated and propagate onward through the downstream queues [55, 56]. The effective bandwidth approach holds the promise of analyzing stochastic packet switched networks as if they were simpler circuit switched networks. The effective bandwidth function $a(\theta)$ is defined such that the linear sum of effective bandwidths entering a queue must be less than the queue’s capacity $\sum a(\theta) \leq c$, much like the sum of circuits passing through a given circuit switched node [27, 57].

The difficulty in analytically or numerically analyzing large networks of queues stems from the statistical interaction among the queues and traffics on a global scale. Nodal decomposition attempts to analyze sub-systems or single queues by disregarding global statistical effects in order to make the analysis tractable although necessarily inaccurate. The amount of inaccuracy is determined by how much statistical interdependency is ignored while decomposing the problem into smaller pieces and is

a highly intuitive and non-systematic technique [43]. One set of experiments investigated the conditions necessary so that nodes can be analyzed in statistical isolation. It was found that provided the traffic entering all queues has peak rate less than 5% of the service rates, and that the output stream is dispersed with no more than 10% being directed to any one destination, nodal decomposition could be used to analyze the network [58]. Conditions for nodal decomposition are also examined from the effective bandwidth approach in [57]. For this purpose a modified effective bandwidth called decoupling bandwidth $a_D(\theta) > a(\theta)$ was used to characterize the traffic. It was found that provided the linear sum of decoupling bandwidths entering the queues are less than the capacities, nodal decomposition would be valid throughout the network.

Clearly the development of system level principles for network analysis must begin with a sound traffic characterization, and at this point it is not at all clear that the marginal-autocorrelation methodology of characterization meets this requirement. It is notable that the above exploratory works do not presume the marginal-autocorrelation methodology upon which to build their ideas. In one case the power spectrum and queue service capacity, and in the other the notion of effective bandwidth with respect to queue capacity, is used for describing the traffic.

Chapter 3

Experimental Design

3.1 Traffic Specification

In recent years it has been observed that compressed video and packet data traffic exhibit self-similarity or long range dependence (LRD) in the autocorrelation structure [3, 8, 10]. The steady-state marginal distribution is also likely to be non-Gaussian when considering cases of single or lightly aggregated streams due to the complexity underlying these types of traffic. For the modeling exercise we create a definition of a *generic traffic* which exhibits distribution and time-dependence features representative of real VBR video traffic. In keeping with the marginal-autocorrelation methodology, the design of the models will be based on matching the marginal distribution and autocorrelation function characteristics typical of compressed video.

A specification for the marginal distribution function was derived from empirical measurement of a 15000 sample trace of MPEG video, and was highly non-Gaussian. As seen in the specification (Figure 3.1) there is a large concentration of values at the low end but also a heavy tail extending for many lengths beyond the mean. Such a distribution was expected from the sample, and is not untypical of multimedia and network traffic [10, 30].

The autocorrelation function was based on the formulation used in [7] which specifies an exponential autocorrelation function e^{ak} for short lags followed by a slow

Figure 3.1: Steady-State Marginal Distribution Specification for Traffic

Figure 3.2: Autocorrelation Specification for Traffic

polynomial decay bk^c for long lags. The exponential curve defined over the range of lags $0 \leq k \leq 30$ was fit with parameter $a = -0.0415$ and based on the autocorrelation observed from the sample. (The empirical measurement of autocorrelation was made after cyclic components were removed from the MPEG sample. Attention has been focused in this thesis on the non-Gaussian and LRD aspects of VBR video rather than its periodic behavior.) For lags $k > 30$ the polynomially decaying tail was specified in terms of a desired amount of second-order self-similarity. The Hurst parameter was set to $H = 0.75$ which is a fairly typical amount of self-similarity for video and data traffic [15]. The autocorrelation function tail was defined by the curve bk^c with parameters $b = 1.5571$ and $c = -0.5$ chosen to produce an LRD effect with the desired Hurst value. The completed autocorrelation function specification is shown in Figure 3.2.

Figure 3.3: Autocorrelation of PseudoEmpirical Trace

While the marginal distribution and autocorrelation functions have provided a hypothetical description of traffic it was also desired to embody the traffic with a realization that one could refer back to as a control or reference trace. A 500000

sample long trace, denoted “PseudoEmpirical” because of its very natural looking yet artificially generated sample path, was generated by the NG-LRD generator. It matches the marginal specification in Figure 3.1 very well and the autocorrelation specification in Figure 3.3 satisfactorily given that statistical noise is present. However the main thrust of the experiment will not make use of “PseudoEmpirical” but it is a potentially useful device in further investigations. A comparison between the synthetic models amongst each other rather than with respect to the PseudoEmpirical control trace is the main purpose of the investigation.

3.2 Choice of Models

The chosen set of stochastic models comprises TES, E-TES, SRP and NG-LRD, where the first three produce SRD time-dependence while the last one is LRD by nature. A large range of model classes is represented, given that the TES model is Markovian, the E-TES model is autoregressive, the SRP is a Renewal process and the NG-LRD model is based on self-similar Gaussian Noise. The models differ from each other in interesting ways, for example the sample path of ETES versus that of SRP, and the true self-similar nature of NG-LRD versus the other three models. The first important criteria for choice of traffic model was the ability to match and generate any arbitrary marginal distribution. On this count all the four models under consideration were accepted. The second criteria, the reproduction of an autocorrelation function with LRD characteristics, was easily met by the NG-LRD model. The E-TES and SRP models were able to produce very long SRD autocorrelation, and proved to be satisfactory by matching the specification well up to 5000 lags in practice. Only the TES model with its nearly exponential autocorrelation decay could only approximate the required autocorrelation for a few hundred lags before dying out. In summary only the E-TES, SRP and NG-LRD models were selected for the core study however the TES model was not discarded in order to make some secondary comparisons to

the E-TES model, and subsequently a modified SRP model denoted SRPMOD was introduced.

3.3 Simulation Experiment

Synthetic traffic generated from each model will be put to test in a single queue simulation by computer. The simulation experiments consist of a fixed number of independently seeded runs, each run consuming a trace of fixed length generated by the model. The source-queue system does not attempt to model a specific transport mechanism such as an ATM link nor individual arrival times of cell-by-cell traffic. Rather it operates in a slotted-time mode, or equivalently, in discrete-time where each slot interval corresponds to an integer time.

The time axis is divided into consecutive slots of $\frac{1}{30}$ second length. The traffic generator produces a frame size value for each slot, which is interpreted as the amount of data produced during the slot period, and a quantity of cells are transmitted to the queue. Cell transmission, queue arrivals and queue service events take place only at the slot boundaries. Figure 3.4 illustrates by example the arrivals and departures of cell batches of varying size to a slotted “Lindley” queue. The decision to construct a slotted time queueing system as opposed to a system which would emulate a realistic (e.g. ATM) link was based on the gain in simulation speed and for compatibility with the autocorrelation and “normalized log moment generating” function measurements of the synthetic traffic. Appendix A provides more detail into the design and operation of the slotted Lindley queue simulation.

Figure 3.5 depicts the simulation experiment. The synthetic cell traffic is generated by the cell source using the traffic model and sent to the queue. Marginal distribution, autocorrelation and the log moment generating function measurements are taken from the traffic stream before entering the queue. The queue has an infinite buffer and a constant service rate which is set to provide a utilization factor of $1/3$.

Figure 3.4: Illustration of Slotted Queueing System

Figure 3.5: Block Diagram of Experiment

The random process $Q_t \in \{0, 1, 2, \dots\}$ records the size of the queue at time t , but because the system will be assumed to be in steady-state the subscript t is hereafter dropped. The key output result is the buffer occupancy statistic $\Pr[Q > x]$ which is estimated from the simulation runs in a time-averaged sense. The measurements are taken at various occupancy levels x and the results will be presented on a plot of $\log \Pr[Q > x]$ versus x .

EXPERIMENTAL PARAMETER	TYPE	VALUES
Model Type	Input (Controlled)	E-TES, SRP, or NG-LRD (Variable)
Marginal Distribution	Input (Controlled)	Non-Gaussian (Nominally Fixed)
Autocorrelation	Input (Controlled)	Exponential/Pareto Combination (Nominally Fixed)
Queue Behavior	Input (Controlled)	Infinite Buffer with Constant Rate Service (Fixed)
Run Length	Input (Controlled)	Fixed length generated traffic with 10% warmup (Fixed)
Buffer Occupancy Probability	Output (Observed)	Probability Measurements down to 10^{-5}
Log Moment Generating Function	Output (Observed)	Two Dimensional Measured Surface

Table 3.1: Parameters Considered in Main Experiment

The first experimental parameter is *model type* which is a controlled variable that assumes one of TES, ETES, SRP or NG-LRD. As far as experimentally practicable the other controlled input parameters were fixed. These include the *marginal distribution* and the *autocorrelation* functions of the traffic and the generated traffic trace *run length*. The *queue behavior*, limited to a constant service rate as its only parameter, is the last controlled input parameter. The measured output variable *buffer*

occupancy probability $\Pr[Q > x]$ is observed from the queue and is uncontrolled. The *log moment generating function* measured from the traffic is the second output variable and is likewise an observed parameter. Table 3.1 summarizes the variables under consideration.

In the core simulation experiment the E-TES, SRP and NG-LRD traffic models are studied. This experiment has been designed to demonstrate that the marginal-autocorrelation pair of traffic descriptors do not adequately characterize traffic and that different input models with the same apparent marginal-autocorrelation characteristics will produce vastly different queue effects. The potential of using the log moment generating functions to characterize traffic will be demonstrated in contrast to the weakness of the marginal-autocorrelation method. A secondary experiment will compare the output queue effects produced by the TES and E-TES models, providing more indirect evidence to suggest that marginal-autocorrelation does not relate well to queue effects. A third sub-experiment will compare the autocorrelation statistics of the NG-LRD to that of the SRPMOD traffic.

Chapter 4

The TES Model

The Transform-Expand-Sample (TES) model was developed by Melamed to capture the characteristics of a stationary stochastic process and generate synthetic trace versions of it [13, 59]. The modeling procedure is computationally expensive, but once a model has been created synthetic data can be generated quite efficiently. The two characteristics of the stochastic process that are captured by TES are the steady-state marginal distribution and the autocorrelation co-efficient function. By design a TES model is able to generate a stochastic process with any desired marginal distribution allowing it to exactly match that of any empirical trace in this respect. On the other hand TES cannot exactly match any given autocorrelation function (whether arbitrarily specified or measured from an empirical trace) so the synthetic model can only approximately capture and reproduce the desired autocorrelation characteristic. While these two characteristics are numerically quantifiable, the author points out that a third characteristic, the look of the sample path to the human eye, also requires attention during the overall modeling process even though it is not explicitly part of the TES model.

4.1 Review of TES⁺ Processes

The TES model comes in two flavors, TES⁺ and TES⁻. It was found from practical experience that the TES⁺ version is suitable for modeling network traffic while the

Figure 4.1: Example of TES Sample Path

TES⁻ version generally is not because it produces a zig-zag sample path. For this reason the TES⁺ version was selected. The background TES⁺ process is a sequence of random variables $\{U_n^+\}_{n=0}^\infty$ defined by the recursive formula:

$$U_n^+ = \begin{cases} U_0^+, & n = 0 \\ \langle U_{n-1}^+ + V_n \rangle, & n > 0 \end{cases} \quad (4.1)$$

where $\{V_n^+\}_{n=0}^\infty$ is a sequence of IID random variables known as the innovation. The sequence U_n^+ is by construction uniformly distributed on $[0, 1)$ and is a Markov Chain. The angular brackets $\langle \rangle$ denote the modulo-1 function which serves to constrain the state space of U_n^+ to the unit circle. From this background sequence the output process of the model referred to as the foreground sequence $\{X_n^+\}_{n=0}^\infty$ is created by running each U_n^+ through a distortion function D .

$$X_n^+ = D(U_n^+) = H_X^{-1}(S_\epsilon(U_n^+)) \quad (4.2)$$

$$S_\epsilon(y) = \begin{cases} y/\epsilon, & 0 \leq y < \epsilon \\ (1-y)/(1-\epsilon), & \epsilon \leq y < 1 \end{cases} \quad (4.3)$$

where $0 \leq \epsilon \leq 1$

The function D is composed of a stitching transformation S_ϵ followed by the inverse of the function $H_X(x)$. The function H_X is in fact the marginal cumulative distribution function of the desired output process, and can be arbitrarily specified

or based on a given empirical trace. As Melamed points out in [13] that since the U_n^+ sequence is uniform on $[0, 1)$ it is transformed into an output process X_n^+ with distribution described by H_X and corresponding density function as shown in Equation 4.4.

$$f_X(x) = \frac{d}{dx} H_X(x) \quad (4.4)$$

The autocorrelation co-efficient function $\rho_X(k)$ of X_n can be calculated once the distortion function D and the innovation random variable PDF f_V are defined. The analytical formula is:

$$\rho_X(k) = \frac{2}{\sigma_X^2} \sum_{\nu=1}^{\infty} \Re[\tilde{f}_V^k(i2\pi\nu)] |\tilde{D}(i2\pi\nu)|^2 \quad (4.5)$$

The functions $\tilde{f}_V$ and $\tilde{D}$ are Laplace transforms of the innovation and distortion functions respectively, and σ_X^2 is the steady-state variance of X_n which can be easily deduced from H_X .

One should further interpret and illustrate the role played by the $\tilde{f}_V^k$ term in Equation 4.5 because it will prove useful in developing an analogous formula for Extended TES. Equation 4.1 implies that any two background values U_i and U_j with $0 \leq i < j$ are related by

$$U_j = \langle U_i + V_{i+1} + V_{i+2} + \cdots + V_j \rangle \quad (4.6)$$

There are $j - i$ innovation variables summed together to produce a total innovation from U_i to U_j . This total innovation V_T can be interpreted as the “innovation distance” between any two background variables U_n separated by a lag distance of $k = j - i$. Visually the innovation distance is the sum of all innovation variables contained within the vertical bars as can be seen in Table 4.1 for the case of lag $k = 5$. In Equation 4.5 the function $\tilde{f}_V^k$ loosely corresponds to the characteristic function of the total innovation V_T because the sum of k IID variables V_n is given by convolving the PDF f_V by itself k times.

$$\cdots \quad V_{n-2} \quad V_{n-1} \quad \left| \quad V_n \quad V_{n+1} \quad \begin{matrix} V_{n+2} \\ = V_t \end{matrix} \quad V_{n+3} \quad V_{n+4} \quad \right| \quad V_{n+5} \quad \cdots \quad (4.7)$$

Table 4.1: Background Innovation Process

The process of creating a TES model is centered around Equations 4.4 and 4.5. The steady-state cumulative distribution H_X is defined using empirical data or other means provided that it is a valid CDF. A value between 0 and 1 is chosen for the stitching factor ϵ . Most of the difficulty takes place in choosing an innovation function f_V such that the autocorrelation function calculated in Equation 4.5 is made to match the desired autocorrelation.

The TES output process is not strictly Markovian but for practical intents and purposes can be considered as such due to the inheritance of Markov-like properties from its background process. Practical experience has confirmed that TES models produce autocorrelation functions that are exponential/sinusoidal in shape and that decay exponentially in magnitude for large lags. This is acceptable for modeling empirical processes that have short range dependence in time, characterized by finite autocorrelation energy. However it has been observed that much traffic is long range dependent (LRD) having heavy tailed curves. One possible solution for this is to develop models that are inherently long range dependent with polynomially decaying correlation in time, such as Huang's model [7]. Another way is to enhance exponentially decaying models into hyper-exponential ones that decay quickly at early lags but slowly as lags increase. One suspects that the TES model is theoretically capable of generating output with such hyper-exponential decay due to the similarity of the background process of TES, which is a continuous state Markov Process, with that of S. Q. Li's discrete circulant-matrix MMRP [21]. The latter model can generate traffic with potentially very large low frequency correlation that presumably produces long tailed hyper-exponential autocorrelation. But experience has shown that it is impractical to extract hyper-exponential autocorrelation behavior out of TES in its present form.

4.2 Specification of Experimental Study Model

A TES model was created against the marginal distribution and autocorrelation specifications introduced in Chapter 3. The software tool TEstool automates the process of creating a TES model given the marginal and autocorrelation requirements [59]. The program features a fully automatic search algorithm that creates the model. It applies a global search followed by a local search formula to find an optimal solution for the three components making up the model: a marginal distribution function H_X , an innovation function f_V and a stitching value ϵ .

Figure 4.2: TES Model Innovation PDF and Stitching Functions

Some of the default parameters pertaining to the search algorithm were changed from their default settings to improve the search-space based on an extensive experience with the TEstool program. The innovation function step count was reduced from 100 to 32, while the innovation value quantization was increased from 2 to 4. The fully automatic search algorithm was run for several hours in order to produce the model. Since the TES model by construction matches the marginal distribution specification exactly, the distribution function H_X fully responsible for this

aspect was assumed to be optimal in the created model. The components responsible for producing autocorrelation, f_V and ϵ are shown after optimization by TESTool in Figure 4.2, and the model parameters are summarized in Table 4.2.

$f_V =$	Step function on $[-\frac{1}{2}, \frac{1}{2})$
$H_X =$	Empirically exact CDF
$\epsilon =$	0.7253

Table 4.2: TES Model Input Parameters

Figure 4.3: Autocorrelation Spec versus TES Analytical

The analytically predicted autocorrelation function of the model is shown in Figure 4.3. The match between the required (Specification) and predicted (Analytical) autocorrelation curves is clearly poor, especially for greater lags where the model autocorrelation decays to zero. However given the limitations of the search algorithm and computer processing power the result is as good as can be expected. Yet this poor result presents an opportunity for comparison with the E-TES model which is based on the TES model but greatly improves the autocorrelation fit.

Chapter 5

The E-TES Model

The TES model has been recently extended by Melamed into the more general Autoregressive Modular Process [60]. The ARM family of models is much more flexible than TES because it allows complex and potentially long range autocorrelation structures to be modeled while TES is limited to short range exponential autocorrelation.

5.1 Description

We restrict our discussion to the ARM⁺ version and define the background process U_n^+ similarly to that of TES⁺

$$U_n^+ = \begin{cases} U_0^+, & n = 0 \\ \langle U_{n-1}^+ + V_n \rangle, & n > 0 \end{cases} \quad (5.1)$$

The difference between this equation and Equation 4.1 is that the innovation sequence V_n can no longer be assumed IID nor even stationary. Throughout this section, though, it shall be assumed that V_n is a stationary sequence. The change to V_n does not affect Equations 4.2, 4.3 or 4.4 used in the TES model which pertain to the steady-state marginal distribution of the output X_n . The marginal distribution can be modeled in exactly the same way as is done in TES. The change however does affect Equation 4.5 and the autocorrelation function formula. In [60] the f_V^k PDF function is replaced by f_{S_k} where S is the partial sum of the innovation variables as

Figure 5.1: Example of Moving-Average Filter

defined in Equation 5.2 and used in Equation 5.3.

$$S_k = \sum_{i=1}^k V_i \quad (5.2)$$

$$\rho_X(k) = \frac{2}{\sigma_X^2} \sum_{\nu=1}^{\infty} \Re[\tilde{f}_{S_k}(i2\pi\nu)] |\tilde{D}(i2\pi\nu)|^2 \quad (5.3)$$

The ARM⁺ model in contrast to the TES⁺ model is no longer a Markov process because of the statistical dependence in time of the innovation sequence V_n . This frees the ARM model from the exponentially decaying short range dependent structure that TES is limited to. The problem at hand is to design a stationary, time dependent series of innovations V_n that produce a highly hyper-exponential autocorrelation structure in the output X_n . The choice should be easy to implement and efficient in the same vein as the TES model. One possible solution considered was to describe and generate V_n using a Markov chain but it was not pursued in search of a method simpler to analyze. We chose a method that begins with an IID sequence Z_n which is fed into a causal moving-average filter as defined in Equation 5.4 and pictured in Figure 5.1 resulting in the innovation sequence V_n .

$$V_n = \sum_{i \geq 0} \alpha_i Z_{n-i} \quad (5.4)$$

Since the formula describes a finite impulse response filter the resulting system is guaranteed to be stable. In addition, the computational effort required to generate

$\alpha_1 Z_{n-1}$	$\alpha_1 Z_n$	$\alpha_1 Z_{n+1}$	$\alpha_1 Z_{n+2}$	$\alpha_1 Z_{n+3}$	$\alpha_1 Z_{n+4}$	$\alpha_1 Z_{n+5}$	$\alpha_1 Z_{n+6}$
$\alpha_2 Z_{n-3}$	$\alpha_2 Z_{n-2}$	$\alpha_2 Z_{n-1}$	$\alpha_2 Z_n$	$\alpha_2 Z_{n+1}$	$\alpha_2 Z_{n+2}$	$\alpha_2 Z_{n+3}$	$\alpha_2 Z_{n+4}$
$\alpha_3 Z_{n-6}$	$\alpha_3 Z_{n-5}$	$\alpha_3 Z_{n-4}$	$\alpha_3 Z_{n-3}$	$\alpha_3 Z_{n-2}$	$\alpha_3 Z_{n-1}$	$\alpha_3 Z_n$	$\alpha_3 Z_{n+1}$
V_{n-1}	V_n	V_{n+1}	V_{n+2}	V_{n+3}	V_{n+4}	V_{n+5}	V_{n+6}
			$= V_T$				

Table 5.1: Background Innovation Process for Extended TES

V_n is small and the mathematical analysis of the autocorrelation (to be demonstrated in the next section) is fairly mechanical.

5.2 Autocorrelation Co-efficient Function

This section derives an analytical formula for the autocorrelation co-efficient function of the Extended TES (E-TES) model with innovations generated using the causal filter of Equation 5.4. The key to applying the general analytical formula for $\rho_X(k)$ introduced in Equation 5.3 is to find the PDF function of the partial sum random variable S_k . To emphasize the notion of innovation distance we will substitute the variable notation S_k with $V_T(k)$ and likewise replace $\tilde{f}_{S_k}$ with $\tilde{f}_{V_T}$.

For the purpose of illustration assume that the innovation variable V_n has been constructed using a filter with $N = 3$ terms

$$V_n = \sum_{j=1}^N \alpha_j Z_{n-j}, \quad (5.5)$$

where the innovation filter co-efficient list Λ is

$$\begin{aligned} \Lambda &= \{(\epsilon_1, \alpha_1), (\delta_2, \alpha_2), (\delta_3, \alpha_3)\} \\ &= \{(0, \alpha_1), (2, \alpha_2), (5, \alpha_3)\} \end{aligned} \quad (5.6)$$

Each V_n is the weighted sum of three Z variables tapped at different delay times and can be visually represented by the array in Table 5.1. The innovation distance V_T between two background sequence values U_i and U_j is demonstrated here for the case of lag $k = j - i = 6$ by the positioning of two vertical bars six columns apart. The

innovation distance is the sum of all the V , or equivalently the weighted sum of all the Z variates within the bars.

$$\begin{aligned}
 V_T(6) &= V_n + V_{n+1} + V_{n+2} + V_{n+3} + V_{n+4} + V_{n+5} \\
 &= (\alpha_1)Z_{n+4} + (\alpha_1)Z_{n+5} + \\
 &\quad (\alpha_1 + \alpha_2)Z_{n+1} + (\alpha_1 + \alpha_2)Z_{n+2} + (\alpha_1 + \alpha_2)Z_{n+3} + \\
 &\quad (\alpha_1 + \alpha_2 + \alpha_3)Z_n + \\
 &\quad (\alpha_2 + \alpha_3)Z_{n-2} + (\alpha_2 + \alpha_3)Z_{n-1} + \\
 &\quad (\alpha_3)Z_{n-5} + (\alpha_3)Z_{n-4} + (\alpha_3)Z_{n-3}
 \end{aligned} \tag{5.7}$$

Equation 5.7 expresses V_T as a sum of terms composed of variables Z_{n-5} through Z_{n+5} with various weighting constants. A term is defined as a random variate Z multiplied by a sum of co-efficients of the form $(\alpha_i + \dots + \alpha_j)$. The PDF function of the random variable V_T is calculated from the convolution of the individual PDFs corresponding to each term as shown below in Equation 5.8. The same expression in the Laplace transform domain is shown in Equation 5.9.

$$\begin{aligned}
 f_{V_T} &= f_{\alpha_1 Z} * f_{\alpha_1 Z} * f_{(\alpha_1 + \alpha_2) Z} * f_{(\alpha_1 + \alpha_2) Z} * f_{(\alpha_1 + \alpha_2) Z} * \\
 &\quad f_{(\alpha_1 + \alpha_2 + \alpha_3) Z} * \\
 &\quad f_{(\alpha_2 + \alpha_3) Z} * f_{(\alpha_2 + \alpha_3) Z} * f_{\alpha_3 Z} * f_{\alpha_3 Z} * f_{\alpha_3 Z}
 \end{aligned} \tag{5.8}$$

$$\begin{aligned}
 \tilde{f}_{V_T} &= (\tilde{f}_{\alpha_1 Z})^2 (\tilde{f}_{(\alpha_1 + \alpha_2) Z})^3 \tilde{f}_{(\alpha_1 + \alpha_2 + \alpha_3) Z} \\
 &\quad (\tilde{f}_{(\alpha_2 + \alpha_3) Z})^2 (\tilde{f}_{\alpha_3 Z})^3
 \end{aligned} \tag{5.9}$$

Algebraic simplification can be performed by collecting identically distributed random variable terms and expressing them as powers of Laplace transform functions, as can be seen in the progression from Equation 5.7 through to Equation 5.9. For example, the term $(\alpha_2 + \alpha_3)Z$ appears twice in Equation 5.7 once with time index $n-2$ and once with index $n-1$, which combined together result in the power $(\tilde{f}_{(\alpha_2 + \alpha_3) Z})^2$

as a Laplace transform expression. The algebraic simplification step can be visually interpreted in Table 5.1 as the identification of Z variates grouped by weighting constant. In this concrete example there can be seen one term of $(\alpha_1 + \alpha_2 + \alpha_3)Z$, three terms of $(\alpha_1 + \alpha_2)Z$ and two terms of $(\alpha_2 + \alpha_3)Z$, two terms of $\alpha_1 Z$ and finally three terms of $\alpha_3 Z$. The key problem is to automatically identify how many terms exist of each weighting constant of the form $(\alpha_i + \dots + \alpha_j)$ where $i \leq j$. The solution furthermore depends on the value of k . Configurations for various k and filter size N were manually analyzed this way on paper until a pattern leading to a solution could be developed.

Given a lag k , and a range of integers $\{i \dots j\}$ such that $1 \leq i \leq j \leq N$, the following basic formula gives the number of independent variates of form $(\alpha_i + \dots + \alpha_j)Z$ contained inside the total innovation $V_T(k)$:

$$Q_{ij}(k) = [k - (\delta_j - \delta_i) - R_{ij}(k)]^+ \quad (5.10)$$

Variable $Q_{ij}(k)$ is a non-negative integer specifying the number of independent terms that exist and $R_{ij}(k)$ is a count adjustment value. The problem of “double counting” terms arises when a particular term is a “subset” of a larger term and is counted in both categories. For example there are a total of three $(\alpha_2 + \alpha_3)Z$ terms that can be identified in Table 5.1 but one of them (Z_n) belongs to the term $(\alpha_1 + \alpha_2 + \alpha_3)Z$ and must be explicitly discounted from the former set. While it is possible to calculate by hand the values of $Q_{ij}(k)$ and $R_{ij}(k)$ for every lag $k \geq 0$, it became necessary to generalize and automate the calculation for practical application! The trial calculations by hand uncovered a pattern in the way the sums of co-efficients α were formed to produce the Z variate weighting constants.

5.2.1 General Algorithm

We used a directed tree graph with numbered nodes, and assigned a mapping from every possible weighting constant of form $(\alpha_i + \dots + \alpha_j)$ to a unique node p where

Table 5.2: Pascal's Triangle and Mapping

$1 \leq p \leq N(N+1)/2$. Table 5.2.1 shows a graph constructed with $N = 3$ as in the previous example, and lists the corresponding mapping table. In general for a co-efficient set $\{\alpha_1, \dots, \alpha_N\}$ the triangle node mapping function $\Gamma_N : \{i, j\} \rightarrow p$ is defined to be

$$\Gamma_N(i, j) = ((N - (j - i) - 1)(N - (j - i))/2 + i \quad (5.11)$$

$$1 \leq i \leq j \leq N$$

With this mapping the old nomenclature $Q_{ij}(k)$ is replaced with $Q_p(k)$ where the subscript p refers to the node that uniquely identifies a term $(\alpha_i + \dots + \alpha_j)Z$. $R_{ij}(k)$ is similarly replaced by $R_p(k)$. The Pascal's Triangle graph is conveniently represented by a connectivity matrix C of size of $N(N+1)/2$ rows by $N(N+1)/2$ columns.

$$C_{ab} = \begin{cases} 1 & \text{there is path from node } a \text{ to } b \\ 0 & \text{otherwise} \end{cases} \quad (5.12)$$

In other words the row vector $\vec{C}_p$ identifies all the nodes that are in the forward path of node p . This is significant in that it equivalently specifies which $(\alpha_i + \dots + \alpha_j)Z$ terms already contain the $(\alpha_i + \dots + \alpha_j)Z$ term in question. By using this method to determine how many terms have been doubly or multiply counted, one can subtract the adjustment value R_p from the total count $k - (\delta_j - \delta_i)$. The computation of the $\vec{Q}(k)$ vector for a given lag k relies on iteratively computing the following two

formulas starting with $p = 1$.

$$R_p = \vec{C}_p \cdot \vec{Q} \quad (5.13)$$

$$Q_p = [k - (\delta_j - \delta_i) - R_p]^+ \quad (5.14)$$

$\vec{Q}$ is a vector of length $N(N+1)/2$ of Q_p values. The calculation of the elements in $\vec{Q}$ and $\vec{R}$ is recursive in nature but the form of the graph allows one to iteratively calculate the vector elements starting from $p = 1$ to $p = N(N+1)/2$.

Given	$\Lambda = \{(\delta_1, \alpha_1), \dots, (\delta_N, \alpha_N)\}, k, f_Z(z)$
$C \leftarrow$	Graph connectivity matrix $N(N+1)/2$ nodes
$\vec{Q} \leftarrow$	$\vec{0}$
$\vec{R} \leftarrow$	$\vec{0}$
for	$p = 1 \dots N(N+1)/2$
	$(i, j) \leftarrow \Gamma^{-1}(p)$
	$R_p \leftarrow \vec{C}_p \cdot \vec{Q}$
	$Q_p \leftarrow [k - (\delta_j - \delta_i) - R_p]^+$
	$\tilde{f}_p \leftarrow \mathcal{L}\{f_{(\alpha_i + \dots + \alpha_j)Z}\}$
end	
	$\tilde{f}_{V_T} \leftarrow \prod_{p=1}^{N(N+1)/2} (\tilde{f}_p)^{Q_p}$
	$\rho_X(k) \leftarrow \frac{2}{\sigma_X^2} \sum_{\nu=1}^{\infty} \Re[\tilde{f}_{V_T}(i2\pi\nu)] \tilde{D}(i2\pi\nu) ^2$

Table 5.3: Analytical Calculation of Autocorrelation

The PDF $f_Z(z)$ of the random variate Z and its Laplace transform $\tilde{f}_Z$ is known. The Laplace transforms of the density functions of the random variables $(\alpha_i + \dots + \alpha_j)Z$ can thus be evaluated beforehand, by applying the relationship $f_{\alpha Z}(z) = \frac{1}{|\alpha|} f_Z(z/\alpha)$. An expression for the Laplace transform of the total innovation distance $V_T(k)$ is given by Equation 5.15.

$$\tilde{f}_{V_T} = \prod_{p=1}^{N(N+1)/2} (\tilde{f}_p)^{Q_p} \quad (5.15)$$

where $f_p(x)$ is the PDF function of the random variable $(\alpha_i + \dots + \alpha_j)Z$ and $\tilde{f}_p$ is its Laplace transform. The reverse mapping $\Gamma^{-1} : p \rightarrow \{i, j\}$ can be used to find $(\alpha_i + \dots + \alpha_j)$. The numerical evaluation of Equation 5.15 is roughly of order $O(N^2)$ based on the growth rate of the graph with respect to N .

5.3 Implementation

This is a list of the MATLAB routines created to numerically evaluate the autocorrelation co-efficient function of the Extended TES model using the analysis described in Section 5.2. It is assumed that the steady-state cumulative distribution function H_X , the stitching factor value ϵ , and the stochastic process Z_n have already been defined. The Z_n series, which is used to generate the innovations V_n , is an IID sequence of random variables described by the CDF function F_Z . The innovation filter list Λ is a set of one or more (δ, α) pairs used to transform the sequence Z_n into the innovation variable V_n .

- The function `triConnectMatrix` creates a connectivity matrix C for a directed graph constructed like Pascal's triangle. The input argument N is the size of the graph and the returned matrix has size $N(N+1)/2 \times N(N+1)/2$.
- The function `stitchf` applies the stitching factor ϵ to the inverse of the steady-state distribution H_X^{-1} and returns the stitched distortion function D . Both the input and returned functions are of type `PWfunction`.
- Function `pwftodiscrete` converts a function of type `PWfunction` into a discrete function of type `discrete` with a given number of points. It is specifically used here to discretize the distortion function D .
- The function `scaleInnovCdf` converts the CDF function of the random variable Z_n into the CDF function of the scaled version of the variable $(\alpha)Z_n$. The input

arguments are the CDF function F_Z in `PWfunction` format and a scaling value α . The returned function is also `PWfunction`.

- The routine `cdftomodulopdf` converts a given CDF function of type `PWfunction` into a PDF function which is discrete. The PDF resulting function $f(x)$ is constrained to a period of 1 such that any values outside of the range $x \in [0, 1]$ are “folded over” in a periodic sense.
- Function `fs` performs a Fourier Series transform on the input function. The input function is of discrete type and furthermore is assumed to have a period of 1 although the starting and ending points do not necessarily have to be 0 and 1 respectively. The list of coefficients returned is cut off at the Nyquist frequency which depends on the quantization of the input function.
- The routine `innovMatrix` returns a vector of $N(N + 1)/2$ functions of type `PWfunction`. Every random variable term of the form $(\alpha_i + \dots + \alpha_j)Z$ is converted into its corresponding Laplace-transformed PDF function $\tilde{f}_p$ and stored for future use by the `etesautocov` routine. The input arguments required are function F_Z , a quantization parameter, and the set Λ .
- Routine `etesautocov` numerically calculates the autocorrelation function for a set of lags k ($k \geq 0$) using Equation 5.15. The required input arguments are the set Λ , the innovation vector from `innovMatrix`, the discretized distortion function D and the connectivity matrix C .

The routines as a whole are efficient in their use of MATLAB’s memory space, and can numerically calculate one autocorrelation value in a few seconds running on a fast Sun machine.

5.3.1 Extended TES Generator Program

An Extended TES generator program was written in C language to create synthetic traces from the model. The executable program is run from the UNIX shell using the format:

```
cetes etesmodelfilename tracelength [seed] > outputfile
```

The input model file follows the same format as the TESTool model file format, with the addition of a single line at the file's end to specify the filter co-efficient list Λ . For example:

```
#
#Innovation for TES model: pseudoEmpirical.data.model1 --- created on 9/8/96
#
SIGN = +1
...
...
E-TES INNOVATION COEFFICIENTS = (0,0.900) (40,-0.470) (150,-0.185)
<EOF>
```

5.4 Specification of Experimental Study Model

One of the purposes in developing the Extended TES model was to apply it in a queueing simulation study and compare the output results to that of other models. A standard TES⁺ model had already been designed using the TESTool automated program. However it was unable to capture the LRD heavy-tailed characteristic which became apparent when the specification autocorrelation curve was compared to the one produced by TES back in Figure 4.3.

An Extended TES model was designed based on the TES model, in particular by borrowing the same distortion function D and stitching factor ϵ from TES, but requiring a new definition for the innovation sequence V_n . The autocorrelation produced by the TES model provided a good starting point towards the desired autocorrelation shape. By noting that the V_n process defined in the TES model is an IID series, it was natural to redefine it as the Z_n in Extended TES and then attempt to reshape the resulting autocorrelation by manipulating the filter co-efficients Λ . If Λ

should be set to $\{(0, 1.0)\}$ then $V_n = Z_n$ and the Extended TES model would behave identically to the TES model. After observing the quick exponential decay of the desired pseudoempirical curve and its slowly decaying polynomial tail it was clear that a minimum filter size of $N = 2$ terms would be necessary. To be on the safe side the filter Λ was defined with $N = 4$ terms as shown below:

$$\Lambda = \{(\delta_1, \alpha_1), (\delta_2, \alpha_2), (\delta_3, \alpha_3), (\delta_4, \alpha_4)\}$$

An optimal solution for Λ was sought for by a combination of inspection, intuition and a global search routine that optimized for the autocorrelation function in a mean square error sense. A makeshift routine was written in MATLAB to perform the global search for Λ . An ideal search would have included the whole solution space spanning values of both δ and α , but for simplicity the routine iterates over ranges of α with fixed δ , the expectation being that the human user will adjust δ as necessary. The filter delay co-efficients $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ were chosen by inspection and trial-and-error to the values $\{0, 40, 100, 500\}$. The weighting co-efficients $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ were initially set to search the range $[-1, +1]$. After a series of refinements an optimal solution for Λ as shown in Table 5.4 was found. Figure 5.2 compares the autocorrelation specification with that analytically predicted for the Extended TES model.

$f_Z =$ Same as TES model f_V
$\Lambda = \{(0, 0.875), (40, -0.3985), (100, -0.2505), (500, -0.1235)\}$
$H_X =$ Same as TES model
$\epsilon =$ Same as TES model

Table 5.4: E-TES Model Input Parameters

5.5 Verification

In an early test, an E-TES trace of length 135×10^3 was generated and the empirical autocorrelation statistics (not shown here) were measured from the trace series. The

Figure 5.2: Autocorrelation Spec versus E-TES Analytical

fact that the analytical and empirical autocorrelations were nearly identical provided confidence in the mathematics of Section 5.2, its implementation by the MATLAB routines, and the C program used to generate traces of the model. An example of the sample path generated by the Extended model given in Figure 5.3 can be compared to the similar appearance of the TES sample path in Figure 4.1. For completeness we

Figure 5.3: Example of ETES Sample Path

present the steady-state marginal distribution of a TES sample path and an Extended

TES sample path. As is seen in Figure 5.4 both models produce a nearly equivalent distribution, as it should be because the distribution is defined independently of the innovation process in both TES and ARM models.

Figure 5.4: Histograms of TES and E-TES

An interesting feature seen in the autocorrelation function of the E-TES model is that of the smooth exponential sections which connect at sharp knee points or points of inflection. The knee points of the analytical curve in Figure 5.2 appear to be at lags $k = 40, 100, 400$ and 500 , and make sense in that the composition of the innovation V_T changes qualitatively between lag values k and $k + 1$ as given by:

$$k \in \{ |\delta_i - \delta_j| \mid \forall \{i, j\} \mid 1 \leq i, j \leq N \} \quad (5.16)$$

Formula 5.16 also predicts the knee points at $k = 60$ and 460 which are also observed in the figure.

5.6 Conclusion

The evolution or extension of the TES model to the E-TES model by making use of an FIR filtered innovation sequence seems to be a natural one. The extension has effect on the autocorrelation of the output sequence, accurately producing hyperexponentially decaying autocorrelation which is accurately predicted by analytical method. The design process begins with a TES model exhibiting exponentially decaying autocorrelation and uses a brute-force search algorithm to find appropriate E-TES FIR parameters to modify it.

Chapter 6

The NG-LRD Model

The traffic model developed by Huang [7] is the best suited of the models for producing long range dependent (LRD) traffic. It is also the easiest to use because only three parameters are required to specify the time dependence structure. The model incorporates a Gaussian distributed LRD background process X_k which is responsible for producing time-dependence in the stream. The background is transformed into a foreground process Y_k which assumes the desired marginal distribution. The combination of FGN-like background and distortion function will be referred to as the NG-LRD model. The main drawback to NG-LRD is the computational effort required for its generation, which grows non-linearly with number of samples generated.

Figure 6.1: Example of NG-LRD Sample Path

6.1 Description

The background process X_k is an asymptotically second-order self-similar or LRD process with a Gaussian steady-state marginal distribution. The time-dependence structure is described by the autocorrelation function $\hat{\rho}(k)$ which is composed of an exponentially decaying segment followed by a polynomially decaying tail. The exponential part corresponding to the effects of short range dependence (SRD) is described by the curve e^{ak} for lags $0 \leq k < K_t$ while the tail section bk^c for lags $k \geq K_t$ corresponds to LRD dependence. The constant K_t appears as a “knee” in the curve between the two sections.

The foreground or output process is derived from the background process by distorting generated values through a function h producing an output Y_n with arbitrary marginal distribution. The distortion function h is a tandem combination of a Gaussian CDF F_X followed by the output process CDF F_Y :

$$Y_k = h(X_k) = F_Y^{-1}(F_X(X_k)) \quad (6.1)$$

The output process inherits the time-dependence and LRD nature of the background process. Despite the fact that the autocorrelation and time-dependence structure is degraded by the transformation through h , provided the transformation is reasonably well behaved the error can be compensated for. The output process autocorrelation function is given by:

$$\rho(k) = \begin{cases} \alpha_k e^{ak}, & 0 \leq k \leq K_t \\ \alpha_k bk^c, & k \geq K_t \end{cases} \quad (6.2)$$

The autocorrelation $\rho(k)$ decays exponentially up to the lag K_t after which it follows a polynomial decay. The α_k term, introduced here to demonstrate the attenuation effect, is assumed to be a slowly changing function starting from unity and settling to some limiting value $0 < \alpha_\infty < 1$ as k approaches infinity. In other words, the output autocorrelation function $\rho(k)$ may be viewed as a degraded version of the background autocorrelation $\hat{\rho}(k)$ because of the reduction in its magnitude.

a	$=$	-0.03608
K_t	$=$	30
H	$=$	$0.75 \Rightarrow c = -0.5, b = 1.8556$
$F_Y(y)$	$=$	Output Marginal Distribution CDF

Table 6.1: NG-LRD Model Input Parameters

The NG-LRD generator accepts the three parameters: a , K_t and a self-similarity value called the Hurst parameter $\frac{1}{2} < H < 1$. The exponential section e^{-ak} is defined by the a and K_t parameters while the polynomial section bt^c is succinctly described by H where c is given by the relationship: $c = 2H - 2$.

6.2 Specification of Experimental Study Model

The output distribution function $F_Y(y)$ was constructed in a straightforward manner by creating a piecewise function to match the histogram specification. The time-dependence structure of the NG-LRD model, an exponential decay followed by a polynomial decay, was particularly convenient in this exercise because of its similarity to the autocorrelation specification.

The first modeling attempt used parameters borrowed directly from the specification curve itself by setting $a = -0.04150$, $K_t = 30$ and $H = 0.75$. The empirically measured autocorrelation was smaller in magnitude than desired because of the attenuation effect. It was observed that the attenuation α reduced output autocorrelation by a factor of 0.85 at $k = 30$ lags and approached 0.7 as lags became very large. To provide for good matching over short lags at a slight sacrifice to large lags, a compromise was found by assuming a constant attenuation factor of $\alpha = 0.85$, which in turn resulted in a new value for the a parameter in the final refined model shown in Table 6.1.

It was expected that these parameters would provide a (nominally) perfect match for the short lags, while the long range autocorrelation was expected to be attenuated by a factor of $\frac{0.7}{0.85} = 0.82$ which was deemed acceptable.

6.3 Conclusion

Compared to the other models, the given autocorrelation specification would appear to be most closely met by the NG-LRD model, and even though the match is not exact at large lags the time-dependence should still be second-order self-similar. The steady-state marginal distribution match is theoretically perfect, and as will later be seen the empirical results are quite good.

Chapter 7

The SRP Model

The Spatial Renewal Process has not attracted much attention in the literature for use as a synthetic traffic generator, perhaps in part due to the unnatural block-like sample path that it produces. But in its favor SRP can exactly match any given marginal distribution and it can also, in theory at least, even match sub-exponential autocorrelation functions that produce long range time dependent traffic. Whether this traffic can substitute for more sophisticated models in queueing performance simulation studies is an open question although intuition might suggest that it is not. Yet we are motivated to explore the possibilities by developing a Spatial Renewal Process and using it as input to a queueing performance simulation. The model will be a specialized SRP model with arbitrary marginal distribution and an autocorrelation structure constructed of exponential short range decay and long range Pareto decay.

The one advantage that the SRP model enjoys over other traffic generation models is its extremely low computational demands. This opens up the possibility of using SRP as part of more complex hybrid models which take advantage of this speed but disguise the basic SRP process by manipulation and/or aggregation. Two traffic models making use of these ideas will be developed in the appendix sections of the thesis.

Figure 7.1: Example of SRP Sample Path

7.1 Description

The Spatial Renewal Process, introduced by Jelenkovic, Lazar and Semret in [44] is a subclass of sub-exponentially time dependent stochastic processes. Sub-exponentiality implies that the auto-correlation co-efficient function of the stationary process decays not exponentially for large lags, but in a slower polynomial or “Pareto-like” decay. This implies that the time dependence is long range (LRD) with a Hurst parameter greater than 0.5 and an infinite autocorrelation energy.

Consider a point process with IID inter-renewal time sequence $\{T_n\}$ described by the CDF function $F(t)$. This process plays the role of background process in the SRP model and is responsible for the time dependence structure. A second process, independent of the first, is a sequence of IID values $\{X_n\}$ which are distributed according to the desired steady-state marginal distribution of the traffic. The SRP process is a chain of renewal periods where the n th period is T_n in length and the sample path during this period takes on the value of X_n .

The marginal distribution of the output process is identical to the marginal of X because the processes X_n and T_n are independent of one another. This lack of interdependency between marginal distribution and autocorrelation structure means that any marginal can be chosen for X output process. The autocorrelation co-efficient function $\rho(t)$ of the output process is related to the renewal process through

Equation 7.1 or its time derivative shown in Equation 7.2.

$$1 - \rho(t) = \mu^{-1} \int_0^t (1 - F(u)) du \quad t \geq 0 \quad (7.1)$$

$$-\frac{d}{dt}\rho(t) = \mu^{-1}(1 - F(t)) \quad t \geq 0 \quad (7.2)$$

$$\text{where } \mu = \int_0^\infty (1 - F(u)) du$$

The autocorrelation $\rho(t)$ must be a convex function for all time $t \geq 0$ as implied by the constraints listed in Equations 7.3 through 7.5. This restriction ensures that the corresponding renewal distribution function $F(t)$ is indeed a valid CDF function.

$$\rho(t) = 1, \quad t = 0 \quad (7.3)$$

$$\frac{d}{dt}\rho(t) \leq 0, \quad \forall t \geq 0 \quad (7.4)$$

$$\frac{d^2}{dt^2}\rho(t) \geq 0, \quad \forall t \geq 0 \quad (7.5)$$

Provided that the above constraints are met any autocorrelation function having Pareto, exponential, or piecewise linear form can be chosen.

7.2 Specialized SRP Model

In our implementation the SRP model is specialized so that the autocorrelation function is restricted to an exponential curve followed by a polynomially decaying tail. The autocorrelation co-efficient function $\rho(t)$ is specified by the parameters a, b, c and k . To denote the use of these parameters the model is referred to as SRP-ABCKL.

$$\rho(t) = \begin{cases} e^{at} & 0 \leq t < k \\ bt^c & k \leq t < \infty \end{cases} \quad (7.6)$$

where

$$a, c < 0$$

$$k > 0$$

$$a/c \geq 1/k$$

$$b = e^{ak}/k^c$$

The exponential decay for $0 \leq t < k$ is determined by a while the constant c determines the polynomial decay rate when $t \geq k$. The Hurst parameter H is related to the polynomial curve by the relation $c = 2H - 2$. While a , c , and the knee point k can be independently chosen, the constant b is a dependent function of the other parameters. By direct application of Equation 7.2 and taking into account the boundary condition $F(0) = 0$ the renewal function CDF is found to be:

$$F(t) = \begin{cases} -e^{at} + 1 & 0 \leq t < k \\ -\frac{bc}{a}t^{c-1} + 1 & k \leq t < \infty \end{cases} \quad (7.7)$$

At this point a prototype SRP generator using the renewal function above was implemented for test in MATLAB. The parameters were set to produce an autocorrelation curve $\rho(t)$ as shown in Figure 7.2. Trace paths were generated and an empirical measurement of the autocorrelation showed that the results were in agreement with the analytical curve. But two interesting points were noted from these initial tests.

Figure 7.2: Autocorrelation $\rho(t)$ - Early Test

Firstly, the rate of convergence of the empirical autocorrelation to the analytical one was considerably slower than for a model like TES, in the sense that much longer sample paths were needed before the empirical traces produced consistent statistics. This was expected because of the longer time dependence inherent in the SRP model as compared to the relatively short range TES process [59].

Another interesting feature that showed early on was the tendency to generate many short renewal intervals but occasionally produce a completely uncharacteristic

long renewal period many orders of magnitude greater than the mean renewal. The example sample path shown in Figure 7.3 illustrates this. While the generation of such long periods is consistent with and predictable by the definition of SRP, in practical situations it would be clearly undesirable to generate such paths for use in simulation study. The phenomenon implies a very large variability in the sample path which results in high variability in the output measurements.

Figure 7.3: Sample Path - Early Test

The solution to address this concern was simple: to produce a renewal period and hard clip its maximum value at some arbitrarily set limit L , such that renewal times T_n greater than L are replaced by L . The effect of this modification on the autocorrelation was determined by working backwards from the CDF function $F(t)$ to the autocorrelation function $\rho(t)$. The modified CDF function was now defined as:

$$F(t) = \begin{cases} -e^{at} + 1 & 0 \leq t < k \\ -\frac{bt}{a}t^{c-1} + 1 & k \leq t < L \\ 1 & L \leq t \end{cases} \quad (7.8)$$

By performing the integration in Equation 7.1 the autocorrelation function was found to have become:

$$\rho(t) = \begin{cases} \frac{e^{at} - bL^c}{1 - bL^c} & 0 \leq t < k \\ \frac{bt^c - bL^c}{1 - bL^c} & k \leq t < L \\ 0 & L \leq t \end{cases} \quad (7.9)$$

where $L > k$

As the renewal limit L approaches infinity the modified formulas of Equations 7.8 and 7.9 revert back to the form of the original formulas in Equations 7.6 and 7.7, showing that generality has not been lost.

7.2.1 Specification of Experimental Study Model

The goal of this modeling exercise was to produce an SRP model that would closely resemble the hypothetical traffic specification in terms of its marginal distribution and autocorrelation function. The marginal distribution is easily reproduced in the SRP model by generating the IID sequence X_n with the inverse CDF technique based on the histogram specification. To aid in the search for an autocorrelation function, the MATLAB routine `AcSrpABCKL.m` was written to numerically evaluate Equation 7.9 given the set of parameters a, b, c, k and L . A satisfactory match to the specification curve was found as can be seen in comparison to the analytically predicted curve for the model in Figure 7.4. The match for low lags is excellent while in the high lags the model autocorrelation begins to run low due to the effect of L . The fitting exercise resulted in the parameter values listed in Table 7.1. By design the autocorrelation should be zero for lags greater than L which implies that this model is not truly LRD, yet it does reproduce the heavy autocorrelation structure for reasonably long time lags.

$a = -0.0415$
$b = 1.5771$
$c = -0.5$
$k = 30$
$L = 10^5$
$F_X =$ Output Marginal Distribution CDF

Table 7.1: SRP Model Input Parameters

Figure 7.4: Autocorrelation Specification versus SRP Analytical

7.2.2 Implementation

A prototype SRP model generator was built in the MATLAB environment for testing purposes.¹ This prototype was later replaced by a C implementation to make full use of computing efficiency and provide a convenient way to generate SRP traces either using a stand-alone program or through a set of callable C functions which are listed below.

- The `srpModelType` is a structure which contains the full description of an SRP model. The marginal distribution is specified by the inverse CDF which is of type `PWfunction`. The renewal distribution F can either be specified by an inverse CDF or as a set of ABCKL parameters.
- The routine `readSrpModel` reads in an ASCII file and fills in the fields of an `srpModelType` which it then returns to the calling program. The ASCII file is written in a format similar to the input model file used by `TESTool`.

¹The Matworks Inc., Natick, MA, USA.

- The function `srpInverseRenewal` accepts an argument between 0 and 1 and returns the value generated from the function F^{-1} the inverse CDF of the renewal process T_n . F is specified using the input parameters a, b, c, k and L .
- Routine `genSrp` generates the SRP sample path. The input arguments are time and a pointer to the model structure. A pseudo-code listing of this routine is found in Table 7.2.

```

FUNCTION genSrp( time, theModel )
  IF time < theModel.renewalPeriodStart
    Error - Cannot go backwards in time!
  ELSE IF time < theModel.renewalPeriodEnd
    No change. RETURN theModel.value
  ELSE
    DO
      CASE theModel.renewalFunctionType
        ABCKL: renewal = srpInverseRenewal( ABCKL )
        PWF: renewal = PWfunction( renewal CDF )
      END CASE
      theModel.renewalPeriodStart = renewalPeriodEnd
      theModel.renewalPeriodEnd = renewal + renewPeriodStart
      WHILE time >= theModel.renewalPeriodEnd
        theModel.value = PWfunction( marginal CDF )
      RETURN theModel.value
    END IF
  END FUNCTION

```

Table 7.2: Pseudo-Code for Routine `genSrp.c`

7.2.3 Verification

Marginal distribution and autocorrelation measurements were taken and averaged from a set of five independently generated traces of traffic, each 10^5 samples long. The steady-state marginal distribution measured from the synthetically generated trace was found to match that of the of specification. The autocorrelation structure of the synthetic trace over the first 5000 lags was also in good agreement with that

analytically predicted for the given model parameters. It appears that the SRP model with the parameters specified in section 7.2.1 is a satisfactory model. Its behavior in a queue performance context will be observed during the simulation experiments.

7.3 Conclusion

The generator program written in C has demonstrated that SRP traces can indeed be generated quickly and efficiently. The autocorrelation function can be specified by using the ABCKL model parameters, or by providing a piecewise linear description of the renewal function $F(t)$ which in turn implies the required autocorrelation. The steady-state marginal distribution, which is unrelated to the autocorrelation, is specified using a piecewise CDF function. It remains to observe the behavior of the SRP model in a queue performance simulation and note if the model reproduces results similar to other models, and to examine the differences in light of the log moment function measurements.

Chapter 8

Characterization Based on Moment Generating Functions

The *moment generating function* in statistics $M_X(\theta)$ is defined as the expected value of an exponential function of a random variable [43]. This function lends itself to the calculation of the n^{th} order moments of the random variable, a purpose for which it has acquired the name. The moment statistics provided by the function can completely characterize a single random variable. A stationary process X_k may be partially characterized by applying the same moment generating function to its family of partial sum processes A_k^T , where the time parameter T (restricted to positive integers) defines the size of aggregation. The resulting generalized function $M_{AT}(\theta)$ is dependent on two parameters, θ and T . To note, there are many recent papers in the area of traffic and queueing analysis where functions denoted *effective bandwidth* and *minimum envelope processes* are developed, and they make use of a moment generating function of partial sums in their formulations.

The traditional means of characterizing a stochastic process both in analytical as well as empirical approaches has been the pair of steady-state marginal distribution function and normalized autocovariance function. But eschewing these traditional measures there are notable examples of recently published papers that use the moment generating “class” of functions to characterize stochastic traffic and apply

them to the area of queueing analysis. The growing interest possibly signals an endorsement of this class of function as a better method of characterization than the traditional methods. The use of moment generating functions might even replace both the marginal distribution and autocorrelation functions altogether, which is not unreasonable since marginal is captured by the basic moment generating form and the time phenomena is captured by the T parameter. Even more compelling is the suspicion that moment generating function measurement of traffic is more suitable than autocorrelation, because the former measures workload arrival directly while the latter captures frequency characteristics.

8.1 Literature

The moment generating function of a random variable X may be defined as:

$$M_X(\theta) = \mathcal{E}[e^{\theta X}] \quad (8.1)$$

where X is a random variable and in this thesis θ is a positive value in $\mathcal{R}$. The moment generating function of X is used to derive the large deviation rate function $I(y)$ which plays a central role in *Large Deviation Theory* and the associated theorems of Cramér and Gärtner-Ellis [61]. Large Deviation Theory deals with the (exponential) rate at which the sum of a large series of random variables converges probabilistically towards its expected value. This rate has direct consequences on the workload received by a queue over long time spans and can estimate the excess amount of workload accumulated in the queue's buffer. The references that follow take advantage of the relationship between large deviations in a stochastic series and workload arrival to an idealized queue, in order to characterize the traffic in a way that is important for queue performance behavior.

In [28, 26] a stochastic process is characterized by its “envelope process” which envelopes or bounds the partial sum process $A(k, T)$. The partial sum $A(k, T)$ itself

may be interpreted as a random process, and is transformed by the moment generating function in Equation 8.3 into a minimum envelope process $A^*(\theta, T)$. The minimum envelope process provides a raw and informative characterization of the process X_k over various time scales by varying the T parameter. However as a two dimensional function it is relatively difficult to use in analysis, so a key step is performed resulting in a simplification leaving only the θ parameter. This step leads to the definition of the minimum envelope rate function $a^*(\theta)$ in Equation 8.4 which normalizes the envelope process by T and then takes the limiting case of $T \rightarrow \infty$.

$$A(k, k+T) = \sum_{i=k}^{i=k+T-1} X_i \quad (8.2)$$

$$A^*(\theta, T) = \sup_k \frac{1}{\theta} \log \mathcal{E} \left[e^{\theta A(k, k+T)} \right] \quad (8.3)$$

$$a^*(\theta) = \limsup_{T \rightarrow \infty} \frac{A^*(\theta, T)}{T} \quad (8.4)$$

The concept of the minimum envelope rate is further emphasized in other works [25, 27] where it is referred to as the effective bandwidth function of the traffic process. The application of the $a^*(\theta)$ function to queueing analysis begins with an idealized model of a fluid flow queue, emptying its buffer at constant rate c over a period T as illustrated in Figure 8.1. The input stream is assumed to deliver its workload to the queue at a nearly-constant average rate r over a duration of T , causing the buffer to fill up if the average rate is greater than the service. Given a queue of service rate c and a traffic characterized by a function $a^*(\theta)$ the buffer occupancy probability is given as:

$$\lim_{x \rightarrow \infty} \Pr[Q > x] \approx e^{-\theta x} \text{ such that } c = a^*(\theta) \quad (8.5)$$

The underlying and limiting qualitative assumption made by Equation 8.5 is that time is very large and the characterization of traffic applies only to its long term behavior. This implies that the probability $\Pr[Q > x]$ approximated by $e^{-\theta x}$ is only valid for very large buffer sizes (x large) where the expected probability value is very small.

Nevertheless it is often these low likelihood high buffer occupancy probabilities that are of most interest in designing very low-loss high-speed networks which accounts for the interest in the effective bandwidth function.

Figure 8.1: Idealized Fluid-Flow Traffic & Queue

For the purposes of this thesis a moment generating function $M_{AT}(\theta)$ with parameters θ and T is defined as per Equations 8.6 and 8.7. It is equivalent to applying the moment generating function to the partial sum process A_k^T of the series X_k . The partial sum $A(k, k+T)$ as previously defined in Equation 8.2 does not assume that the series X_k is necessarily stationary, but given that assumption it follows that the distribution of A_k^T is dependent only on T .

$$A_k^T = \sum_{i=k}^{i=k+T-1} X_i \quad (8.6)$$

$$M_{AT}(\theta) = \mathcal{E}[\exp(\theta A^T)] \quad (8.7)$$

The function $M_{AT}(\theta)$ is a common pattern that appears in the description of effective bandwidth and in similar analytical works. For example the previous equations for minimum envelope process and effective bandwidth may be re-written in terms of $M_{AT}(\theta)$ in Equations 8.8 and 8.9, respectively.

$$A^*(\theta, T) = \frac{1}{\theta} \log M_{AT}(\theta) \quad (8.8)$$

$$a^*(\theta) = \limsup_{T \rightarrow \infty} \frac{\log M_{AT}(\theta)}{\theta T} \quad (8.9)$$

8.2 The Case for Adopting the Moment Generating Function

It is common in the literature to develop synthetic traffic models for the purpose of emulating real samples of traffic in network/queueing analysis and performance studies. The least controversial method of comparing synthetic models to their empirical counterparts is based on matching the marginal distribution and autocorrelation functions. While marginal would suffice as a description of any IID traffic process, the autocorrelation function needs to be introduced in order to capture the time-dependence structure of the series. As seen in Equation 1.14 the autocorrelation function is equivalent to power spectral density in an information sense, yet the power spectral density only captures the magnitudes of frequency components and not their phase. There is no simple conceptual connection between the spectral density of a traffic flow and workload input into a queueing system, except for a few rules of thumb. First, it is commonly accepted that that lower frequency components in the autocorrelation (and spectral density) function are primarily responsible for greater accumulated workload in a queue system [20]. Second, a long and heavy positive tail on the autocorrelation implies greater burstiness and accumulated workload as compared to a traffic with rapidly decaying autocorrelation, other things being equal. Third, it is usually assumed that an autocorrelation function composed of an exponential decay will produce buffer probabilities $\Pr[Q > x]$ also exponential in decay, and that sub-exponentially decaying autocorrelation would translate into polynomially decaying probabilities. These intuitive observations are reasonable but not strictly quantifiable or definable, and they only tangentially address the existence of a conceptual queue.

In contrast, one can conceptualize the queueing system as a fluid flow system by referring back to Figure 8.1. The graph plots the size (workload accumulation) of the queue Q_k over a time period of length T which begins at arbitrary time k

and ends at $k + T$. At the start of the period the queue system is probabilistically expected to have length $\mathcal{E}[Q]$. An accumulation of workload in the system results when the A_k^T over time period T produces a mean rate of traffic r that is larger than the queue's service rate c . By time $k + T$ the queue length has grown to b units: the probability of such an occurrence is given by the tail-end distribution of A^T and is estimated by Large Deviation Theory. One needs to consider such a scenario for all possible values of k and T . By assuming stationarity of the input traffic X_k the choice of k is not consequential as the steady state statistics are not conditioned on k . The probability of accumulating workload in the queue is however very dependent on the time scale T considered, so that the probability of excess workload $A^T - rT$ causing the buffer to fill up $\Pr[A^T - rT > b|T]$ is conditioned on T . This probabilistic event is roughly equivalent to the probability of buffer occupancy $\Pr[Q > b|T]$. Furthermore the probability $\Pr[A^T - rT > b|T]$ is related to the parameters θ and T in the $M_{AT}(\theta)$ function. The conceptualizations involved, starting from Large Deviation Theory, to workload accumulation, to buffer occupancy probability provide a much more direct and intuitive bridge between traffic and queue than does the autocorrelation technique. Despite some of the conceptual simplifications involved in relating the input traffic stream to queue buffering effects this method appears better suited than marginal-autocorrelation as a statistical measurement technique.

8.3 The $\mathcal{B}(\theta, T)$ Function and its Interpretation

In order to work with a more visually meaningful function, it was decided to negate the skewing effects of the two parameters θ and T and the non-linearity of the exponential function. The moment generating function $M_{AT}(\theta)$ was normalized with respect to its two parameters by division and with respect to the exponential by logarithm to yield the more intuitive “normalized log moment generating” function:

$$\mathcal{B}(\theta, T) = \frac{\log M_{AT}(\theta)}{\theta T} \quad (8.10)$$

The time parameter has not been removed by performing a limit $T \rightarrow \infty$ as was done in the derivation of effective bandwidth (Equation 8.4) because it was desired to consider short time scale phenomenon rather than long time scale. This was necessitated in part by the inability to simulate long time scale queueing performance with limited computing time, without applying importance sampling. So there is a slight difference in approach in this thesis as compared to the bulk of current literature which concentrates on long time scale effects, yet the overall goal is the same: to model traffic characteristics that affect queue performance.

Due to the normalizations involved one can physically interpret the value produced by $\mathcal{B}(\theta, T)$ as the traffic or workload flow rate averaged over various timescales T and skewed by the θ parameter. The T parameter acts to define the time scale over which the traffic and its partial sums A_k^T are aggregated and averaged. The θ parameter acts to modify the weighting of the measurement between mean value and maximum value. This interpretation is also implied by the general behavior of the e^x function inside the moment generating function. As $\theta \rightarrow 0$ the $\mathcal{B}(\theta, T)$ function returns the mean $\mathcal{E}[A^T]/T$ while on the other extreme as $\theta \rightarrow +\infty$ the maximum $\mathcal{E}[\sup A^T]/T$ is returned. Thus the combination of θ and T together provide a complete signature of the traffic's flow behavior into the queue.

The function $\mathcal{B}(\theta, T)$ should produce a reasonably well behaved two-dimensional curved surface that is constrained to values between the mean and the maximum of the series X_k . For illustration purposes an empirically estimated $\mathcal{B}(\theta, T)$ function is shown in Figure 8.2 with the θ and T axes lying horizontally and the function value $\mathcal{B}$ on the vertical axis. The measured series was an IID Binomial series with parameters $n = 5$ and $p = 0.25$, having mean and peak values of 1.25 and 5.0 respectively. The $\mathcal{B}$ function characterizes the input traffic and its partial sums over various time scales T . Figure 8.2 also superimposes a horizontal plane representing the constant service rate c of the queue. This superposition of traffic flow and queue flow leads to a reasonable (but intuitive) hypothesis: $\mathcal{B}(\theta, T)$ points lying below the

c -plane do not result in buffer filling, while points above produce buffer filling in a manner related to the shape of the curve above the plane. The working hypothesis is that traffic series showing similar $\mathcal{B}(\theta, T)$ signatures in the critical region above the plane will produce similar performance in the queueing system, or that at least the critical region will better predict such queue effects.

Figure 8.2: Empirically Measured $\mathcal{B}(\theta, k)$ Function of Binomial Series

8.3.1 Implementation

Although the function $\mathcal{B}(\theta, T)$ can be analytically defined for a given stochastic model it can also be empirically estimated from traces of traffic data. Due to the complexity of deriving analytical expressions of the various synthetic models, the latter method was chosen. The MATLAB routines `Lmg.m` and `Nlmg.m` were created to provide numerical estimates of the two dimensional surface $\mathcal{B}(\theta, T)$. The input arguments are: an input data trace X_k and a set of θ and T values to define the surface range. The output of the function is a two dimensional matrix of the $\mathcal{B}$ function values.

8.4 Conclusion

A generalized moment generating function of the form $M_{A\tau}(\theta)$ characterizes the partial sums of various aggregations of a stochastic process X_k in order to describe it. While this does not capture the full time-dependence nature of the series itself, it does capture what probably are the essential aspects of the series so far as arrival workload quantities that affect queueing systems are concerned.

The normalized log moment generating function $\mathcal{B}(\theta, T)$ described in the previous section provides an alternative method for characterizing traffic data in comparison to the traditional marginal distribution and autocorrelation method. The determination of $\mathcal{B}(\theta, T)$ can be done empirically by measurement of traffic data, although as with the estimation of marginal and autocorrelation functions, the results will be subject to statistical error. Moreover, due to the limitations of traffic generation and model simulation, only the short and medium term characteristics of the traffic will be satisfactorily captured by the $\mathcal{B}(\theta, T)$ measurements. These limitations aside, the queue simulation experiment will attempt to support the hypothesis that $\mathcal{B}(\theta, T)$ is a potentially useful measurement for traffic characterization in addition to the marginal distribution and autocorrelation functions.

Chapter 9

Simulation Results and Measurements

Initially four traffic models were created to provide the synthetic traffic necessary for the investigation of queueing effects and stochastic traffic. The first of these models, the TES model, drew attention at an early stage of model parameter fitting because it could not match the required autocorrelation specification. Nevertheless the process of design and simulation was continued with this model because of the interest in comparing it with a newer derivation of the model denoted E-TES. The three remaining synthetic models E-TES, SRP and NG-LRD, were successfully fitted to the marginal and autocorrelation requirements of the experiment and they formed the core study of traffic, its characterization and queue performance. At a later stage of experimentation, in part motivated by the initial results, a fifth synthetic traffic model denoted SRPMOD was introduced. Its purpose was to set up a comparison between it and the NG-LRD model and to provide more evidence in regards to characterization, queueing effects and the moment generating function.

9.1 Simulation Runs

The discrete-time (i.e. slotted-time) simulation was set to run for a fixed length of 180000 time units. A stream of synthetic traffic 180000 sample values long fed the

queue over the duration of the simulation. The synthetic traffic values were generated on a frame-by-frame basis such where the value represented the size of a frame, and for each a proportionate number of cells were transmitted to the queue. The Lindley queue operated in slotted-time mode in step with the simulation and traffic generator. The warmup of the traffic-queueing system lasted 18000 time units. The ratio of warmup to duration at 10% was somewhat arbitrary as the heavy time dependence structure of the traffic made it difficult to discriminate steady-state from transient operation, but given the aims of the study it was arguably not relevant. The experiments made use of the independent replications approach for estimating means and confidence intervals, based on the t -distribution. Sixty independently seeded runs were executed for each model and the traffic statistics and queue buffer occupancy levels were observed. For buffer occupancy and autocorrelation, mean estimates are presented with 95% confidence intervals.

9.2 The E-TES, SRP and NG-LRD Traffic Runs

The unconditional probability of buffer occupancy $\Pr[Q > y]$ was recorded from the simulation in a time-average sense. Buffer occupancy results from simulation of the E-TES, SRP and NG-LRD models are shown in Figure 9.1. The difference between them is remarkable given the fact that all three models were designed to produce the same (nominal) marginal and autocorrelation structure. The E-TES traffic produces low occupancy levels in the queue buffer evidenced by the rapid exponential decay of the $\Pr[Q > y]$ probability curve, while the SRP (and to a lesser extent NG-LRD) traffic produce very high levels of buffer occupancy. The result is significant in practical terms because the SRP and NG-LRD traffics would be subject to high cell blocking and delay jitter while the E-TES traffic would not.

The marginal distributions measured from the synthetic traffic streams are seen in Figure 9.2 to match the desired specification very accurately. Figures 9.3, 9.4

Figure 9.1: Buffer Occupancy $\Pr[Q > y]$ for E-TES, SRP, NG-LRD Traffic

Figure 9.2: Marginal Distribution of E-TES, SRP, NG-LRD Traffic vs. Spec.

and 9.5 plot the empirically measured autocorrelation functions of the traffic. The upper plots in each figure graph the autocorrelation from 0 to 500 time lags, while the lower ones graph on a larger time scale of 0 to 5000 lags. The nominal specification appears in all the plots as a solid curve to provide reference. For reasons of clarity, only the upper and lower 95% confidence intervals of the empirical curves have been plotted, the mean value lying equidistant between the two. From empirical measurement it was found that the models produced good matching to the autocorrelation

Figure 9.3: Autocorrelation of E-TES Traffic vs. Specification. Top graph 500 lags, bottom graph 5000 lags.

Figure 9.4: Autocorrelation of SRP Traffic vs. Specification. Top graph 500 lags, bottom graph 5000 lags.

Figure 9.5: Autocorrelation of NG-LRD Traffic vs. Specification. Top graph 500 lags. bottom graph 5000 lags.

specification over the first 500 lags. It can be seen that on the larger time scale of 5000 lags the empirically measured results do not exactly match the design specification yet the results do present an opportunity for interpretation. What is striking about these autocorrelation results is a lack of relationship to the buffer results, in particular that the empirically measured autocorrelation tail of E-TES is consistently higher than that of the other models but results in the lowest buffer probabilities. The SRP traffic produces the highest buffer occupancy yet its autocorrelation is roughly on par with that of NG-LRD, both of which have consistently low autocorrelation tails. These early observations suggest that two traffic streams with the same first- and second-order characteristics might in practice produce highly different queueing behavior and performance, and that modeling based on the marginal-autocorrelation methodology is not reliable.

9.3 The TES Traffic Run

The comparison between the TES and E-TES traffic runs also gave interesting results. Figure 9.6 shows the buffer occupancy produced by the TES model traffic. The synthetic traffic generated by the TES model matched the specification exactly as seen in Figure 9.7. However the autocorrelation function as plotted in Figure 9.8 is best described as an exponential decay which dies around 500 lags, and as compared to the polynomial tail which had been prescribed, is not an acceptable match. This was analytically predicted in Chapter 4. While both models matched marginal distribution equally well TES does not reproduce the long autocorrelation tail while E-TES matches the autocorrelation very well up to 5000 lags. Yet the queue occupancy results are surprisingly equal despite the difference in time dependence structure as indicated by the autocorrelation function.

Figure 9.6: Buffer Occupancy $\Pr[Q > y]$ for TES Traffic

Figure 9.7: Marginal Distribution of TES Traffic vs. Specification

Figure 9.8: Autocorrelation of TES Traffic vs. Specification. Top graph 500 lags, bottom graph 5000 lags.

9.4 Characterization Using the $\mathcal{B}(\theta, T)$ Function

Autocorrelation is a linear measurement which emphasizes the power spectral nature of a traffic flow lending itself more to signal theory than queueing analysis. There is no simple conceptual connection between the spectral density of a traffic flow and workload input into the queueing system except for a few rules of thumb. For example, it is commonly accepted that that lower frequency components in the autocorrelation function are primarily responsible for buffer occupancy in a queue system [20]. Or, that a long and heavy positive tail on the autocorrelation implies greater burstiness and buffer occupancy as compared to a traffic with rapidly decaying autocorrelation, other things being equal.

Empirical measurement of the $\mathcal{B}(\theta, T)$ function from the synthetic traffic streams yielded the surfaces shown in Figures 9.9, 9.10, 9.11 and 9.12. The two independent variables, θ and T lie in the horizontal plane and the value of the $\mathcal{B}(\theta, T)$ function is on the vertical. In all the figures a horizontal plane superimposed on the surface represents the constant rate of service of the queue, which was fixed at a traffic rate of 17490 units = 330 cells per time unit, equivalently a utilization ratio of 1/3.

As was seen the SRP and NG-LRD models produced similarly high buffer occupancy results in the queue, which is reflected in the similar shapes of the $\mathcal{B}(\theta, T)$ traffic measurements. The $\mathcal{B}$ measurement of the E-TES traffic is visually very different than the two others and appears to predict the corresponding difference in queueing results. The TES and E-TES models produced virtually identical queueing results and their $\mathcal{B}$ surfaces are likewise identical. To complement the qualitative and rather interpretive nature of the visual observations a more exact numerical comparison was desired. To make the comparison more meaningful the critical region of $\mathcal{B}$ as suggested in Chapter 8 was isolated by truncating values below the c -plane up to the level of the plane:

$$\mathcal{B}^*(\theta, T) = \begin{cases} \mathcal{B}(\theta, T) & \mathcal{B}(\theta, T) \geq c \\ c & \mathcal{B}(\theta, T) < c \end{cases} \quad (9.1)$$

Figure 9.9: $B(\theta, t)$ Characterization of TES Traffic Figure 9.10: $B(\theta, t)$ Characterization of E-TES Traffic

Figure 9.11: $B(\theta, t)$ Characterization of SRP Traffic Figure 9.12: $B(\theta, t)$ Characterization of NG-LRD Traffic

The following two tables present a rudimentary numerical comparison of the measured $\mathcal{B}^*(\theta, T)$ functions of the traffic. Table 9.1 calculates the average difference between the surfaces $\mathcal{B}_X^* - \mathcal{B}_Y^*$ in a pair-wise comparison of generated traffic, calculated with:

$$\frac{1}{n} \sum_{(\theta, T)} \mathcal{B}_X^*(\theta, T) - \mathcal{B}_Y^*(\theta, T)$$

The set of (θ, T) covers the same range as that illustrated on the main axes of the figures. The order and magnitude of the values imply that the SRP traffic would produce highest buffer occupancy followed closely by NG-LRD traffic, and that much lower buffer occupancy would be expected for E-TES. The TES and E-TES traffic show no significant numerical difference. The $\mathcal{B}^*(\theta, T)$ surfaces are also compared in Table 9.2 but in a root mean square sense:

$$\sqrt{\frac{1}{n} \sum_{(\theta, T)} (\mathcal{B}_X^*(\theta, T) - \mathcal{B}_Y^*(\theta, T))^2}$$

The low RMS difference values imply a similarity between the SRP and NG-LRD model traffic in contrast to the E-TES traffic, also consistent with the observed buffer results. There appears to be a strong correlation or relationship between traffic characterization $\mathcal{B}(\theta, T)$ on the input side and buffer occupancy probability results on the output, regardless of stochastic model type. The autocorrelation function was seen to be highly sensitive to the type of synthetic model used, emphasized by the fact that marginal distribution was experimentally held constant.

	TES	E-TES	SRP	NG-LRD
TES	0	23	1300	940
E-TES	-23	0	1200	920
SRP	-1300	-1200	0	-320
NG-LRD	-940	-920	320	0

Table 9.1: Comparison of $\mathcal{B}_X^* - \mathcal{B}_Y^*$ Mean Differences

	TES	E-TES	SRP	NG-LRD
TES	0	81	3000	2500
E-TES	81	0	2900	2400
SRP	3000	2900	0	980
NG-LRD	2500	2400	980	0

Table 9.2: Comparison of $\mathcal{B}_X^* - \mathcal{B}_Y^*$ RMS Differences

9.5 Further Evidence with the SRPMOD Traffic

Another SRP model was introduced specifically for the purpose of comparison to the NG-LRD traffic model. This fifth model was labeled SRPMOD to differentiate it from the previous SRP model. Whereas the methodology employed in the previous modeling exercises began with a set of marginal and autocorrelation specifications, this new model was designed with specific regard to matching the same output buffer results as that produced by the NG-LRD traffic. Optimization of the model was achieved by degrading the magnitude of autocorrelation to be produced in the tail region $b\tau^c$ for lags $\tau > 30$ by reducing the value of c from -0.5 to -1.245 . This ultimately resulted in a new model which could attain a good match in the buffer occupancy to that of the NG-LRD traffic.

As is seen in Figure 9.13 the observed buffer occupancy curve matches closely to that produced by the NG-LRD traffic, and the model's marginal distribution matches NG-LRD in that aspect as well. The autocorrelation function, a variable parameter during the modeling exercise, was measured after-the-fact and seen to have degenerated as seen in Figure 9.15. The autocorrelation is practically zero for lags $\tau > 100$ while the NG-LRD example sustains autocorrelation well up to 5000 lags. That one model (NG-LRD) could produce almost the same buffer occupancy as another model (SRPMOD) with such a vast difference in autocorrelation structure provides yet more evidence to question marginal-autocorrelation based modeling.

Figure 9.13: Buffer Occupancy $\Pr[Q > y]$ with SRPMOD Traffic

Figure 9.14: Marginal Distribution of SRPMOD Traffic vs. Specification

Mean Differences

	NG-LRD	SRPMOD
NG-LRD	0	-33
SRPMOD	33	0

RMS Differences

	NG-LRD	SRPMOD
NG-LRD	0	430
SRPMOD	430	0

Table 9.3: Comparison of $\mathcal{B}_{huang}^* - \mathcal{B}_{srpmo}^*$ Differences

Figure 9.15: Autocorrelation of SRPMOD Traffic vs. Specification. Top graph 500 lags, bottom graph 5000 lags.

Figure 9.16: $B(\theta, t)$ Characterization of SRPMOD Traffic

The $\mathcal{B}^\pi$ characterization of the SRPMOD traffic illustrated in Figure 9.16 visually resembles that of the NG-LRD traffic, correctly inferring the similarity of queue output results. The critical region surfaces $\mathcal{B}_{huang}^\pi$ and $\mathcal{B}_{srpmo}^\pi$ likewise show a numerical closeness in Table 9.3 which is not seen among the previous traffic comparisons in Tables 9.1 and 9.2. Empirical measurements of $\mathcal{B}(\theta, k)$ appear again to characterize traffic more robustly than the marginal-autocorrelation pair.

Chapter 10

Conclusion

10.1 General Discussion

As evidenced by the quantity of literature and study devoted to traffic modeling and the vast range of stochastic model types proposed by such studies it is arguable that the field is still in a state of flux. As Melamed comments in [15] it is difficult to create traffic models when there is a lack of comprehensive characterization of empirically observed traffic, complicated by the fact that such traffic is rapidly changing in scale and content so that the design specifications themselves are speculative. Whether all types of multimedia traffic can be captured by a single class of model with variable parameters, or necessarily by the use of specific custom built models, is yet to be seen. A related issue is that of treating all traffic as either inherently self-similar or exponential in time dependence structure. And the question of traffic characterization needs to be addressed satisfactorily with regard to the above issues. Ultimately the goal is to bridge the gap between theoretician and practitioner, allowing the latter to assess the limitations and applicability of a theoretical framework for solving real problems [62].

The marginal-autocorrelation methodology in modeling presumes that stochastic processes should be characterized by the marginal distribution and autocorrelation functions — at least as a first approximation. The characterization is

conceptually based on a linear system or frequency domain perspective of modeling and is primarily concerned with the traffic but with little regard to the queue. Some authors have explicitly addressed this weakness and taken into consideration the queue service behavior in addition to marginal-autocorrelation of traffic [19]. Effective bandwidth characterization is based on a fresh approach to traffic-queue modeling by considering workload and service rate of the traffic queueing system. The description of accumulated workload is in concept simple: to measure the average input workload presented to a queue over various time scales, parameterized with values of θ that measure a range of mean to peak values of workload. The central question put forth by the thesis was whether the marginal-autocorrelation characterization of traffic is good enough for traffic-queue performance analysis, and if log moment generating function characterization is a robust alternative. The thesis does not question the use of autocorrelation in general as one can argue that given appropriate conditions of traffic aggregation leading to Gaussian traffic the characterization by autocorrelation is sufficient.

The experiment was unadorned: a single stream of traffic from the synthetic models was fed into an infinite buffer queue with constant rate server in order to measure buffer occupancy levels alongside input traffic statistics.

Creation of synthetic traffic followed the standard marginal-autocorrelation practice of modeling, partly in order to test the methodology itself and partly due to its wide acceptance. The choice of very dissimilar traffic models and the definition of a traffic with highly non-Gaussian distribution were intentionally employed to test the limitations of marginal-autocorrelation modeling. The marginal distribution as derived from empirical measurement of compressed video and the composition of an autocorrelation structure with hybrid exponential-polynomial decay was furthermore relevant to LRD traffic. Four classes of traffic model were selected: The TES background process is a strictly pure Markov Chain process and the resulting output TES

process is roughly Markovian for all intents and purposes i.e. with complex exponentially decaying autocorrelation. The E-TES process is autoregressive due to the autoregressive nature of the background process generated by an FIR filter. The SRP process is based on a background point renewal process and as such can theoretically produce LRD autocorrelation. The NG-LRD model is based on a background Gaussian process with inherent LRD time dependence which carries through to its output process. The stochastic series generated by the synthetic models was interpreted as the random flow or data rate of traffic with respect to time which was then translated into a discrete quantity of cells over periodic intervals in time.

The simulation software written as part of the thesis work offered a choice of queue type. There were two versions of a discrete-time Lindley queue operating in cut-through and non-cut-through modes and a third version for continuous-time cell-by-cell operation. It was decided to simulate a Lindley queue in discrete-time due to the gain in processing speed. While a continuous time simulation was also feasible it was noted that the two versions of the Lindley queue would provide upper and lower bounds to the results obtained by a continuous time queue, and some tests during development showed that the bounds were fairly tight.

The experimental observations comprised the marginal distribution and autocorrelation and normalized log moment generating function of the input traffic, and the buffer occupancy probability $\Pr[Q > x]$ of the queue. A comparison of the TES and E-TES traffic runs was revealing in that E-TES, for all its improved hyperexponential autocorrelation tail, did not produce a significant increase in the buffer occupancy. The marginal-autocorrelation characterization failed to detect this anomaly while the $\mathcal{B}$ characterization did. The TES model was henceforth put aside because it could not produce the required heavy-tailed autocorrelation that the other models could. When the E-TES, SRP and NG-LRD traffic runs were comparatively examined, it appeared that autocorrelation is a poor predictor of buffer occupancy

because the relationship between amount of autocorrelation and amount of occupancy was counter-intuitive. The $\mathcal{B}$ measurements in contrast showed a positive correlation to the buffer results. Given the circumstantial nature of the first two observations a third experimental situation was designed from a slightly different angle of attack. A fifth model SRP_{MOD} was introduced for comparison to the NG-LRD model. Both models produced the same buffer occupancy and exhibited similar $\mathcal{B}$ signatures while the autocorrelation functions were seen to be completely different. Taken as a whole the analyses suggest that marginal-autocorrelation characterization is adversely affected by different traffic types in that it does not adequately capture enough information of the traffic. The log moment generating characterization $\mathcal{B}$ was seen to be robust in this important regard.

Some secondary outcomes of the thesis work include the E-TES model and the SRP model. The E-TES model is academically interesting in that it can produce very hyperexponential autocorrelation yet result in the same queue effects (i.e. traffic workload) as a similar TES model. Perhaps the modulo-1 background process accounts for such a phenomenon. The SRP traffic was seen to produce the greatest amount of buffer occupancy for a given amount of autocorrelation among the traffic models, leading one to hypothesize that SRP traffic generally results in an upper bound for buffer occupancy and delay. The SRP generator also shows a potential for fast FGN generation. Unfortunately it was observed that SRP traffic displays a high variability in its structure as evidenced by the wide confidence intervals of its autocorrelation and $\mathcal{B}$ measurements, hinting that simulation results might be less precise than those of the other models. An issue that was not addressed in the thesis was the characterization of cyclic processes which display periodic components in the sample path and autocorrelation such as that seen in the MPEG GOP pattern. The direct and intuitive way that $\mathcal{B}$ characterization measures traffic workload to the queue suggests that it can describe such traffic as well.

The effective bandwidth $a^*(\theta)$ function characterizes the traffic-queue systems in the asymptotic infinite time scale sense only. Treatments of effective bandwidth in the current literature are necessarily analytical since the $a^*(\theta)$ function cannot be empirically estimated from finite length traces. In contrast the normalized log moment generating function characterizes over a finite specified range of time scales and thus is suited to empirical estimation in practical use. An interesting aspect of the two dimensional surface generated by $\mathcal{B}(\theta, T)$ is its behavior for large time values. In theory the function $\mathcal{B}(\theta, T) |_{T \rightarrow \infty}$ should approach the effective bandwidth $a^*(\theta)$ curve, but as was seen in the traffic measurements of Chapter 9 the curves flattened out as the time parameter approached 2^{16} . One can interpret the result as due to the lack of long bursts in the traffic traces, understandable in that they constitute rare events which by definition are undetectable in finite length measurements.

10.2 Achievements

The process of experimentation presented both theoretical and practical challenges. The background literature research led to the formulation of a hypothesis questioning the traditional traffic characterization method and suggesting a new characterization measure. The experiment comprising a select set of traffic models and a simulation engine was implemented. An analysis of the SRP and (particularly) the E-TES models had to be developed before they could be implemented. The simulation was written in a commercially available object-oriented simulation language.

Theoretical

- Showed the failure of the marginal-autocorrelation modeling methodology in a specific context,
- Developed a log moment generating traffic characterization method related to the method of effective bandwidth.

- Showed that the new method provides rigorous and accurate traffic characterization for the end purposes of observing queueing effects.

Implementation

- Suggested the use of an FIR filter to generate ARM innovation sequence. Developed analytical method to solve for autocorrelation of such model. Implemented model generator using C language,
- Constrained the autocorrelation function of the SRP model to exponential-polynomial parameters for ease of use. Implemented such model generator using C language,
- Demonstrated ability of SRP model to generate FGN self-similar traffic,
- Demonstrated a practical algorithm to capture IBP characteristics of empirical MPEG trace. Generated synthetic IBP traffic by modulating an aperiodic traffic model (SRP) with frame pattern. The implementation was in Matlab language.

10.3 A New Modeling Methodology?

A natural extension implied by the work on log moment traffic characterization is to reverse the characterization procedure and create synthetic models given a traffic specification. It may be possible to forward engineer a model to produce any generic traffic described by a set of $\mathcal{B}$ values $\{\mathcal{B}_1(\theta_1, T_1), \mathcal{B}_2(\theta_2, T_2), \mathcal{B}_3(\theta_3, T_3), \dots\}$ over a range of θ and T . While it would be desirable to consider all the information supplied by the whole surface $\mathcal{B} \forall (\theta, T)$ for traffic matching purposes, in practice it might only be necessary to concentrate on a limited region. The “critical region” of the surface was employed in this vein. A more elegant idea worth further investigation is the identification of a succinct set of traffic descriptors drawn out of the $\mathcal{B}$ surface, enabling the practitioner to use a limited set of parameters in characterization and

modeling procedures. The as-yet unidentified parameters would likely comprise the low-order moments of the traffic process X_k and its partial sums A_k^T . Once a set of traffic parameters are defined the process of synthetic model design may begin. The selection and fitting of a stochastic model to the given requirements would be facilitated if one could identify a stochastic model class that can be analyzed once then applied to all types of traffic. Such an archetypical model would need be capable of emulating a wide variety of traffic and be amenable to $\mathcal{B}$ characterization. The SRP stochastic model might be able to fill the role because of its apparent simplicity and efficiency. Techniques developed to these ends would ideally be analytical in nature and lend themselves to closed form solution although the use of numerical search techniques cannot be ruled out.

References

- [1] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison-Wesley Publishing Company, Don Mills, Ontario. 2nd edition, May 1994.
- [2] V. Paxson and S. Floyd. Wide Area Traffic: The Failure of Poisson Modeling. *IEEE Transactions on Networking*, 3(3):226–244, June 1995.
- [3] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson. Self-Similarity Through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level. In *Proceedings of ACM Sigcomm 95*, pages 100–113, October 1995.
- [4] M. Schwartz. *Broadband Integrated Networks*, page 25. Prentice Hall, New Jersey, USA, 1st edition, 1996.
- [5] J. Gordon. Pareto Process as a Model of Self-Similar Packet Traffic. In *Proc. IEEE Globecom*, pages 2232–2236, 1995.
- [6] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Fast Simulation for Self-Similar Traffic in ATM Networks. In *Proc. IEEE ICC*, pages 438–444, 1995.
- [7] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified Approach. In *Proceedings of ACM Sigcomm 95*, pages 114–125, 1995.
- [8] I. Norros. The Management of Large Flows of Connectionless Traffic on the Basis of Self-Similar Modelling. In *Proc. IEEE ICC*, pages 451–455, 1995.
- [9] W. Lau, A. Erramilli, J. L. Wang, and W. Willinger. Self-Similar Traffic Parameter Estimation: A Semi-Parametric Periodogram-Based Algorithm. In *Proc. IEEE Globecom*, pages 2225–2231, 1995.
- [10] R. Addie, M. Zukerman, and T. Neame. Fractal Traffic: Measurements Modelling and Performance Evaluation. In *Proc. IEEE Infocom*, pages 977–984, 1995.
- [11] D. P. Heyman and T. V. Lakshman. Source Models for VBR Broadcast Video Traffic. *ACM Transactions on Networking*, pages 40–48, February 1996.
- [12] N. Duffield, J. Lewis, N. O’Connell, and et al. Predicting Quality of Service for Traffic with Long-Range Fluctuations. In *Proc. IEEE ICC*, pages 473–477, 1995.

- [13] B. Melamed and J. R. Hill. A Survey of TES Modeling Applications. Technical report, Rutgers Center for Operations Research, New Jersey, USA, 1995. <http://rutcor.rutgers.edu/>.
- [14] L. Kulkarni and S. Q. Li. Traffic Modeling: matching the Power Spectrum and Distribution. In *Proc. IEEE Globecom*, pages 1701–1706, 1995.
- [15] V. S. Frost and B. Melamed. Traffic Modeling for Telecommunications Networks. *IEEE Communications Magazine*, pages 70–81, March 1994.
- [16] I. Lambadaris, M. Devetsikiotis, A. R. Kaye, M. R. Ismail, C. M. Sharon, Y. Fang., and C. C. Huang. Traffic Modelling and Design Methodologies for Broadband Networks. *Can. Journ. Elect. Comp. Eng.*, 20(3), 1995.
- [17] P. Pruthi and A. Erramilli. Heavy-Tailed ON/OFF Source Behavior and Self-Similar Traffic. In *Proc. IEEE ICC*, pages 445–450, 1995.
- [18] N. T. Plotkin and P. P. Varaiya. The Entropy of Traffic Streams in ATM Virtual Circuits. In *Proc. IEEE Infocom*, pages 1038–1045, 1994.
- [19] C. L. Hwang and S. Q. Li. On Input State Space Reduction and Buffer Noneffective Region. In *Proc. IEEE Infocom*, pages 1018–1028, 1994.
- [20] H. D. Sheng and S. Q. Li. Second Order Effect of Binary Sources on Characteristics of Queue and Loss Rate. In *Proc. IEEE Infocom*, pages 18–27, 1993.
- [21] C. Hwang and S. Q. Li. On the Convergence of Traffic Measurement and Queueing Analysis: A Statistical Match Queueing SMAQ Tool. In *Proc. IEEE Infocom*, pages 602–612, 1995.
- [22] S. Chong and S. Q. Li. Sigma-Rho Characterization Based Connection Control for Guaranteed Services in High Speed Networks. In *Proc. IEEE Infocom*, pages 835–844, 1995.
- [23] C. Chang, H. Chi, and R. Cheng. A Power-Spectrum Based Connection Admission Control for ATM Networks. In *Proc. IEEE ICC*, pages 637–641, 1996.
- [24] S. Chong and S. Q. Li. Fundamental Limits of Input Rate Control in High Speed Networks. In *Proc. IEEE Infocom*, pages 662–671, 1993.
- [25] C. S. Chang and J. A. Thomas. Effective Bandwidth in High-Speed Digital Networks. *IEEE Journal on Selected Areas in Communications*, 13(6):1091–1099, August 1995.
- [26] C. S. Chang. Stability, Queue Length, and Delay of Deterministic and Stochastic Queueing Networks. *IEEE Transactions on Automatic Control*, 39(5):913–931, May 1994.

- [27] G. Kesidis, J. Walrand, and C. S. Chang. Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources. *IEEE ACM Transactions on Networking*, pages 424–428, August 1993.
- [28] C. S. Chang, P. Heidelberger, S. Juneja, and P. Shahabuddin. The Application of Effective Bandwidth to Fast Simulation of Communication Networks. Technical report, IBM Research Division, Yorktown Heights, NY, USA. May 1993. RC 18877.
- [29] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*, pages 336–337. Addison-Wesley Publishing Company, Don Mills, Ontario, 2nd edition, May 1994.
- [30] J. Gordon. Long Range Correlation in Multiplexed Pareto Traffic. In *Proceedings of International IFIP-IEEE Conference on Broadband Communications*, pages 28–39. 1996.
- [31] A. Corte, A. Lombardo, and G. Schembra. Modeling Superposition of ON-OFF Correlated Traffic Sources in Multimedia Applications. In *Proc. IEEE Infocom*, pages 993–1000, 1995.
- [32] F. Kamoun and M. M. Ali. Queuing Analysis of ATM Tandem Queues with Correlated Arrivals. In *Proc. IEEE Infocom*, pages 709–716, 1995.
- [33] S. B. Kim, M. Y. Lee, and M. J. Kim. Sigma-Matching Technique for MMPP Modeling of Heterogeneous ON-OFF Sources. In *Proc. IEEE Globecom*, pages 1090–1094, 1994.
- [34] K. Qian and D. McDonald. An approximation method for complete solutions of markov modulated fluid models. In *Proc. IEEE Globecom*, pages 1406–1411, 1995.
- [35] Y. Fang, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye. Exponential Bounds for the Waiting Time Distribution in Markovian Queues with Applications to TES/GI/1 Systems. In *Performance Evaluation Review SIGMETRICS 95*, pages 108–115, 1995.
- [36] N. Rananand. Approximating a Variable Bit Rate Source by Markov Processes. In *Proc. IEEE Globecom*, pages 1107–1112, 1994.
- [37] N. L. Fonseca and J. A. Silvester. Modelling the Output process of an ATM Multiplexer. In *Proc. IEEE ICC*, pages 721–725, 1994.
- [38] M. Conti, E. Gregori, and A. Larsson. On the Relevance of Long Term Correlation in MPEG-1 Video Traffic. In *Proceedings of the International IFIP-IEEE Conference on Broadband Communications*, pages 52–63, 1996.
- [39] M. E. Anagnostou, J. A. Sanchez, and I. S. Venieris. A Multiservice Structured Markovian Traffic Source Model. In *Proc. IEEE ICC*, pages 1008–1013, 1994.
- [40] W. Hsin and K. Sohraby. Cycle Modulated Sources and their Correlational Structure. In *Proc. IEEE ICC*, pages 1619–1623, 1996.

- [41] Z. Zhang and A. S. Acampora. Equivalent Bandwidth for Heterogeneous Sources in ATM Networks. In *Proc. IEEE ICC*, pages 1025–1031, 1994.
- [42] A. Hatono. Evolution of the Subspace Method to ATM Traffic Analysis. In *Proc. IEEE ICC*, pages 1596–1600, 1996.
- [43] P. G. Harrison and N. M. Patel. *Performance modelling of Communication Networks and Computer Architectures*. Addison-Wesley Publishing Company, Don Mills, Ontario, 1st edition, 1992.
- [44] P. R. Jelenković, A. A. Lazar, and N. Semret. Multiple Time Scales and Subexponentiality in MPEG Video Streams. In *Proceedings of International IFIP-IEEE Conference on Broadband Communications*, pages 64–75, 1996.
- [45] D. Kouvatsos and R. Fretwell. Correlated Traffic: Explicit Effects and the Creation of Burstiness from a Simple Queue. In *Proceedings of IFIP Workshop TC6*, pages 39/1–12, July 1995.
- [46] H. Lee and J. Mark. Characterization of Statistical Multiplexing of Heterogeneous ATM Sources. In *Proc. IEEE Infocom*, pages 345–353, 1995.
- [47] Q. Ren and H. Kobayashi. Diffusion process Approximations of a Statistical Multiplexer with Markov Modulated Bursty Traffic Sources. In *Proc. IEEE Globecom*, pages 1100–1106, 1994.
- [48] S. Giodano, M. Pagano, R. Pannocchia, and F. Russo. A New Call Admission Control Scheme Based on the Self Similar Nature of Multimedia Traffic. In *Proc. IEEE ICC*, pages 1612–1618, 1996.
- [49] A. Adas and A. Mukherjee. On Resource Management and QoS Guarantees For Long Range Dependent Traffic. In *Proc. IEEE Infocom*, pages 779–787, 1995.
- [50] W. Lau, A. Erramilli, J. L. Wang, and W. Willinger. Self-Similar Traffic Generation: The Random Midpoint Displacement Algorithm and its Properties. In *Proc. IEEE ICC*, pages 466–472, 1995.
- [51] J. Mata, G. Pagan, and S. Sallent. Multiplexing and Resource Allocation of VBR Video over ATM Networks. In *Proc. IEEE ICC*, pages 1401–1405, 1996.
- [52] J. Wu, Y. Chen, and C. Shiu. Traffic Modelling and Bandwidth Allocation for MPEG Video Sources in ATM Networks. In *Proc. IEEE Globecom*, pages 2237–2241, 1995.
- [53] C. Cario and B. Nelson. Numerical Methods for Fitting and Simulation Autoregressive-To-Anything Processes. Technical report, Delphi Packard Electric Systems, Ohio, USA, August 1996. Preprint.
- [54] T. Czachorski, J. Fourneau, and L. Kloul. Diffusion Approximation to Study the Flow Synchronization in ATM Networks. In *Proceedings of IFIP Workshop TC6*, pages 28/1–10, July 1995.

- [55] J. Pruneski and S. Q. Li. The Linearity of Low Frequency Traffic Flow: An Intrinsic I/O Property in Queueing Systems. In *Proc. IEEE Infocom*, pages 613–623, 1995.
- [56] S. Chong, C. Hwang, X. Zhao, and S. Q. Li. Link Capacity Allocation and Network Control by Filtered Input Rate in High Speed Networks. In *Proc. IEEE Globecom*, pages 744–750, 1993.
- [57] G. deVeciana, C. Courcoubetis, and J. Walrand. Decoupling Bandwidths for Networks: A Decomposition Approach to Resource Management. In *Proc. IEEE Infocom*, pages 466–473, 1994.
- [58] W. Lau and S. Q. Li. Traffic Analysis in Large-Scale High-Speed Integrated Networks: Validation of Nodal Decomposition Approach. In *Proc. IEEE Infocom*, pages 1320–1329, 1993.
- [59] P. R. Jelenković and B. Melamed. Automated TES Modeling of Compressed Video. In *Proc. IEEE Infocom*, pages 746–752, 1995.
- [60] B. Melamed. Autoregressive Modular (ARM) Processes. Technical report, Enterprise Technology Solutions, Inc., Warren, New Jersey, 1996. Draft Copy.
- [61] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation*. John Wiley & Sons, Inc., Toronto, 1990.
- [62] İ. Şahin. Queues and Related Stochastic Systems: A Unified Equilibrium Theory, Numerical Analysis and Approximations. Technical report, Department of Operations Research and Planning, University of Ottawa, July 1972.
- [63] M. Krunz, R. Sass, and H. Hughes. Statistical Characteristics and Multiplexing of MPEG Sources. In *Proc. IEEE Infocom*, pages 455–462, 1995.

Appendix A

Simulation Using MODSIM

A.1 Introduction

MODSIM is an object oriented language which allows the user to create customized objects and run them in a simulation engine setting.¹ A set of custom objects were created to implement the queueing simulation, organized under the four top-level classes:

- GenTypeObj objects to provide a source of pure traffic such as TES or NG-LRD.
- SourceTypeObj objects to read the output from GenTypeObj objects and produce an output stream of network packets/cells.
- QueTypeObj objects which are queues that accept cells from a source, serve them at a given rate, and then forward them to other objects, and
- the RecordToolObj object, which not directly called by the programmer, is automatically created by any of the above objects to record data while running.

A.2 Class Types

Figure A.1 show the top level classes and their related child classes. The diagram illustrates the static definition of object types and their inheritance from each other.

¹CACI Products Company, La Jolla, CA, USA.

Figure A.1: Class Hierarchy of Objects

The MODSIM class definition files that were created are `Dgentypeobj.mod`, `Dsource-typeobj.mod`, `Dquetypeobj.mod` and `Drecordtoolobj.mod`. They form the class definition half of the source code and contain comment sections giving more detail than is given here on the use of the objects. Paired with each class definition file is an implementation file that provides the source code underlying the procedures/methods referred to in the definition file.

A.2.1 GenTypeObj Objects

These generator type objects provide a pure source of traffic such as TES, E-TES, NG-LRD or SRP. Upon request the object returns a real numerical value from the traffic model which is in this context interpreted as the size of one video frame or similar variable bit rate traffic. The underlying traffic generator functions are all written externally in C (or C++ for NG-LRD) and are called from MODSIM by the generator object to read and initialize the model and to generate frame traffic. The files `libModels.c`, `libFunctions.c`, `libReadModels.c` and `cdfetc.h` comprise the source code of the TES, E-TES and SRP traffic generators.

A.2.2 SourceTypeObj Objects

A source type object provides the source of network packets or cells. In order to produce an output, a fully initialized GenTypeObj traffic generator must be plugged into the SourceTypeObj. The traffic generated by the former is converted into cells/packets which are sent out to a receiving object such as a queue.

The PointCellObj evenly spreads out cells over the length of the defined period and assumes that each cell is a point arrival. A more accurate model is that of the DiscreteCellObj which takes into account the cell line rate in order to discretize the interarrival time points, resulting in a better emulation of a real stream of ATM cells. A spreading factor of 1.0 implies maximum spreading of cells throughout the period, while 0.0 causes cells to be sent out end-to-end according to the line rate. A third source object is the LindleySrcObj which operates in slotted-time such that all the cells generated for a given slot (period) are sent instantaneously at the beginning of the slot. Although this method is less realistic than the cell spreading method, it runs much faster and it is consistent with Lindley's Queue Equation for discrete (slotted) time. Figure A.2 plots the cumulative packet arrival count as a function of time, and compares the output of the continuous-time DiscreteCellObj source to that of the slotted-time LindleySrcObj source.

Figure A.2: Test run of DiscreteCellObj and LindleySrcObj Objects

A.2.3 QueueTypeObj Objects

Queue type objects receive an input cell from a source, place the cell into a FIFO queue, and schedule a departure when appropriate. The output cells can then be sent to another queue or discarded. The service time is constant by default but can in principle be replaced by a more complex function. The queue gathers statistics while it is running such as the cell count received, cell count serviced, cells lost, and time-averaged buffer occupancy.

The continuous-time queues, which were created for use with the continuous-time sources are named `ThreshQueueObj` and `BlockingQueueObj`. The former has an infinite buffer and collects statistics on buffer size while the latter measures ratio of blocked (discarded) cells arriving to its finite buffer. Figure A.3 plots the queue size versus time for the traffic input which was given in Figure A.2. Operating in a slotted or discrete-time mode, the `LindaQueueObj` is fed by the `LindleySrcObj` and obeys the “cut-through” version of Lindley’s Equation: $Q_t = \langle Q_{t-1} - c + A_t \rangle^+$. The `LindBQueue-`

Figure A.3: Test run of Continuous-Time Source Objects

Figure A.4: Test run of Lindley Source Objects

Obj obeys Lindley's Equation in the form: $Q_t = \langle Q_{t-1} - c \rangle^+ + A_t$ thus produces a slightly higher queue occupancy. The operation of the two slotted-time queues is compared in Figure A.4.

A.2.4 RecordToolObj Object

The recording tool object attaches itself to any other object when requested and keeps a running log of the host object's input, output or state to a named ASCII file. All objects of class GenTypeObj, SourceTypeObj and QueTypeObj provide the commands startRecorder, pauseRecorder, resumeRecorder and stopRecorder to control recording.

A.3 Data and Performance Measures

Cell loss ratio and time-averaged buffer occupancy $\Pr[Q > b]$ are two of the most commonly observed measurements. For the purposes of such statistics, cells or packets can be treated as completely anonymous, indistinguishable objects and need virtually no memory space for storage nor time-consuming routines to manipulate them into and out of various objects. Other statistics such as cell delay, cell jitter or cell-arrival-order require that each cell occupy a location in memory to identify itself in some way. If these statistics are required the existing SourceTypeObj and QueTypeObj objects can be modified (by inheritance) into new child objects that support any new cell format.

A.4 Design of the Simulation Program

In MODSIM a main program needs to be written that creates and initializes all the objects needed to start the simulation, and then executes the command Start-Simulation. The simulation engine then schedules and executes the various objects over the duration of the simulation. When all pending events have been cleared,

control passes back to the main program at which time the statistics gathered in the objects can be printed out.

Figure A.5: Component Context Diagram for Mlindleysim.mod

The main simulation program `Mlindleysim.mod` reads in one of the traffic model types (TES, E-TES, NG-LRD, SRP or ASCII file) and creates an appropriate generator object. The generator is then plugged into a Lindley slotted-time cell source which sends its output to a Lindley queue. All three objects may have their recorders enabled for the duration of the simulation. A snapshot of the components assembled together just before calling `StartSimulation` is shown in Figure A.5, and Table A.1 describes the main program in pseudocode. Chapter 9 summarizes the results produced by this simulation.

```

PROGRAM Lindleysim
Input command line arguments
Initialize seeds for the drand48 and NG-LRD runiform functions
Create and initialize a GenTypeObj frame generator
  Ask GenTypeObj to startRecorder
Create LindleySrcObj cell source and initialize it
  Ask LindleySrcObj to attach GenTypeObj
  Ask LindleySrcObj to startRecorder
Create queue LindQueObj and initialize it
  Ask LindQueObj to startRecorder
  Ask LindleySrcObj to send output to LindQueObj
Tell LindleySrcObj to run(duration)
Tell theSimulation to endWarmup(warmuptime)
StartSimulation
Gather statistics from the objects
Dispose of LindQueObj, LindleySrcObj, GenTypeObj, theSimulation
END

```

Table A.1: Pseudo-code for Mlindleysim.mod

Appendix B

An MPEG Generation Scheme

The predicted importance of MPEG traffic makes it a good candidate for traffic modeling, albeit complicated by the fact that it is a non-stationary cyclic process. A technique was required to address two issues. The first was to isolate the aperiodic time dependence structure of the traffic by removing its periodic components, in order to measure its non-cyclic autocorrelation decay. To replicate this autocorrelation the SRP model was chosen for the role of background process generator. The second and more fundamental issue was to address the problem of analyzing and then reproducing the periodic time structure of the IBP pattern. The method employed measures the joint distributions of neighboring I, B and P frame sizes from the empirical sequence, later using the results to modulate the monotonic traffic produced by the background generator.

The MPEG-SRP generator was put to test by modeling a trace of frame sizes from an empirical source, the “BBC video”. The 26000 frame long empirical trace consisted of I, P and B frames repeating in a regular GOP 12 pattern. The time dependence structure seen in the empirical trace was characterized by its autocorrelation coefficient function. It exhibited both a cyclic component, a by-product of the MPEG compression scheme, and an aperiodically decaying component presumably imparted by the information content of the video signal itself. The periodic component reflected the repetitive pattern of the IBBPBBPBBPBB frame sequence

and was very large in amplitude. Partially obscured by the period component, the underlying aperiodic time dependence structure asymptotically decays to zero. In an MPEG modeling exercise [7] this underlying aperiodic decay was described by an autocorrelation function with exponential decay up to a lag k followed by a polynomial decay.

A method was desired to capture and model only the underlying autocorrelation component with a simple model and later introduce the periodic elements, as opposed to capturing both types of components together. It had been noted that each of the I, P, B streams individually exhibit an autocorrelation function that decays monotonically following the underlying long-term dependence structure as described above, and that furthermore the autocorrelations of the three types of streams are nearly identical in shape and decay rate when plotted on a common time scale [44]. For this reason the modeling was based on the autocorrelation function measured only from the primary (I frame) stream. The autocorrelation of the empirical trace I frames is shown in Figure B.1 as a solid curve. The curve does not decay smoothly because of statistical noise but otherwise displays no periodic component.

Figure B.1: Autocorrelation of I-Frames

The SRP model can match any autocorrelation curve provided that it is a perfectly convex, continuously decreasing function from lag 0 to ∞ . As the empirical

curve was not perfectly convex, an idealized curve $LMOD(k)$ was fitted (dash curve) that the SRP generator took as the desired autocorrelation match. This idealized curve was made roughly exponential in shape. but in general could also have been of polynomial or linear piecewise construction. Later experiments would show that the synthesized composite MPEG stream faithfully carries the underlying dependence described by the $LMOD$ curve. Incidentally, if long range time dependence (LRD) had been desired a polynomial shaped tail rather than an exponential one could have been constructed to provide the necessary autocorrelation at large lags.

The marginal distributions of I frames, P frames and B frames were separately measured from the empirical trace. Frame size values were generated using the inverse cumulative distribution (CDF) method but conditionally defined to ensure that I, P and B frame values were statistically related to each other in a joint probability sense.

B.1 Generation of I, P and B Frames

The I frame generator creates I stream values and also acts as master to the P and B generators. The generator implements the SRP technique as described in [44] and matches both the marginal distribution of the empirical I frame sequence and its autocorrelation.

The SRP model is a semi-Markov renewal process in continuous time with a renewal period given by the cumulative distribution function $F1_CHG(x)$. This distribution is obtained from Equation B.1 and is related to the autocorrelation $LMOD$. The scaling factor k_{scale} is introduced to match the time scale of the lag curve to the GOP size and the constant μ is chosen such that $F1_CHG(0)=0$.

$$F1_CHG(t) = 1 - \mu \frac{d}{dt} \left(1 - LMOD \left(\frac{t}{k_{scale}} \right) \right) \quad (B.1)$$

At each renewal time a new I frame value is generated by the inverse CDF technique using the function G_CHG which was obtained from the I frame empirical distribution.

The I frame generator repeats the same value over the length of the renewal interval resulting in a step-function like sample path.

Figure B.2: Block Diagram of MPEG-SRP Generator

The P renewal process will use the same renewal intervals as that of the I Frame generator, hence only an inverse CDF function for generating P frame values needed to be developed. To ensure a proper statistical relationship between generated I and P values a matrix of I-P pairs is measured from the empirical data to create a joint probability density matrix. Creating such a matrix first requires the discretization of the continuous space of I and P values, typically resulting in a continuous space divided into 16 or more evenly spaced intervals. The empirically measured joint probability matrix is then integrated and normalized along each row to yield the conditional cumulative probability matrix $CCPMIP(i, j)$ with rows i corresponding to I intervals and columns j to P intervals. The distribution $\Pr[P \leq x \mid Interval(I) = i]$ is given by row i of the matrix. The selected row forms a piecewise distribution function that is used to generate a P frame value with the inverse CDF technique. At every I stream renewal time (i.e. instants the I frame stream changes value) a new P

frame value is generated using the above process thus creating a P sample path with marginal distribution conditioned on the I stream and with identical renewal periods. Like the I process the P sample path resembles a series of step functions.

The B frame stream is generated from the P frame stream in the same manner that the P stream is generated with reference to the I. The B process uses the same renewal periods as the P process, and generates B values with distribution conditioned on the P stream value.

B.1.1 Experimental Results

Earlier tests with single stream stochastic processes have shown that the SRP method matches the autocorrelation and marginal distribution quite well. But of more concern is whether it will match autocorrelation and marginal for an artificially constructed composite stream like MPEG. It is necessary that I, P and B values are distributed correctly not only with respect to individual marginal distributions but also with respect to the overall frame size distribution. The highly periodic nature of the autocorrelation function should be reproduced at the same frequencies and amplitude as seen in the empirical data, while the long range decaying component must also be present. Twenty independent batch runs of synthetic data each 64000 frames long were generated. Figure B.3 shows the sample path of the empirical stream compared to one of the generated streams. As predicted, the synthetic path P and B frame renewals follow the I frame renewals exactly. The P and B frames also appear to have values which are sensibly related to the I and P frame values respectively, due to the conditional distributions described by the matrices CCPMIP and CCPMPB.

Each of the twenty runs was independently seeded so that the sample histogram and autocorrelation results could be combined over all runs for more accuracy. The multiple runs also provided a way to intuitively confirm the stability of the results even though a rigorous treatment using confidence intervals was not done. The marginal distribution of the empirical path illustrated by the bars in Figure B.4 is

Figure B.3: Sample Path Empirical IBP and Synthetic IBP

compared to that of the synthetic trace shown by plus signs. The agreement is quite good over the entire range. It was found that the number of segments comprising the underlying CDF functions had an effect on the accuracy of the overall marginal distribution output.

The autocorrelation function of the empirical process is compared against that of the synthetic trace in the next two graphs. Figure B.5 plots the autocorrelation of the empirical (line) and synthetic (+) paths for lags 0 through 24 while Figure B.6 plots over lags 0 through 300. The autocorrelation functions compared in Figure B.5 match very well which is promising. They indicate that short term time dependence is being faithfully reproduced, and in particular that the high frequency cyclic component in the autocorrelations match. The cyclic autocorrelation component is also seen in Figure B.6 by the envelope it forms around the autocorrelation functions for both the empirical and synthetic cases. Interestingly the magnitude of the envelope remains fixed. But also seen in this graph is the long term slide or decay of the two functions from lag 0 to 300 due to the decay of the underlying aperiodic autocorre-

Figure B.4: Histogram of IBP Empirical and Synthetic

Figure B.5: Autocorrelation of IBP Empirical and Synthetic

lation. As expected the empirical and synthetic functions do not decay at the same rate since the generator is designed to match the idealized function LMOD and not the true empirical function. In summary, the high frequency components resulting from the GOP compression pattern are accurately reproduced both in frequency and in magnitude, and the underlying long-term non-periodic decay is also reproduced.

Figure B.6: Autocorrelation of IBP Empirical and Synthetic

B.2 MPEG Scene Changes

The term “scene-change” applied to video denotes sudden changes in video signal or information rate typically due to a cut from one scene to another. More specifically in this chapter, the term will refer to points where the MPEG compressor has decided to break the regular GOP pattern at an arbitrary point in the stream by restarting a new pattern due to a scene-change or other event. Both the empirical MPEG samples available in the lab and the synthetic MPEG generator as described so far do not implement a scene-change feature and this results in an unbroken repeating GOP pattern. Consequently the envelope formed around the autocorrelation function (Figure B.6) has a magnitude which does not decay over time. Alternately the GOP pattern and its harmonics appear as large high frequency spikes in the frequency domain. But it is likely that realistic MPEG compressors will implement scene-change detection and produce MPEG streams with occasional breaks in the regular GOP pattern. Such an MPEG stream would be expected to yield an autocorrelation function looking something like that in Figure 6 of reference [63] where the periodic envelope dies away as the lag increases from 0 to 120 time units.

By adding a new section of code at the top of the main loop, the MPEG

generator began injecting random disruptions into the GOP pattern. Scene durations were determined by an exponential renewal process with parameter γ . At a scene-change instant the GOP pattern is immediately reset to generate an I frame followed by a P frame after which the pattern continues normally but out of phase with the old pattern. Implementing a scene-change mechanism in this manner was expected to dampen the cyclic components in the autocorrelation especially at long lags, and more importantly not to affect the underlying non-cyclic long-term correlation.

B.2.1 Experimental Results

A new run of synthetic data was generated with input parameters similar although not identical to the previous run in Section B.1.1, and with a scene-change mechanism set to $\gamma = 600$ for a mean scene length of 20 seconds. The autocorrelation function of the synthetic path is shown in Figure B.7-top from lags 0 through 600. The autocorrelation function around 0 lags starts out with an envelope roughly positioned and of same height as that in Figure B.6, but then the envelope decays exponentially as the periodic GOP component disappears from the autocorrelation. Unfortunately a comparable empirical MPEG sample path implementing scene-change was not available for comparison. However, the underlying non-periodic autocorrelation can still be isolated by considering the GOP-level autocorrelation function as is done in Figure B.7-bottom for the empirical stream (solid curve) versus the synthetic stream (plus signs). This clearly shows that the scene-change mechanism while damping the periodic components (of the GOP pattern) does not affect the aperiodic long-term decay component inherent to the video itself.

B.3 Implementation

All the calculation and generation was done in the MATLAB environment. The first procedure `mpeg.m` needs to be run once to perform the preliminary calculations. The

Figure B.7: Autocorrelation Functions with MPEG Scene Changes

second procedure `mpegrun.m` is called at run-time to generate a synthetic stream with the variables and functions provided to it by the first routine.

- The routine `mpeg.m` reads in an empirical IBP path and separates the streams by I, P and B type. From the I stream a cumulative distribution function `G_CHG` is created. Matrices `CCPMIP` and `CCPMPB` are created to specify the joint distributions of the P and B streams. The autocorrelation function `LMOD` is used to calculate the renewal distribution `F1_CHG` for the I frame generator. All functions are described by piecewise linear curves stored in 2-dimensional arrays.
- The main loop of function `mpegrun.m` is executed once for each frame generated. Inside the main loop, a determination is made of the frame type to generate according to the GOP pattern and scene-change state and an appropriate sub-section is executed.

B.4 Discussion

The calculation and generation routines written in MATLAB were prototypes for experimentation and were not intended to be fully automatic. The routine does automatically capture and process the IBP pattern from the given empirical specimen, however the autocorrelation measurement and fitting requires human input.

The generation algorithm is potentially very fast because it only requires significant calculation at SRP renewal instants, while within a renewal period the output streams stay constant. At the renewal instants, new values for I, P and B streams and a new SRP renewal period are generated with light-weight inverse CDF techniques. Similarly the scene-change events only occur once in a while and the processing at these times is also light-weight. By contrast more sophisticated mathematical models, such as FGN by Hosking's method which is briefly illustrated in [7] are computationally expensive and grow non-linearly in cost as sample path length grows.

The MPEG-SRP generator outlined in this chapter is constructed in a robust way without reliance on conceptually or computationally difficult techniques. The synthetic output faithfully matches the marginal distributions and the autocorrelation structures, both periodic and aperiodic. Choice of different MPEG GOP patterns is not expected to pose difficulties to the model, and the scene-change feature (if implemented) is independently controllable.

An issue inviting further investigation is the applicability of the MPEG generator technique with other sources of background traffic besides SRP. The system block diagram in Figure B.2 has isolated the operation of the background traffic source within the SRP Module block, such that it can be replaced with another generator design. The IBP pattern and scene change mechanisms are driven by this module but otherwise the interdependence between them is limited. Traffic models such as TES and NG-LRD produce natural-looking traces with controllable marginal and autocorrelation characteristics, and thus may be better candidates for the role than SRP.

Appendix C

An FGN Generator Using SRP

The Fractional Gaussian Noise (FGN) process is used at the core of a self-similar traffic generator described in a recent paper by Huang et al. [6]. The FGN process is exactly second-order self-similar and is characterized by the Hurst parameter H . Hosking's method, used to generate an FGN sample path, increases in computational effort with order $O(n^2)$ as the sample path length grows. In practice this has limited simulation runs to a maximum of about 5×10^5 samples. An alternative to Hosking's method was sought to play the role of the FGN generator and do it quickly and efficiently while allowing some loss in fidelity. The SRP generator model developed in the first section seemed like a good candidate because it could both match marginal, in this case a Gaussian $N(0, 1)$, and reproduce the convex shaped autocorrelation function of FGN for reasonably long lags.

C.1 Description

The autocorrelation co-efficient function $\rho(k)$ of an exactly second-order self-similar FGN process with parameter $0.5 < H < 1$ is given in discrete-time as:

$$\rho(k) = \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) \quad k = 1, 2, 3, \dots \quad (\text{C.1})$$

It was desired to produce a traffic with interpolated autocorrelation function described by the continuous-time function (C.2) but it was noted that the convexity conditions

for the renewal process were not being maintained in the region $t \in [0, 1)$. The problem was overcome by redefining the autocorrelation function in Equation C.3 which substitutes a linear segment between lag 0 and 1 to maintain convexity of the curve.

$$\rho(t) = \frac{1}{2}(|t+1|^{2H} - 2|t|^{2H} + |t-1|^{2H}) \quad t > 0 \quad (\text{C.2})$$

$$\rho(t) = \begin{cases} \frac{1}{2}(2 + (2^{2H} - 4)t) & 0 \leq t < 1 \\ \frac{1}{2}((t+1)^{2H} - 2t^{2H} + (t-1)^{2H}) & 1 \leq t \end{cases} \quad (\text{C.3})$$

The renewal function was calculated by applying Equation 7.2 to (C.3) with the constant μ chosen such that $F(0) = 0$.

$$F(t) = 1 + \mu \begin{cases} 2^{2H-1} - 2 & 0 \leq t < 1 \\ H(t+1)^{2H-1} - 2Ht^{2H-1} + H(t-1)^{2H-1} & 1 \leq t \end{cases} \quad (\text{C.4})$$

$$\mu = -(2^{2H-1} - 2)^{-1} \quad (\text{C.5})$$

Figure C.1 plots the autocorrelation function of an FGN process with Hurst parameter $H = 0.75$. The renewal function $F(t)$ is given by Equations 7.2 and C.4 and is plotted in Figure C.2. A piecewise linear approximation of $F(t)$ was used to specify the renewal distribution to the SRP generator program. The marginal distribution was described as a piecewise approximation to the Gaussian $N(0, 1)$ distribution.

Incidentally, the fact that the required output was Gaussian allows one to aggregate multiple SRP streams together, producing a better marginal distribution as well as mitigating the variability seen in the sample path and empirical autocorrelation curves as noted in section 7.2.

C.2 Results

A modified SRP generator engine was created to produce 10 independent streams in parallel, each stream producing an approximately Gaussian marginal distribution and an autocorrelation based on the renewal function F . One test trace 300000 samples

Figure C.1: Analytical Autocorrelation Curve with $H=0.75$ Figure C.2: Renewal CDF Function with $H=0.75$

Figure C.3: Empirically measured Marginal Distribution of Gaussian10 process

long was generated and its marginal distribution as seen in Figure C.3 is very close to true Gaussian $N(0, 1)$. The autocorrelation empirically measured from the test trace is compared to the predicted analytical curve in Figure C.4 and it looks quite good over the range of lags shown. Perhaps more interesting is the variance-time plot made of the test trace, showing that the rate of decay of the variance with increasing (partial sum) aggregation size is decreasing at a rate of $\hat{\beta} = 0.5$ for much of the range shown. As a check an estimate of the Hurst parameter is given by $\hat{H} = 1 - |\hat{\beta}|/2$. By comparison an IID Gaussian series also included in the V-T plot decays with rate 1.0. Beyond the aggregation size of about 2^{12} statistical noise begins to creep in.

C.3 Discussion

There is room to fine tune the generator, with regard to specifying the marginal distribution function and the renewal function more accurately to improve the marginal and autocorrelation results, respectively. Yet even at this stage it appears that this SRP based generator might replace Hosking's method when truly exact self-similar traffic is not needed. The test program `GaussianSrp.c` was written to generate the traffic. It reads in the SRP input file `GaussianSrp.model` and generates 10 identical but independent SRP streams which it then aggregates and divides by the nominal standard deviation $\sqrt{10}$ to produce pseudo-FGN output.

Figure C.4: Empirical Vs Analytical Autocorrelation of Gaussian10 process

Figure C.5: Variance-Time Plot of IID process vs Gaussian10 process

IMAGE EVALUATION TEST TARGET (QA-3)

APPLIED IMAGE, Inc.
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved