Traffic Modelling for Broadband Networks -with Emphasis on Self-Similarity

Dissertation Submitted to
JAWAHARLAL NEHRU UNIVERSITY
in partial fulfillment of requirements
for the award of the degree of
Master of Technology
in
Computer Science

by
Polimetla Daiva Kumar

SCHOOL OF COMPUTER & SYSTEMS SCIENCES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI - 110 067
January 1998

CERTIFICATE

This is to certify that the dissertation entitled **Traffic modeling for broadband networks - with emphasis on self-similarity** being submitted by **Polimetla Daiva Kumar** to the school of computer and systems sciences, Jawaharlal Nehru University, New Delhi in partial fulfillment of the requirements for the award of the degree of **Master of Technology in computer science** is a bonafied work carried by him under the guidance and supervision of **Prof. Karmeshu**.

The matter embodied in the dissertation has not been submitted for the award of any degree or diploma.

Prof. Karmeshu 7/1/98
SC&SS, JNU,
New Delhi.

(P. DAIVA KUMAR)

Prof. P.C.Saxena
Dean,
SC&SS, JNU,
New Delhi.

ACKNOWLEDGMENTS

I wish to convey my heartful gratitude and sincere acknowledgments to my guide **Prof. Karmeshu**, School of Computer and Systems Sciences, Jawaharlal Nehru University for his wholehearted, tireless and relentless effort in helping me throughout in completing this project.

I would like to record my sincere thanks to my dean **Prof. P.C.Saxena**, School of Computer and Systems Sciences, JNU for providing the necessary facilities in the centre for successful completion of this project.

I would take this opportunity to thank all of my faculty members and friends for their help and suggestions during the course of my project work.

P.Daiva Kumar

..... dedicated to

my beloved parents, brother and sister .....

Contents

Figure contents

Abstract

Broadband networks, utilizing ATM, are expected to provide the information transport for a rich mixture of services (voice, video, and data etc.) and applications. These are associated with broad spectrum of traffic types and transport performance requirements. It is therefore necessary to develop traffic models to capture the statistical features of actual traffic characterized by marginal distribution, auto-correlation structure, traffic delays and cell loss probabilities. Such models will facilitate network performance studies in realistic situations.

Recently, statistical analysis of different types of network traffic has shown that network traffic is self-similar or fractal in nature. A characteristic feature of self-similarity is the existence of a heavy-tailed distribution which behaves like a power law. It is proposed to develop a new generalized traffic model mimicking the power law behaviour. We have also developed a self-similar traffic model based on random-walk with transition probabilities having infinite first moment.

The dissertation consists of five chapters. The first two deal with brief review of ATM characteristics and traffic models respectively. Chapter 3 is concerned with self-similarity aspects and power law behaviour in traffic models. In chapter 4, we present a new generalized model for traffic distribution. The last chapter deals with random walk based self-similar traffic model. Finally, the dissertation ends with conclusions.

1

Introduction

In the evolution from the current telecommunication networks towards the Broadband Integrated Digital Services Network (B-ISDN), some important directions and guidelines have recently been made. The recent directions opened up by the B-ISDN are influenced by a number of parameters, the most important being the emergence of a large number of tele-services with different, sometimes yet unknown requirements. In this information age, customers are looking for an ever increasing number of new services. The most promising teleservices to appear in the future are HDTV (High Definition TV), video conferencing, high speed data transfer, videophony, video library, home education, and video on demand. Each of these services will generate other requirements for the B-ISDN. This large set of requirements introduces the need for one universal network which is flexible enough to provide all of these services in the same way. The other parameters which are affecting the directions suggested by the B-ISDN are the fast evolution of semiconductor and optical fiber technology [1].

The traditional circuit switching, currently used in telephone network has some limitations because it is incapable of satisfying the wide range of bandwidth required for varying services such as video, voice, and multimedia services. It is also inefficient for switching bursty traffic because it establishes a dedicated communication path for the duration of the connection without regard to whether the information is being actually transmitted or not. Fast packet switching is complicated, requiring node-to-node protocols to deliver error free packet and can result in large delay and increase in variance of the packet transmission time. Fast packet switching is a concept that covers several alternatives, all with the same basic characteristic, i.e. packet switching with minimal functionality in the network [2].

Asynchronous Transfer Mode (ATM) is the fast packet switching and multiplexing technology that has been standardized for the B-ISDN. Here, asynchronous refers to the manner in which bandwidth is allocated among connections and users. Bandwidth is divided into time slots of fixed length. These time slots are allocated for user on demand [3].Traffic mode is a term intended to signify that it uses the multiplexing and switching technique. There are many reasons for choosing packet switching over circuit switching. Firstly, packet switching is highly flexible and can handle both constant rate traffic (audio, video) and variable rate traffic (data) easily. Secondly, looking to transmission at very high speeds (Gbps), digital switching of cells is easier than using traditional multiplexing techniques, especially using fiber optics. Thirdly, broadcasting is essential for television distribution, which cannot be possible with cell switching [4].So, ATM is a compromise between pure circuit switching and packet switching.

1.1 Asynchronous Transfer Mode (ATM)

The basic idea underlying ATM is transmission of all information in small, fixed size packets called *cells*. The length of ATM cell is 53 bytes, consisting of 5 bytes of header information followed by 48 bytes of data. The main reason for choosing 48 bytes of data length was the reduction of packetization delay of voice, reduction of the internal buffers in the switching nodes, and limiting the queuing delays in those buffers [1]. A typical ATM cell format is shown in the fig.1.1.

5 bytes	48 bytes
Header	User data

Figure 1.1 ATM cell format

ATM is connection oriented and supports two types of connections, such as point-to-point and point-to-multipoint. The connection is half-duplex. Then, for two way communication it needs two connections. If any end station wants to set up the connection with the network, then initially it has to send a message for the connection. So, if the resources are available, then the connection can be given. ATM is guaranteed for the cell delivery order but there can be cell loss. The intended speeds for ATM networks are 155 Mbps and 622 Mbps with the possibility of Gigabit per second speeds at a later stage. The 155 Mbps speed was chosen to transmit high definition television. And the 622 Mbps speed was chosen so that four 155 Mbps channels could be sent through it.

The advantages of ATM are :

- Flexibility to support existing services and unforeseen future services;
- Dynamic bandwidth allocation;
- Integrated transport of all information types;
- Efficient utilization of network resources by statistical sharing.

The disadvantages are that cells may suffer variable delays through the network or may be lost. The network must ensure the quality of service, in terms of cell delays and cell loss rate, required by the users. In addition, considerable processing is involved in converting user information into and from the ATM cell format, and in carrying cells at high rates through each switch [4].

1.2 The B-ISDN ATM Protocol Reference Model

Broadband ISDN using ATM has its own reference model based on standards developed by the ITU. The protocol reference model based for ATM is divided into three layers: the physical layer, the ATM layer, and ATM adaption layer plus whatever the users want to put on top of that [4] as shown in the fig 1.2.

1.2.1 Physical layer

The physical layer defines a transport method for ATM cells between two ATM entities. It encodes and decodes the data into suitable electrical/optical waveforms for transmission and reception on the communication medium used. It also provides cell

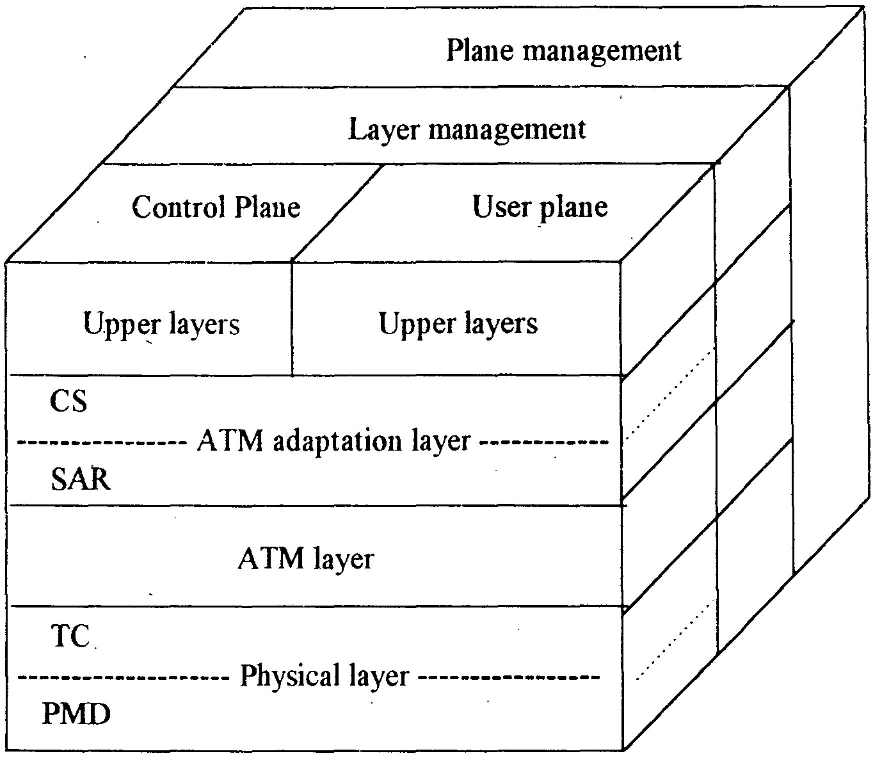

Figure 1.2 The B-ISDN ATM reference model.

transmission and reception on the communication medium used. It also provides cell delineation functions, header error check (HEC) generation and processing, performance monitoring and payload rate matching of the different transport formats used at this layer. It is divided into two sublayers: Transmission Convergence (TC) sublayer and Physical Medium Dependence (PMD) sublayer. TC sublayer is responsible for mapping of the ATM cells to the transmission system used. And the PMD sublayer is responsible for the correct transmission and reception of bits on the physical medium [5].

1.2.2 ATM layer

The ATM layer is responsible for cell relaying between ATM layer entities, cell multiplexing of individual connections into composite flow of cells, cell demultiplexing of composite flow into individual connections, cell rate decoupling or unassigned cell insertion and deletion, priority processing and scheduling of cells, cell loss priority marking and reduction, cell rate pacing and peak rate enforcement, explicit forward congestion marking and indication, cell payload type marking and differentiation, and generic flow control access. The functionality of the ATM layer is defined by the fields present in the ATM cell header. A typical ATM cell header is shown in fig. 1.3.

Cell (53 bytes)					
Header					Information
Generic flow control	VCI/VPI field	Payload type indicator	cell loss priority	Header check-sum	Payload
4 bits	24 bits	2 bits	2 bits	8 bits	48 bytes

Figure 1.3 ATM cell header.

The cell header consists of a generic flow control (GFC) field, the VCI/VPI fields, a payload type indicator (PTI) field, a cell loss priority (CLP) field, and a header checksum field. The GFC field supports the implementation of a flow control mechanism on the user network interface. The VCI/VPI fields are used for channel identification and simplification of the multiplexing process. The PTI field is used to distinguish between user cells and control cells. The CLP field is used to indicate whether a cell may be discarded during periods of network congestion. The header checksum field is used to protect the header field from transmission errors.

1.2.3 ATM adaption layer (AAL)

The ATM adaption layer enhances the service provided by the ATM layer to the higher network layers. It performs functions for the user, control and management planes which depend on higher layer requirements. The AAL is subdivided into two layers: the segmentation and reassembly sublayer (SAR) and the convergence sublayer (CS). The SAR sublayer is used for segmentation of the higher layer information into a size suitable for the payload of consecutive ATM cells of a virtual connection, and the inverse operation, reassembly of contents of the cells of a virtual connection, into data units to be delivered to the higher layers. The convergence sublayer performs functions like message identification, time/clock recovery, etc. [1].

Upto now four service classes have been defined, based on three parameters: time relation between the source and the destination, constant or variable bit rate and connection mode.

The service classes are: class A, class B, class C, and class D.

Class A - a time relation exists between the source and the destination, the bit rate is constant, and the service is connection oriented (e.g. a voice channel).

	class A	class B	class C	class D
Timing between source and destination	required		not required	
Bit rate	constant	variable		
connection mode	connection-oriented			connection-less

Figure 1.4 service classes in B-ISDN.

Class B - a time relation exists between the source and the destination, the bit rate is variable, and the service is connection oriented (e.g. a video or audio channel).

Class C - no time relation exists between the source and the destination, the bit rate is variable, and the service is connection-oriented (e.g. a connection oriented file transfer).

Class D - no time relation exists between the source and the destination, the bit rate is variable, and the service is connectionless (e.g. LAN interconnection and electronic mail).

Initially four types of AAL protocols were recommended by ITU to support the four service classes A, B, C and D, called types 1,2,3 and 4 respectively. Later, AAL3 and AAL4 were merged into a single type (called AAL type 3/4), since the differences between them were minor. After that, a fifth AAL type has also been defined, in view of the high complexity of AAL3/4 [5]. The AAL also plays a key role in the internet-working of different networks and services. So, cell transmission in ATM networks can be done by passing through the devices known as ATM switches.

1.3 ATM switching

An ATM switch is composed of a number of input ports and output ports and a network which connects input ports to output ports called *switch fabric*. As shown in

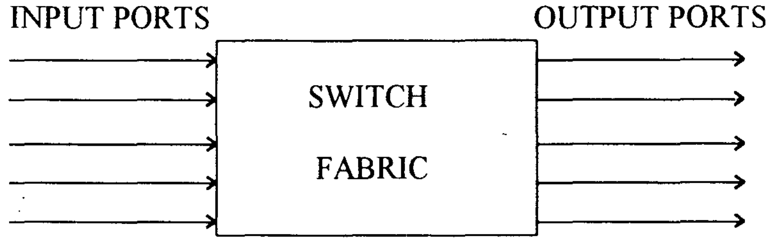

Figure 1.5 ATM switch.

fig. 1.5, cells arrive at the input port asynchronously, pass through the fabric and are eventually transmitted on the appropriate output line. A number of different switching fabrics exist including simple crossbars and Batcher Banyan networks.

All ATM switches have buffers, so that they can store the cells, if they are arriving too quickly or there is congestion in the switch fabric. Buffers can be at the input ports (input buffering) or at the output ports (output buffering) [6]. Most of the ATM switches use a combination of input and output buffering. ATM switches use the VPI and VCI fields of the cell header to identify the next network segment that a cell needs to transmit on its way to final destination. A virtual channel is equivalent to a virtual circuit - that is, both terms describe a logical connection between the two ends of a communications network. A virtual path is a logical grouping of virtual circuits that allows an ATM switch to perform operations on groups of virtual circuits.

When an end station connects to the ATM network, it is essentially making a contract with the network based on Quality of Service (QoS) and parameters such as peak bandwidth, average sustained bandwidth and burst size. It is the responsibility of the ATM device to adhere to the contract by means of *traffic shaping,* which will use the queues to constrain data bursts will limit the peak data rate, and smooth jitter so that the traffic meet the prescribed QoS parameters. ATM switches have the option of using *traffic policing* to enforce the contract. The switch can measure the actual traffic flow and compare it against the agreed upon QoS parameters. If it finds that the traffic is outside these parameters, the switch can discard the cell by setting the CLP (cell loss priority) bit during the periods of congestion.

Congestion will occur when there is contention for limited resources (i.e. buffer, bandwidth and processors); it is exhibited when the offered traffic load exceeds the network's designed capacity. Because ATM is connection oriented and packet switched, ATM networks may exhibit congestion at both the levels of connection and cell. At the level of connections, call processors will become preoccupied with unsuccessful call attempts. At the level of cells, transmission links will become saturated with traffic, and buffers will overflow with cells. Hence, to eliminate the congestion we need to control the traffic. The primary role of traffic control in B-ISDN is to protect the network and the user in order to achieve predefined network performance objectives (e.g. in terms of cell loss probability or cell transfer delays) and to optimize the use of network resources [1].

1.4 Need to develop Traffic Models

One of the most important aspects concerning ATM networks is the issue concerning traffic in the transmission line. The idea is to develop traffic models which can reproduce traffic patterns observed in reality. To achieve better performance of the network, we need to develop good traffic models. Accurate traffic models can capture the statistical characteristics of actual traffic. If traffic models do not accurately represent the actual traffic, one may overestimate or underestimate the network performance. Since a model is viewed as a faithful representation of the target system, instrumentation is required in order to collect statistics and formulate performance predictions. More often, the desired response of the system is a function of a parameter vector. For example, the cell loss rate in an ATM switching network is a function of a set of congestion-control parameters. Correlation in the measured observations must be taken into account when forming statistical performance estimates. For example, positive auto-correlation in a sequence of delay times would cause bursts of long (short) delays. Thus, if message k has experienced a long (short) delay, then long (short) delay for message $k+1$ becomes probable. Furthermore, rare events, such as ATM cell losses, are key issues that characterize the performance of emerging broadband networks [7]. Thus, the models that capture the auto-correlated nature of traffic are essential for predicting the performance of emerging broadband networks.

The aim of this project is to develop some new traffic models, by considering the self-similarity nature of traffic.

2

Overview of Traffic Models

The efficient flow of information is a key element in today's technological environment. As discussed in the first chapter, this is possible with the high-speed networks, such as ATM, if it is properly designed and operated. Traffic modelling is a key element in designing and operating the communication networks. Hence, it is necessary to understand the nature of traffic, in order to select the appropriate random traffic model for the better performance of the network. This overview closely follows the articles [7] and [8].

2.1 Introduction

In this chapter, the main discussion is about the general traffic models in broadband networks. Basically, traffic models can be stationary or non-stationary. In general, stationary traffic models can be divided into two types: short-range and long-range dependent. Short-range dependent models consists of significant correlation structure for small lags. Examples are Renewal processes, Markov processes and Regression models. Long-range dependent traffic models have significant correlations for large lags. Examples are Fractional AutoRegressive Integrated Moving Average (F-ARIMA) and Fractional Brownian motion.

Traffic models can be analysed on the basis of the number of parameters required to describe the mode, parameter estimation, analytical tractability and perfect-fit. To evaluate perfect-fit, it is necessary to determine how "close" the model is to the actual data., i.e. how far the model is able to give marginal distributions, auto-correlations structure and ultimately to predict delays and cell loss probabilities [8].

For calculating performance measures, the most common modelling context is queueing, where traffic is offered to a queue or to a network of queues [7]. Simple traffic consists of single arrivals of discrete entities like packets. This can be described as a point process, consisting of a sequence of arrival instants $T_1, T_2, \ldots, T_n \ldots$ measured from the origin 0 (i.e. $T_0 = 0$). These point processes have two additional equivalent descriptions -- counting processes and interarrival time processes. A counting process $\{N(t)\}_{t=0}^{\infty}$ is a continuous-time, non-negative integer-valued stochastic process, where $N(t) = \max\{n: t_n \leq t\}$ is the number of (traffic) arrivals in the interval (0,t]. An interarrival time process is a non-negative random sequence $\{A_n\}_{n=1}^{\infty}$, where $A_n = T_n - T_{n-1}$ is the length of the time interval from the n^{th} arrival to the $(n-1)^{th}$. Since $T_n = \sum_{k=1}^{n} A_k$, the equivalence of these descriptions follows from the evident relation :

$$\{N(t) = n\} = \{T_n \leq t \leq T_{n+1}\} = \left\{\sum_{k=1}^{n} A_k \leq t \leq \sum_{k=1}^{n+1} A_k\right\} \quad (2.1)$$

2.2 Renewal Traffic Models

There are two special types of these models based on Poisson process and Bernoulli process. These models are analytically simple, when A_n are independent, identically distributed (IID); however their distribution is allowed to be a general one. Unfortunately, the superposition of independent renewal processes does not always yield a renewal process. These processes also have a serious limitation in modelling as the auto-correlation function of $\{A_n\}$ vanishes similarly for all non-zero lags.

2.2.1 Poisson Process

This renewal process has the inter-arrival times $\{A_n\}$ distributed exponentially with rate parameter λ such that $P\{A_n \leq t\} = 1\text{-}\exp(-\lambda t)$. It is also a counting process and number of arrivals in subsequent intervals is statistically independent.

Poisson process has some special analytical properties. The superposition of independent Poisson processes results in a new Poisson process whose rate is the sum of the component rates. This process has independent increment property rendering it a memory-less process, which in turn, greatly simplifies queueing problems involving Poisson arrivals. On the basis of Palm's theory, this process support many traffic applications, which comprise large number of independent traffic streams. But the main limitation is, that, traffic aggregation (multiplexing) may not always result in a Poisson stream [7].

2.2.2 Bernoulli Process

This process is the discrete-time analog of Poisson process. Thus, the probability of an arrival in any time slot is p, independent of all others. So, the number of arrivals in slot k follows binomial distribution such that, $P\{N_k = n\} = \binom{k}{n} p^n (1-p)^{k-n}$, $0 \le n \le k$ (see [7]).

2.2.3 Phase-Type Renewal Process

These processes will occur when the inter-arrival times follows the phase type. For modeling phase-type inter-arrival times we consider the time to absorption in a continuous time Markov process $C = \{C(t)\}_{t=0}^{\infty}$ having the state space {0,1,..,m} with state 0 as the absorbing state while all other states being transient [7]. The advantage of using these type of models is that the analysis of the problem becomes tractable.

2.3 Markov and Embedded Markov Models

In many situations, a source can be modeled by a finite number of states characterizing finite different activities. For example, a finite state model in voice telephony as shown in fig. 2.1 is widely used. In this model, a voice source can either be in idle state or in busy state with the source transmiting the packets only in busy state (i.e. speech activity). In general, we can increase the number of states, for more accurately describing a situation, but one has to pay the price by increasing computational complexity.

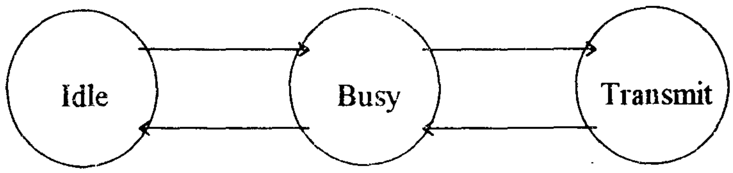

Figure 2.1 Finite state model for voice.

Defining by the random variable X_n, the state from the given state space S = $\{s_1,s_2,....,s_M\}$ at time n. The set of random variables $\{X_n\}$ will form a discrete Markov chain, if the probability of the next value $X_{n+1} = S_j$ depends only on the current state [8,9]. The Markov chain can be called as *discrete time Markov chain*, if state transitions occur at integer time epochs (0,1,....,n,...). When transitions occur in continuous time, the Markov chain is called the *continuous time Markov chain*. This means, that the future value depends on the current state but not on previous states and not on the time delay spent in the current state. Because of this, the time spent in a state is restricted to geometric distribution in discrete case and to an exponential distribution in the continuous case.

The stochastic process is called a Semi-Markov process when the time between state transitions follows an arbitrary probability distribution. An embedded Markov chain can be obtained by ignoring the time distribution between transitions and the sequence of states visited by the Semi-Markov process will be a discrete time Markov chain [8].

2.3.1 ON-OFF Models

The ON-OFF source model is widely used for the voice traffic with transitions shown in fig. 2.2. In this model, packets are generated only during talk spurts (ON state)

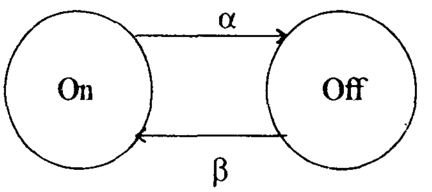

Figure 2.2 ON-OFF model.

with fixed A_n. The time spent in ON and OFF states is exponentially distributed with mean α^{-1} and β^{-1}, respectively [8]. This model provides a simplest description for the voice traffic. However, in realistic situations this model fails to capture bursty nature of traffic.

2.3.2 IPP Models

The Interrupted Poisson Process (IPP) can be in states such as active and idle as shown in fig. 2.3. and arrivals occur only during the active state. Thus IPP model differs from ON-OFF model in inter-arrival time during the active state [8].

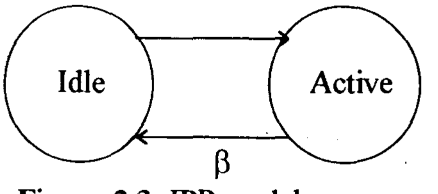

Figure 2.3 IPP model.

2.3.3 Markov Modulated Poisson Process

Morkov Modulated Poisson Process (MMPP) is the most commonly used model for examining the performance measures. This process can be considered as a doubly stochastic Poisson process. This model combines the simplicity of the modulating (Markov) process with that of the modulated (Poisson) process as shown in fig. 2.4. The Markov process uses the current state to control (modulate) the probability distribution of the traffic [7].

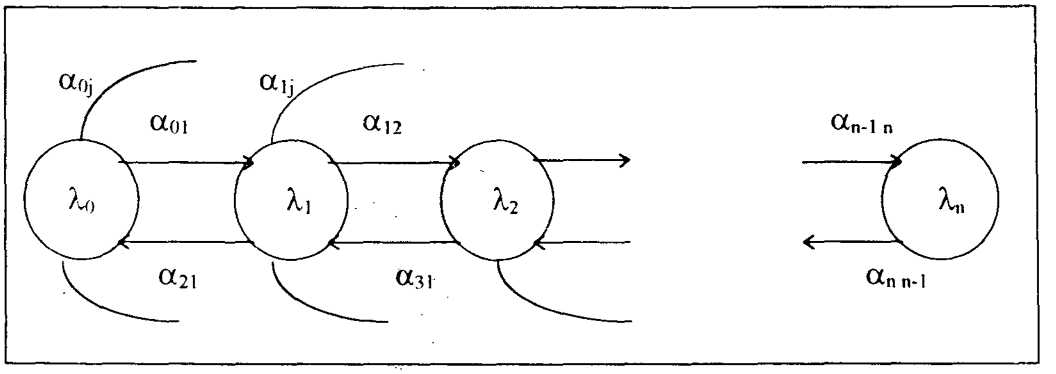

Figure 2.4 MMPP process

In this model, the arrivals occur according to a Poisson process at the rate λ_k, while in state s_k. Thus the rate changes according to the state. The MMPP parameters can be estimated by quantizing the arrival rate into finite number of rates associated with the number of states. This means each rate is associated with a state in Markov chain. For example, if we denote q_{ij} as the transition rate from state i to state j, then q_{ij} can be estimated by quantizing the empirical data and by calculating the fraction of times state i switched to state j [8].

2.3.4 Morkov Modulated Fluid Models

The fluid traffic models treat traffic as a stream of fluid, characterized by a flow rate (such as bits/sec). These models are appropriate when we consider situations with heavy such that individual units such as traffic packets have little impact on the performance of the network. It may be noted that, all models that distinguish between cells and consider the arrival of each cell as a separate event, require vast amount of memory and CPU resources. In contrast, these fluid models characterize the incoming cells by a flow rate [7].

In Markov modulated fluid models, the current state of the underlying Markov chain specifies the traffic flow rate. Generally, this model is employed in modeling the VBR video sources [10]. For simulating fluid model, one assumes that the incoming fluid (traffic) flow remains (roughly) constant over much longer time periods and traffic fluctuations can be modeled by events reflecting a change of flow rate. In a queueing context, the buffer waiting time corresponds to the time it takes to serve (clear) the current buffer, and loss probabilities (at a finite buffer) can be calculated in terms of overflow volumes. The disadvantage of this model is the existence of many parameters, and exponentially decaying auto-correlation function. Besides these disadvantages, queueing performance of this model is still analytically tractable.

2.4 Regression Models

These models express the random variable in the sequence in terms of previous ones in the presence of a white noise. We briefly discuss some of the commonly used regression models [8].

2.4.1 Linear Autoregressive Models

The linear Autoregressive model of order p, denoted by AR(p), has this form:

$$X = \phi 1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \qquad (2.2)$$

where X_t's are correlated random variables, ϕ_j's are real constants and ε_t is white noise. Accordingly, X_t's will be normally distributed random variables, if ε_t is a white Gaussian noise with variance $\sigma^2_{\varepsilon_t}$. In terms of lag operator B as $X_{t-1} = BX_t$, we can rewrite (2.2) as,

$$\phi(B)X_t = \varepsilon_t \tag{2.3}$$

where $\phi(B) = (1-\phi_1 B-\ldots-\phi_p B^p)$. Multiplying eq. (2.2) with X_{t-k} and taking the expectation, we get,

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}; for \quad k \geq 0 \tag{2.4}$$

Where ρ_k is auto-correlation function.

Thus the general solution is,

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k, \tag{2.5}$$

where G_i^k's are roots of $\phi(B)$. There fore, in general, the auto-correlation function of AR(p) process will consist of damped exponentials depending on the roots (real or imaginary). These AR models have been used to model the output bit rate of VBR encoder, because successive video frames are not likely to vary much visually. Accordingly, the output bit rate within a frame period is constant and changes from frame to frame based on the following AR(1) model :

$$\lambda[n] = \phi\lambda[n-1] + b\varepsilon[n] \tag{2.6}$$

Here $\lambda[n]$ is the bit rate during frame n and $\varepsilon[n]$ is a Gaussian white noise. $\varepsilon[n]$ is chosen so that the probability of $\lambda[n]$ being negative is very small. For computational aspects, whenever $\lambda[n]$ is negative, it is set to zero in eq. (2.6), as the number of bits in frame n cannot assume negative values. Thus, the bit rate of frames within the scenes can be modeled as AR process and the scene changes can be modeled as Markov chain. The limitation of this framework is that it fails to capture the sudden changes in the frame bit rates that occur due to scene changes in video.

In a more general setting, simple AR(2) has been used to model variable bit rate (VBR) coded video. One plausible model for VBR video traffic is

$$X_n = Y_n + Z_n + V_n C_n \tag{2.7}$$

where Y_n and Z_n are two independent AR(1) processes and $V_n C_n$ the product of a simple Markov chain and an independent normal random variable. With a view to achieving a better fit to empirical auto-correlation function, it has been suggested to employ two AR schemes; the third term, V_nC_n is proposed to capture sample path spikes due to video scene changes.

Since the auto-correlation function decays exponentially, this model is still unable to capture the slow decay in comparision to exponential decay. It needs to be emphasized that, AR processes with Gaussian distribution cannot capture VBR video traffic probability distribution, since VBR video traffic distribution exhibits a power law behaviour which is not depicted by Gaussian [8].

2.4.2 Discrete Autoregressive Models

A discrete autoregressive model of order p, denoted by DAR(p), generates a stationary sequence of a discrete random variables with an arbitrary probability distribution yielding auto-correlation structure similar to that of an AR(p). It is plausible that, the number of cells per frame of teleconferencing VBR video be modeled by DAR(1) with negative binomial distribution.

The advantage of this model is the simplicity in parameter estimation with less number of parameters. Though, the resulting process distribution is arbitrary, the analytical queueing performance is tractable. On the other hand, since the auto-correlation function decays exponentially as in AR(p), it cannot used to model the traffic with a slower auto-correlation decays [8].

2.4.3 Autoregressive Moving Average Models

An Autoregressive Moving Average model of order (p,q), denoted by ARMA(p,q), is used for modeling VBR traffic. In this case too the auto-correlation function decays exponentially. The duration of a video frame is divided into m time intervals equally, such that the number of cells in the n^{th} time interval can be modeled as :

$$X_n = \phi . X_{n-m} + \sum_{k=0}^{m-1} \theta_k \varepsilon_{n-k} \tag{2.8}$$

The AR part is used to model the re-correlation effect and the parameters θ_k are adjusted to fit the correlation structure. Since each frame in video data is correlated due to temporal correlation, the auto-correlation function has peaks at all lags. For this model, the parameter estimation and analytical solutions are difficult to obtain.

2.4.4 Autoregressive Integrated Moving Average Models

This model denoted by ARIMA(p,d,q) is an extention to the ARMA(p,q). It is obtained by constructing the polynomial $\phi(B)$ to have d roots equal to unity and the remaining roots lie outside the unit circle. The ARMA(p,d,q) can be represented as :

$$\varphi(B)\nabla^d X_t = \theta(B)\varepsilon_t \tag{2.9}$$

Here ∇ is a difference operator, defined as $(X_t - X_{t-1}) = \nabla X_t$, and $\varphi(B)$ is a polynomial in B. The ARIMA(p,d,q) can be used for capturing the non-stationary in the time series [8].

Till now, traditional stationary traffic models have been discussed, in which, it is observed that the auto-correlation function decays exponentially. Recently, it has been found in the traffic measurements of Ethernet LAN traffic that the auto-correlation function decays at a much slower rate than exponential. So, the processes that exhibits this type of slow decay is called long-range dependent processes.

3

Self-Similarity and Power Law Behaviour in Traffic Models

Recently several authors [11] have carried out extensive studies with a view to understand the realistic features of local area and wide area network traffic. These studies show that, network traffic exhibits variability at a wide range of scales [12]. This gives to a phenomenon of self-similarity is the existence of heavy-tailed distribution asymptotically behaving like a power law. We shall briefly review these aspects in this chapter.

The phenomenon of self-similarity is observed in traffic related studies in broadband networks and one can observe structural similarities across wide range of time scales. In the case of packet traffic, self-similarity is observed at every time scale ranging from a few milliseconds to minutes and hours by similar looking bursts [13]. These bursts will occur when the interarrival processes appear to form visual clusters, i.e. several small interarrival times followed by a relatively bigger one. The major cause of burstiness is the presence of strong positive auto-correlations [7].

For the purpose of illustration of this phenomenon, we reproduce here the graph in fig. 3.1 from [6]. This shows the same Ethernet traffic trace at various levels of aggregation. As observed by Morin and Nailson "the first plot shows 3 seconds of the original trace where each point plotted represents the number of bytes of traffic during the corresponding 10 msec interval. The second plot shows 30 seconds of the original trace when each point plotted represents the number of bytes of traffic during the corresponding 100 msec interval divided by 10. The third plot shows 300 sec of the trace where each point plotted represents the number of bytes of traffic during the corresponding 1 second interval divided by 100. Lastly, the fourth plot shows 3000 seconds of the trace where each point plotted represents the number of bytes of traffic during the corresponding 10 sec interval divided by 1000. Surprisingly, the plots *look similar*, even as the level of aggregation increases. This is in contrast to traditional traffic models, whose plot shows *smoothness* as the level of aggregation increases". This means, that, there exists natural length of bursts in Ethernet traffic. Crovella and Bestavros [14] noted that, the world wide web network traffic also exhibits the self-similar nature.

Based on these observations it has been found that, the high speed network traffic is bursty over wide range of time scales.

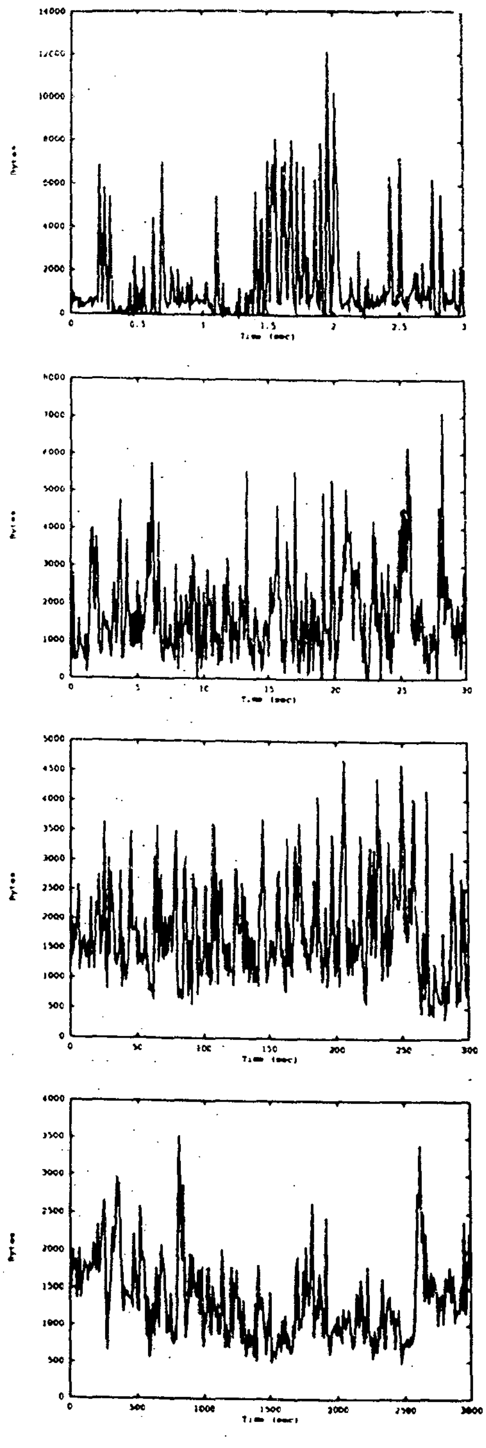

Figure 3.1 Ethernet traffic data at various levels of aggregation.

3.1 Definition of Self-Similarity

Let X(t) be a wide-sense stationary stochastic process with mean μ, variance σ^2, and auto-correlation $\rho(k)$, $k \geq 0$. The stochastic process {X(t)} is said to be "exactly self-similar", if the correlation structure,

$$\rho_k = E\left[\left(X(t)-\mu\right).\left(X(t+k)-\mu\right)\right]/\sigma^2 \tag{3.1}$$

is preserved across different time scales. The process {X(t)} is said to be "asymptotically self-similar", if the corresponding aggregated processes $X_k^{(m)}$ given by $X_k^{(m)} = 1/m\left(X_{km-m+1}+...+X_{km}\right)$, $k \geq 1$ are same as X. i.e.,

$$\rho_k^{(m)} \rightarrow \rho_k, \quad as \quad m \rightarrow \infty \tag{3.2}$$

Earlier in this chapter, it was said that the stochastic self-similar processes retain the same statistical characteristics over a range of time scales and they satisfy the following relation in distribution sense; viz.:

$$\left\{X(at)\right\} = a^H \left\{X(t)\right\} \tag{3.3}$$

where H is self-similar parameter [8]. For more details see [11].

3.2 Heavy-Tailed Distribution and Power Law Behaviour

A random variable X is said to follow a heavy-tailed distribution if,

$$P[X > x] \sim x^{-\alpha} \quad as \quad x \to \infty \tag{3.4}$$

Here α is a parameter such that 0 < α < 2. The tail probabilities depict a power law behaviour which means that there is a non-negligible probabilities for random variable X taking large values. This, in the context of bursty phenomenon in network traffic implies occurrence of transmission of large number of packets or cells with finite probability.

The well known heavy tailed distribution is *pareto* distribution with p.d.f. given as,

$$p(x) = \alpha k^{\alpha} x^{-\alpha-1}, \quad x \geq k \tag{3.5}$$

where α and k are positive parameters [12]. The c.d.f. of (3.5) behaves as (3.4).

It is interesting to note that heavy-tailed distribution can be obtained by resorting to maximum entropy principle.

3.3 Maximum Entropy-Rationale for Power Law

The principle of maximum entropy provides a rationale for choosing a probability distribution when partial information about a random variable is available. According to this principle we choose that probability distribution which besides being consistent with the given information is maximally unbiased in relation to what is not given. In other words the probability distribution chosen has the maximum entropy amongst all distributions consistent with the given information.

Mathematically the problem of maximum entropy distribution can be stated as follows (see for details in [15]): let X be a continuous random variable with p.d.f. f(x). The measure of uncertainty is given by its entropy,

$$H = -\int f(x).\ln f(x)dx \tag{3.6}$$

Further we are given the constraints,

$$\left.\begin{aligned} &\int_a^b f(x)dx = 1, \\ &\int_a^b g_r(x).f(x) = \overline{g_r}, \quad r = 1,2,..m. \end{aligned}\right\} \tag{3.7}$$

The principle of maximum entropy states that:

$$\left.\begin{aligned} &Max \quad H \\ &subject \quad to \quad constraints \quad (3.7) \end{aligned}\right\} \tag{3.8}$$

The maximum entropy probability distribution is then formed to have the p.d.f.,

$$f(x) = N.exp[-\lambda_1 g_1(x)-\lambda_2 g_2(x) \ldots -\lambda_m g_m(x)] \tag{3.9}$$

where N, λ_1, λ_2, .., λ_m are to be determined inorder to satisfy the constraints (3.7). It is interesting to note that when the first two moments of a random variable are prescribed, then M.E.P.D. turns out to be Gaussian distribution.

As discussed in Kapur [15], one obtains the generalized Cauchy distribution when, $E\{ln[1+x^2]\}$ is prescribed. Then the M.E.P.D. turns out to be

$$f(x) = \frac{N}{\left(1+x^2\right)^b}, \quad -\infty < x < \infty \tag{3.10}$$

The Cauchy density function

$$f(x) = \frac{1}{\pi}.\frac{1}{1+x^2}, \quad -\infty < x < \infty$$

being particular case , b=1.

It is worth noting that f(x) given in (3.10) provides power law behavior for large x. i.e.

$$f(x) = K.x^{-2b} \tag{3.11}$$

Similarly by restricting the random variable x to positive values, one can generate Zipf and Pareto like distributions by prescribing $E[ln\ x]$.

The M.E.P.D. turns out to be,

$$p(x) = \frac{N}{x^{\lambda_1}}, \quad x > 0$$

where N is the normalization constant. One can say that power law maximum diversity when the constraint specifies the average logarithm of the variate. As noted by West [16], this constraints implies a scaling relation for the variable of interest. This scaling can be called as self-similarity and is one of the defining properties of *fractal* processes [16,17]. The power law tails of the fractal process distribution indicated that rare events occur in clustered form.

4

Generalized model for traffic distribution with power law

In recent years, it has been established that typical traffic in Ethernet local area network and wide area network appears to be self-similar across all or at least over a wide range of time scales. The crucial feature of self-similar processes is that they exhibit long range dependence (LRD). with auto-correlation function decaying hyperbolically, which is much slower than exponential decay, and is non-summable [18]. This decay corresponds to power law behaviour and is in contrast to traditional models, which exhibit short range dependence (SRD), i.e. have an auto-correlation function that decays exponentially. The serious implication for ATM network design is that, conclusions based on traditional models may not be applicable under self-similar traffic. Recent studies on self-similar traffic have shown that the LRD structure may have a significant impact on queueing performance [13]. So, the earlier attempts to formulate the traffic distribution in terms of Markov Modulated Process does not appear to be realistic as it has auto-covarience which can be approximated by $C(\tau) = C.\exp[-a\tau]$ [8]. Recently, Simonian [19] has also formulated the traffic model in terms of Ornstein-Uhlenbeck process with exponentially decaying correlation structure.

Denoting by X(t), the instantaneous packet rate into transmission system at time t, the correlation function is found to be,

$$E[(X(t+\tau)-K).(X(t)-K)] = \sigma^2 .e^{-\mu|\tau|} \quad (4.1)$$

where K is the mean traffic. Simonian [19] has employed O.U.process model for describing the content of fluid queue.

4.1 Ornstein-Uhlenbeck Process

For the sake of completeness we are giving brief details of the O.U. process (see [20]). O.U. process is described as the stochastic differential equation,

$$dx = -\mu.X.dt + \sigma_0.Z(t).\sqrt{dt} \quad (4.2)$$

where X(t) is the speed of a packet at time t and W(t) is the Wiener process which can be expressed as equal to $Z(t)\sqrt{dt}$. Here, Z(t) is a purely a Gaussian random process with zero mean and unit variance. Conditioning on X(t) = x, we may write (4.2) as,

$$dX(t) = -\mu.x.dt + \sigma_0.dW(t) \quad (4.3)$$

Comparing (4.3) with the general diffusion process, we get X(t) as a diffusion process with infinitesimal mean and variance as $-\mu x$ and σ^2 respectively. Accordingly, the forward Kolmogorov equation is,

$$\frac{1}{2}\sigma_0^2.\frac{\partial^2 p}{\partial x^2} + \mu.\frac{\partial}{\partial x}(x,p) = \frac{\partial p}{\partial t} \quad (4.4)$$

where, $p \equiv p(x,t)$ is the p.d.f. of the process X(t) at time t. The initial value condition is given by,

$$p(x,0) = \lim_{t \to 0} \quad \delta(x - x_0)$$

Defining m.g.f. of X(t) as

$$\phi(\theta,t) = \int_{\infty}^{-\infty} e^{-x\theta} p(x,t).dx$$

eq. (4.4) reduce to

$$\frac{1}{2}\sigma^2\theta^2\phi - \mu\theta\frac{\partial\phi}{\partial\theta} = \frac{\partial\phi}{\partial t} \tag{4.5}$$

Expressing ϕ in terms of cumulant generating function $\psi(\theta,t) = \log \phi(\theta,t)$, we obtain from (4.5)

$$\phi(\theta,t) = -(x_0.e^{-\mu t})\theta + \frac{\sigma_0^2(1-e^{-2\mu t})}{2\mu}(\frac{1}{2}\theta^2) \tag{4.6}$$

from which we get,

$$p(x|x_0,t) = \frac{1}{\sqrt{2\pi}.\sigma_1(t)}\exp\left[-\{x-\mu_1(t)\}^2 \Big/ \sigma_1^2(t)\right] \tag{4.7}$$

where,

$$\sigma_1^2(t) = \frac{\sigma_0^2(1-e^{-2\mu t})}{2\mu} \quad \text{and} \quad \mu_1(t) = x_0.e^{-\mu t}$$

It may be noted that in the limit of large times, the density function in (4.7) reduces to a Gaussian distribution with $\sigma_1^2(\infty) = \sigma_0^2/2\mu$ and $\mu_1(\infty) = 0$. Further it can be proved from equation (4.3) that the auto-correlation decays exponentially. It needs to be emphasised that, the exponentially decaying correlation does not represent the realistic traffic models for ATM networks. However, in reality, the traffic models should depict the power law behaviour.

4.2 Generalized Model

It is proposed to generalize the Simonian [19] model, to capture the realistic aspects of multiple sources like audio, video, voice and multimedia services. As mentioned in the first chapter, that any type of information will be transmitted in the form of packets.

Simonian [19] treated the input process {X(t)} being described by a stationary O.U. process. The corresponding stochastic differential equation is

$$dX(t) = \eta.[K-X(t)].dt + \sigma_0.dW(t) \tag{4.8}$$

where K, η, σ_0 are positive parameters and dW(t) is a standard Wiener increment process. It is easy to show that X(t) has a stationary Gaussian distribution with steady-state mean K and variance $\Sigma^2 = \sigma^2 / 2\eta$. writing K-X(t) = Y(t), equation (4.8) becomes,

$$dY(t) = -\eta.Y(t) + \sigma_2.dW(t) \tag{4.9}$$

It can be argued that, in reality, the rate η cannot be treated as a constant parameter. However η should vary in stochastic manner as a result of packets originating from voice, video, and data services. It may be convenient to explain these services in the present context.

Data : Basically, there are two types of applications such as low speed (keyboard input) and high speed (file transfer, image transfer for CAD etc.) applications. The delay will be more in low speed applications as compared to high speed applications. So, the correlation structures for both applications will be different.

Voice : Since voice is a constant bit rate (CBR)service, packets can be sent for transmission only when it is completely filled. This results in the introduction of packetization delay. Traffic bursts will be generated only during the active (ON) periods of voice.

Video : In video communications, variable bit rate compression algorithms transmit at a higher rate during high-activity (motion) scenes and at a low rate when there is less motion. Different types of video traffic results in fast decaying short term correlation and slow decaying long term correlation. Based on experimental work Sen *etal.*[10] have indicated the following correlation structure in the context of other types of video traffic, such as broadcast television, videoconferencing, and longer videotelephone sequences (showing persons talking and listening). Considering an environment where the video sources feeding the network are a mix of these types, then two important correlations exist, viz. : a relatively fast-decaying short-term correlation corresponding to uniform activity levels, with a time constant of the order of a few hundred milliseconds, and a slow decaying long-term correlation corresponding to sudden changes in the gross activity level of the scene (e.g., scene changes in broadcast TV or changes between listener and talker modes in a video telephone conversation), with a time constant on the order of a few seconds (see [10]).

Thus the parameter η, conditioned by the random traffic resulting from different sources, behave as stochastic process. So, η(t) consists of two components; viz.:

$$\eta(t) = k + \Lambda(t), \tag{4.10}$$

where k denotes the average component and Λ(t) denotes the statistical fluctuations around the mean rate k. The fluctuations around the mean rate is assumed to be characterized by Gaussian white noise process. The assumption of Λ(t) being Gaussian is justified on the grounds that the fluctuations arise from a large number of random factors, and, hence, by central limit theorem, the total effect obeys the Gaussian law [21]. We further assume that the time scale of fluctuations in Λ(t) is much smaller in comparison to the macroscopic time scale, which also can be assumed to be represented by a delta correlated process. Thus Λ(t) is modelled as a Gaussian white noise process. i.e.

$$E[\Lambda(t_1).\Lambda(t_2)] = \sigma_3^2.\delta(t_1-t_2) \tag{4.11}$$

By substituting (4.10) in (4.9), we get the stochastic differential equation (SDE), characterizing the traffic becomes,

$$dY(t) = -k.Y(t).dt - \Lambda(t).Y(t).dt + \sigma_2 dW(t) \tag{4.12}$$

This SDE can be recast in the form :

$$dY = f(Y).dt + \sigma.g(Y).dW \tag{4.13}$$

where $f(Y) = -k.Y(t)$ and $g(Y) = (\sigma_3.Y + \sigma_2)$. This SDE can be interpreted in two ways, viz.: Stratonovich sense or Ito sense and these two prescriptions are formally equivalent, i.e. one can go from the results obtained under one prescription to those under the other via a transformation formula. However, the Stratonovich prescription is preferable for modelling the traffic [22]. The Fokker-Planck equation corresponding to eq. (4.13),

$$\frac{\partial p}{\partial t} = \frac{-\partial}{\partial y}[\{f(y) + \frac{1}{2}.\sigma^2.g(y).g'(y)\}p] + \frac{\sigma^2}{2}.\frac{\partial^2}{\partial y^2}[g^2(y).p] \tag{4.14}$$

Here $p \equiv p(y,t)$ denotes the probability density function. This equation has to be solved with initial condition ,

$$p(y,0) = \lim_{t\to 0} \quad \delta(y - y_0) \tag{4.15}$$

and appropriate boundary conditions. Using (4.13) and (4.14), the Fokker-Planck equation by computing eq. (4.12) becomes,

$$\frac{\partial p}{\partial t}(y,t) = \frac{-\partial}{\partial y}[(-ky + \frac{\sigma_3^{\ 2}}{2}y)p] + \frac{\partial^2}{\partial y^2}[(\frac{\sigma_2^{\ 2}}{2} + \frac{\sigma_3^{\ 2}}{2}y^2)p] \tag{4.16}$$

$$\frac{2}{\sigma_2^{\ 2}}.\frac{\partial p}{\partial t}(y,t) = \frac{2}{\sigma_2^{\ 2}}.(k - \frac{\sigma_3^{\ 2}}{2}).\frac{\partial}{\partial y}(yp) + \frac{\partial^2}{\partial y^2}[(1 + \frac{\sigma_3^{\ 2}}{\sigma_2^{\ 2}}y^2)p] \tag{4.17}$$

Rescaling the variable y and introducing, $(\sigma_3/\sigma_2)y = z$,

we find,

$$\frac{\partial}{\partial y} = \frac{\sigma_3}{\sigma_2}.\frac{\partial}{\partial z}; \quad \frac{\partial^2}{\partial y^2} = \left(\frac{\sigma_3}{\sigma_2}\right)^2.\frac{\partial^2}{\partial z^2}.$$

Noticing $\quad p(y,t).dy = p(z,t).dz = p(z,t).\frac{\sigma_3}{\sigma_2}$,

equation (4.17) yields,

$$\frac{2}{\sigma_3^{\,2}}.\frac{\partial p}{\partial t}(z,t) = (\frac{2k}{\sigma_3^{\,2}} - 1).\frac{\partial}{\partial z}(zp) + \frac{\partial^2}{\partial z^2}[(1+z^2)p] \qquad (4.18)$$

Introducing a new time scale τ,

$$\tau = t.\frac{\sigma_3^{\,2}}{2} \text{ and writing } \alpha = \frac{k}{\sigma_3^{\,2}},$$

then (4.18) reduces to the standard form as given in Wong [23]. We get

$$\frac{\partial^2}{\partial z^2}\left[\left(1+z^2\right)\right] + \left(2\alpha - 1\right)\frac{\partial}{\partial z}\left(zp\right) = \frac{\partial p}{\partial \tau} \qquad (4.19)$$

This parabolic equation has to be solved with reflecting barrier at the origin and the initial condition (4.15).

4.2 Solution of Fokker-Planck equation

Following Wong [23] solution $p(z|z_0, \tau)$ of above Fokker-Planck equation is,

$$p(z|z_0,\tau) = (1+z^2)^{-(\alpha+\frac{1}{2})}\left\{\frac{1}{\pi}\sum_{n=0}^{\infty}\frac{(\alpha-n)}{n!\Gamma(2\alpha+1-n)}.e^{-n(2\alpha-n)\tau}.\theta_n(z_0).\theta_n(z)\right.$$

$$\left. + \frac{1}{2\pi}\int_0^{\infty} e^{-(\alpha^2+\mu^2)\tau}\left[\Psi(\mu,z_0)\Psi(-\mu,z) + \Psi(-\mu,z_0)\Psi(\mu,z)\right]d\mu\right\}$$

where,

$$\theta_n(z) = 2^{\alpha-n}\Gamma(\alpha - n + \frac{1}{2})(-1)^n(1+z^2)^{\alpha+\frac{1}{2}}.\frac{\partial^n}{\partial z^n}\left[(1+z^2)^{n-\alpha-\frac{1}{2}}\right]$$

are polynomials of degree n, and Ψ(μ,z) is given by,

$$\Psi(\mu,z) = (z+\sqrt{1+z^2})^{i\mu}(1+z^2)^{\frac{1}{2}}\,{}_2F_1(-\alpha,\alpha+1;1+i\mu;\frac{1}{2}+\frac{1}{2}\frac{z}{\sqrt{1+z^2}}).$$

Here ${}_2F_1$ is the Gaussian hypergeometric series.

It is interesting to note that a complete time dependent solution for the stochastic variate z(τ) has been obtained. This in turn will provide stochastic description of traffic input rate, $X(t) = \left(\Lambda - \frac{\sigma_2}{\sigma_3}z\right)$. The transient solution can also be utilized in obtaining expressions for equivalent capacity for broadband networks [24]. It can be easily seen that the stationary solution turns out be,

$$P^*(z) = \lim_{\tau\to\infty} p(z,\tau) = (1+z^2)^{-(\alpha+\frac{1}{2})}$$

which for large z behaves as

$$p(z) \approx z^{-(2\alpha+1)} \qquad as\ z \to \infty$$

This gives rise to power law behaviour which is a desirable feature for realistic network traffic.

5

Self-Similar Traffic model

In broadband networks, the aggregate traffic can be visualized as comprising two parts, viz: constant bit rate (CBR) and variable bit rate (VBR). Generally, CBR can be described by average time dependent function, whereas VBR is bursty in character. This VBR generally depicts the *fractal* like character. This aspect of VBR can possibly be described by discrete or continuous random walk. To this end we can adopt the analysis discussed by Bendler and Shlesinger [25] for our purpose. This means that the waiting time probability distribution, governing the time interval between jumps has infinite mean.

5.1 Montroll-Weiss Random Walk Type Traffic Model

We denote by $P_n(x)$ the probability for packet executing a random walk such that at the n^{th} time epoch the number of packets is x. From this, we can recursively obtain $P_{n+1}(x)$ as follows:

$$P_{n+1}(x+1) = \sum_{x} p(x')P_n(x-x') \tag{5.1}$$

i.e. upto n^{th} step or time epoch we have (x-x') packets and then the remaining packets x' arrive during $(n+1)^{th}$ step.

In an interesting paper Montroll and Weiss (1965) established a connection between P(x,t) and $P_n(x,t)$ in terms of generating function for $P_n(x)$ defined as,

$$G(x,z) = \sum_{n=0}^{\infty} P_n(x) z^n \tag{5.2}$$

We find from eq. (5.1)

$$G(x,z) - z\sum_{x'} p(x')G(x-x') = \delta_{x,0} \tag{5.3}$$

In a similar way one can write down the continuous version of eq. (5.1) as,

$$P(x,t) = \int_0^t Q(x,t-\tau)R(\tau)d\tau \tag{5.4}$$

where P(x,t) is the probability of having x packets at time t. Denoting by Q(x,t) the p.d.f. to receive x packets at time t and R(τ) denotes the probability that several packets arrive in a time interval τ after the previous packets have been received.

In terms of probability ψ(t)dt that a jump occurs in (t,t+dt), we can write

$$R(t) = 1 - \int_0^t \psi(\tau)d\tau \tag{5.5}$$

and

$$Q(x,t) = \sum_{x'}\int_0^t \psi(\tau).p(x').Q(x-x',t-\tau).d\tau + \delta_{t,0}.\delta(t) \tag{5.6}$$

Defining Laplace transform of f(t) as

$$\bar{f}(s) = L\{f(t)\} = \int_0^\infty e^{-st}.f(t).dt,$$

we find,

$$\bar{R}(s) = [1 - \psi(s)]/s \tag{5.7}$$

$$P(x,t) = L^{-1}\left[G(x,\psi(s)).\frac{1-\psi(s)}{s}\right] \tag{5.8}$$

where G represents the generating function as in eq. (5.2). For obtaining behaviour of P(x,t), we have to specify ψ(t).

Noting that,

$$\psi(s) = \int_0^\infty e^{-st}.\psi(t).dt \approx 1 - s\bar{t} + \cdots \quad as \quad s \to 0, \tag{5.9}$$

we can obtain the behaviour of P(x,t).

In contrast to first moment being finite moment, we consider the situation described by,

$$\psi(t) \approx At^{-1-\alpha} + \cdots \quad as \quad t \to \infty, \quad 0 < \alpha < 1, \tag{5.10}$$

such that the first moment is infinite as can be seen from,

$$\psi(s) \approx 1 - \frac{A}{\alpha}\Gamma(1-\alpha)s^{\alpha}$$

As observed by Bendler and Shlesinger [25] that, "a stochastic process with a $\psi(t)$ with $\bar{t}$ infinite has been called a fractal time process, because it can be shown that, since the jumps do not occur on the average, at regular intervals they occur in self-similar bursts. These bursts have the structure of a Cantor set of fractal dimension α" [26]. Thus we find the traffic generated by random walk for $\psi(t)$ given by eq. (5.10) is of fractal nature. This traffic when superposed on the traffic generated by CBR may provide a more realistic model for the network traffic.

This modeling process needs to be investigated further with a view to making comparisons with the empirically observed data.

Concluding Remarks

After briefly reviewing earlier traffic models, we have proposed a new model for characterizing traffic in communication networks. This model has been formulated in terms of stochastic differential equation. The advantage of this formulation is that it is amenable to analytical treatment. The complete time-dependent solution of the probability density function for the traffic has been obtained. A significant aspect of this model is that it depicts power law behaviour. This feature has also been found to be present in local as well as wide area networks. This power law behaviour is gaining a lot of importance in communication network studies as its influences on performance measures is enormous. Simulation studies show that packet delay, cell loss rate increases substantially. It is proposed to investigate in future the effect of traffic model which we have proposed in the performance measures of broadband networks.

We have also formulated a model for fractal like traffic behaviour in terms of continuous random walk with transition probability having infinite mean. This aspect appears to be quite promising.

References

[1] MARTIN de PRYCKER, "*ASYNCHRONOUS TRANSFER MODE Solution for Broadband ISDN*",2nd edition,Ellis Horwood,1993.

[2] Jain N., "*Transient Analysis of Index of Dispersion in Arrival Processes for ATM Multiplexer*", a project report,JNU,1997.

[3] Chen T.M and Liu S.S., "*ATM Switching Systems*", The Artech House Tele-Communications Library, 1995.

[4] Tanebaum A.S., "*Computer Networks*", Printice Hall of India, 3rd Edition, 1997.

[5] Wang P. *et al.*, "*ATM*", Communications of the ACM, vol. 38, No. 2, February 1995, pp. 31-42.

[6] Morin P.R. and Nailson J., "*The Impact of Self-Similarity on Network Performance Analysis*", Technical report, Computer Sci. 95.495, Carlton University, 1995.

[7] Frost V.S. and Melamed B., "*Traffic Modeling for Tele-communications Networks*", IEEE Commun. Mag., Mar. 1994, pp. 70-81.

[8] Adas A., "*Traffic Models in Broadband Networks*", IEEE Commun. Mag., July 1997, pp. 82-89.

[9] Kleinrock L., "*Queueing Systems*", vol. 1, John Wiley and Sons, 1975.

[10] Sen P., Maglaris B., Rikli N. and Anastassiou D., "*Models for Packet Switching of Variable-Bit-Rate Video Sources*", IEEE JSAC, vol. 7, No. 5, June 1989.

[11] Leland W.E., Taqqu M.S., Willinger W. and Daniel V., "*On the Self-Similar Nature of Ethernet Traffic (Extended Version)*", IEEE/ACM Trans. Networking, vol. 2, No. 1, Feb. 1994.

[12] Park K., Kim G. and Crovella M., "*On the relationship between file sizes, transport protocols, and self-similar network traffic*", Technical report 96-016, Boston University, Computer Science Department, 1996.

[13] Willinger W., Taqqu M.S., Sherman R. and Daniel V., "*Self-Similarity Through High-Variability : Statistical Analysis of Ethernet LAN Traffic at the Source Level*", SIGCOMM 1995, pp. 100-113.

[14] Crovella M.E. and Bestavros A., "*Self-Similarity in World Wide Web Traffic Evidence and Possible Causes*", SIGMETRICS 1996, pp. 160-169.

[15] Kapur J.N., "*Maximum Entropy Model in Science & Engineering*", John Wiley, New Delhi, 1989.

[16] West B. J. and Salk J., "*Complexity, organization and uncertainty*", European J. Operational Research, vol. 30, 1987, pp. 117-128.

[17] Mandelbrot, B.B., "*The Fractal Geometry of Nature*", Freeman, San Francisco, CA, 1982.

[18] Huang C., Devetsikiotis M., Lambadaris I. and Kaye A.R., "*Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video : A Unified Approach*", SIGCOMM 1995, pp. 114-125.

[19] Simonian A., "*Stationary Analysis of Fluid Queue with Input Rate Varying as an Ornstein-Uhlenbeck Process*", SIAM J. Appl. Math., vol.51, No. 3, June 1991.

[20] Cox D.R. and Miller H.D., "*The Theory of Stochastic Processes*", Methuen & Co. Ltd., London, 1965.

[21] Karmeshu and Jaiswal N.K., "*Heavy traffic queue with stochastic arrival and service rates*", Int. J. Systems Sci., vol. 12, No 5, 1981, pp. 101-110.

[22] Wong E. and Zakai M., "*On the Relation Between Ordinary and Stachastic Equations*", Int. J. Engg. Sci., vol. 3, No. 2, 1965, pp. 213-229.

[23] Wong E., "*The Construction of a Class of Stationary Markoff Processes*", Proc. Simposia in Appl. Math., American Mathematical Society, vol. XVI, 1964.

[24] Guerin R., Ahmadi H. and Naghshineh M., "*Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks*", IEEE JSAC, vol. 9, No. 7, Sep. 1991.

[25] Bendler J.T. and Shlesinger M.F., "*Dielectric Relaxation via the Montroll-Weiss Random Walk of Defects*", The Wonderful World of Stochastics A Tribute to Elliot W.Montroll, North-Holland, vol. XII, 1985, pp. 33-45.

[26] Shlesinger M.F. and Hughes B.D., *Physica A* 109 (1981) 597.