UNIVERSITÉ DU LUXEMBOURG

PhD-FSTC-2015-18
The Faculty of Sciences, Technology and Communication

DISSERTATION

Defense held on 30/04/2015 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN INFORMATIQUE

by

Yu LI
Born on 22 June 1979 in Guangzhou, (China)

MATHEMATICAL MODELING OF NETWORK TRAFFIC

Dissertation defense committee

Dr. Ulrich Sorger, dissertation supervisor
Professor, Université du Luxembourg

Dr. Fernando Gomez-Cubillo
Professor, University of Valladolid, Spain

Dr. Raymond Bisdorff, Chairman
Professor, Université du Luxembourg

Dr. Zbigniew Michna
Professor, Wroclaw University of Economics, Poland

Dr. Zdzislaw Suchanecki, Vice Chairman
Université du Luxembourg

TABLE OF CONTENTS

DEDICATION

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents. Thank you, Mom and Dad, for the beautiful example of love that you have given me, even though you are not with me in Luxembourg. Thank you also for instilling in me a great work ethic that was necessary to succeed in doctoral school. Without you, I could not have succeeded in graduate school. Thank you for your support, your understanding and your patience. Thank you for believing that I could do this, even when I wasn't certain. Most of all, thank you for your unconditional love. I also dedicate this dissertation to my cousin and his wife, Lai Li and Weiti Guo, for raising such a caring daughter and sharing her with me. To my aunt, for your support and encouragement throughout the process.

I dedicate this work and give special thanks to Catherine Guerder, my good friend and my favorite therapist, for your kindness during my hospitalization and help during two years' rehabilitation. Thanks to you, my French has been greatly improved as you don't speak English. I will never forget what you have done for me.

ACKNOWLEDGMENTS

I owe my advisors, Zdzislaw Suchanecki and Ulrich Sorger, more than words can express, for all that you have done for me over the last four years. I thank you for your patience, for sharing with me their enthusiasm for mathematics and computer science, and for helping me grow immensely as a mathematician and computer scientist. Your enthusiasm is contagious even when mathematics is frustrating, and a conversation with you always gives me the motivation I need to continue. Thank you for the advice and support you have provided for me, both academically and personally. A special thank you to Raymond Bisdorff for your helpful remarks. I would also like to thank my committee: Fernando Gomez-Cubillo and Zbigniew Michna. I am especially grateful to Massimo Chenal and Vincenzo Iovino for their comments which helped improve my thesis. Thank you to Christian Frank for his tutelage as I began to study LyX for scientific writing.

My experience as a graduate student at University of Luxembourg has been amazing. Not only do I have an exceptional advisor, but the entire SnT Department has been supportive and encouraging. The faculty and staff truly care about graduate students and their academic success as well as their emotional stability. Thank you each for your support and encouragement. The department also provides many opportunities for the development of graduate students as teachers, as academics, and as future faculty members. Thank you to the faculty which provides these opportunities for graduate students. I also owe many thanks to my fellow graduate students for their friendship and support during my time at University of Luxembourg. For the last several years, I have had the chance of being an office neighbor of Massimo Chenal, Miguel Urquidi, Afonso Delerue, Marjan Skrobot, Masoud Tabatabaei, and other SnT member. Courtney, it has been wonderful to grow scientifically with someone who is also a close friend. I also owe Xihui Chen a very specific thank you related to this dissertation: your comments were trenchant and perceptive. Foued Melakessou, thank you for all of your advice.

Finally, to each and every person mentioned herein I would like to once more sincerely say thank you.

LIST OF FIGURES

E_i	Erlang distribution
F_X, G_X	Distribution function of random variable X
f_X, g_X	Density function or probability function of random variable X
$\overline{F}$	Tail of a given distribution F, i.e. $\overline{F} = 1 - F$
F^{*n}	n-fold convolution of distribution F
H	Homotopy
M	Markov operator or matrix
$\mathbb{N}$	Set of natural numbers
$N(t)$	Counting process
$P[\cdot]$	Probability measure
$P(\cdot)$	Polynomial
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of non negative real numbers
X, Y, Z	Random variables
X_t	Time series or random process
φ_X	Characteristic function of random variable X
M_X	Moment generating function
m_X^k	k-moment of random variable X

$p(x)$ power-law function

Abstract

Increasing access to the Internet is producing profound influence around the World. More and more people are taking advantage of the Internet to obtain information, communicate with each other far away and enjoy various recreations. This largely increased demand for the Internet requires better and more effective models. During the 1990s, a number of studies show that due to a different nature from telephonic traffic, in particular a bursty nature, traditional queuing models are not applicable in modeling of modern traffic. This work presents some alternative rigorous models that can be used in studying the behavior of the Internet traffic.

In the thesis, we propose several new models to explain bursty nature of network traffic.

Many practical problems in computer science, natural science and insurance are highly complex, since they typically involve a huge amount of random factors. Conventional queuing theory cannot handle such problems and the high variance and long memory features are not yet well understood.

In the first part of our research, we start with investigation of random sum as well as random product and derive a single path model of network traffic and a multiple path model of network traffic.

In the second part of this thesis, we present a homotopic approach to model network traffic.

Preface

The work that follows presents the results of my research in the last four years at the University of Luxembourg.

In the first two years of my Phd studies, I delved deeper in the convolution problems in the field of probability theory without apparent relation to network traffic modeling. In the studies I introduced a convolution invariant parameter and derived some properties of random sum.

A fortunate involvement with the research group led by professor U. Sorger provided my first exposure to network traffic modeling. The classical modeling fails in modeling network traffic because of the bursty nature of network traffic. To my surprise, in the light of a multiplicative law derived in the previous work of M. Foued, U. Sorger and Z. Suchanecki, the convolution invariant parameter hints to a novel way to elaborate better models of Internet traffic and provides an effective way for analyzing heavy tail phenomena.

Finally, I built a bridge between the convolution invariant parameter and network traffic modeling. What's more, this approach can not only be used in analysis of data transfer but it can also be used in some other fields, such as financial modeling.

Chapter 1

Introduction

1.1 Background

Internet has nowadays a profound impact on the lives of everyone and provides various services related to various different fields. Internet provides information in an easy and quick manner that is updated by second, which plays an important role in everything from social networking to worldwide stocks and currencies. Secondly, it makes people become more and more connected to each other, even to the ones who are in another city or even halfway across the world. Meanwhile, it changes the way of business and entertainment, for instance, it allows people to have meetings per Skype, watch television programs on demand and play games on line.

In the evolution of Internet there are at least four trends for the last two decades [1].

-Greatly increasing Internet users

The number of internet users has been increasing exponentially and reached three billion by the end of 2014, according to the United Nations' telecommunications agency.

-More and more widely used wireless and mobile communication

The International Telecommunication Union disclosed in 2014 that the number of mobile-cellular subscriptions would grow to almost seven billion within the year of 2014 and wireless and mobile communication finds an increasingly wide utilization in all fields.

-Continuously evolving devices

New Web-friendly devices have been emerging in an endless flow, such as smartphone, e-tablet, Netbook, eReader, etc. Each new device contributes further network traffic.

-Multiple services by using a common network infrastructure

Cloud computing has been growing in popularity and the basic concepts of cloud computing are common services and common infrastructure.

All these trends lead to a huge amount of internet traffic.

Circuit Switching and Packet Switching Networks

Circuit switching is a conventional technology of implementing a telecommunications network, in which two points establish a dedicated line (circuit) through the network before the points may communicate. Early telephone exchanges are a classic example of circuit switching. In circuit switching, a physical path from source to destination is reserved for users as long as being used. The circuit functions as if the points were physically integrated as an electrical circuit. Circuit switching can be relatively inefficient because capacity is guaranteed on connections which are set up not in continuous use, but rather momentarily. However, the connection is immediately available while established [2].

Packet switching is by contrast a technology that has long been used to send data from one computer to another. It groups all transmitted data into individual and suitably sized packets. When traversing network adapters, switches, routers, and other points, packets are buffered and queued, resulting in variable delay and throughput depending on the network's capacity and the traffic load on the network [2].

In circuit switching network, the Erlang model is proved to be effective. However, applying the traditional telephonic traffic model to packet switching network, such as the Internet, fails due to its different nature [3, 4].

> "Normal science possesses a built-in mechanism that ensures a relaxation of restriction that bound research whenever the paradigm from which they derive ceases to function effectively. At that point scientist begin to behave differently and the nature of their research problems changes."
>
> Thomas Samuel Kuhn

Modeling of packets switched network is not a trivial task. In order to elaborator Internet model in a more accurate way, we must consider some certain network constraints, such as network topology, device, algorithms, stochastic properties of data packets,etc. [3].

Packet switched networks have been studied for over 40 years [1]. During this time many models have already been proposed. Early attempts focus on Markovian models, such as the Markov-Modulated Poisson Process [5, 6]. In recent analyses of traffic measurements, non-Markovian effects have been observed in CCS7 signaling traffic [7] and variable bit-rate video [8–11], Ethernet LAN traffic [12–14] ATM cell traffic [15], MAN traffic [16] and WAN traffic [17]. In the work [12] self-similarity is claimed to be one of the most characteristic features of both Ethernet traffic [12] and the World Wide Web [14].

Recent measurements of traditional telephony traffic have shown signs of heavy-tailed delay times [18, 19] and many non-standard characteristics, such as heavy tail property [20] and long range dependence [21], have been carefully studied. In [4, 22, 23] a log-normal model is proposed to study the observed "self similarity" and "bursty" features. Recently it is demonstrated [24, 25] that LogPH model achieves the greater accuracy than the other heavy tail models.

1.2 Problem Statement and Methodology

In the last 20 years computer networks have been intensively studied. The objective was to understand the nature of the Internet, describe the behavior of network traffic and provide a comparatively accurate estimation of traffic variables. In circuit switched networks, Erlang model is proved to be a powerful tool. However, conventional queuing models are not suitable for modeling packets switching network because of their high complexity and different nature, meaning that the packets arrival process is no longer poissonian, the inter delay times are not exponentially distributed. Under this circumstance, new mathematical tools are required.

In the recent studies of network measurements, bursty nature (high variance) has been observed, and the self-similarity model has therefore been proposed. However, the basic mechanism behind this model is unclear. Moreover, it provides a very limited information about probabilistic characteristics of the considered processes and consequently, they are unconvincing and unreliable [4].

In this thesis we present novel ways to model network traffic and explain the mathematical laws underlying the phenomenon of heavy tail.

This thesis is divided into two parts.

We start in the first part with the study of a powerful mathematical tool, the random sum, and introduce a convolution invariant parameter which plays a central role in our modeling. By applying this convolution invariant parameter, we derive a single path model and a multiple path model and illustrate the bursty nature of internet traffic.

The second part focuses on positive correlation structure of network traffic and presents a homotopic approach to model network traffic.

Random Sum Models of Network Traffic

> "Sharpening your axe will not delay your job of cutting wood."
>
> "Good tools are prerequisite to the successful execution of a job"
>
> Old Chinese proverbs

In order to analyze Internet traffic, we should devote some time to studying in depth a powerful mathematical tool: random sum.

The random sum originated from the famous Cramï¿œr–Lundberg model in insurance Mathematics:

An insurance company experiences two opposing cash flows: incoming cash premiums and outgoing claims. So for the insurer, the risk process is of the form

$$U(t) = u + ct - S(t), \quad S(t) = \sum_{i=1}^{N(t)} X_i, t \geq 0 \tag{1.1}$$

where u stands for the initial capital, c for the incoming premium rate and the total claim amount $S(t)$ consists of a random sum of independent and identically distributed claims X_i. Here $N(t)$ stands for the number of claims until time t. In (1.1) both $N(t)$ and X_i are random variables, so this term is called random sum.

Next we will explain the reasons why we choose this random sum to model network traffic.

In view of the high complexity of Internet, many characteristics in the Internet analysis are highly random and unpredictable. On the other hand, many measurements in internet, such as packets count, packets sizes, or arrival time, are accumulative. Therefore, many practical problems can be formulated in the form of a random summation of random variables. Hence, random sum is a useful tool for understanding the nature of the Internet and provides a reasonable description of its behavior.

From the statistics point of view, network traffic exhibits often high variance, therefore, packets arrival process is not poissonian , inter delay times are not exponential and standard models are not applicable. Under such circumstances, random sum as well as its variant, random product, provides a natural and accurate way of network traffic modeling.

A homotopic approach of modeling network traffic

Internet traffic exhibits a strong positive correlation [26]. However, most models do not describe the positive correlation structure of the process, and they are therefore not applicable for modeling network traffic with high correlation structure. In this thesis we present a homotopic approach to handle this time-scale-dependent correlation structure.

Homotopy is a topological concept which describes a continuous deformation between two objects. In mathematical language, a **homotopy** is a continuous function $H : [0,1] \times \mathbb{R} \to \mathbb{C}$ of two given continuous functions $f_1, f_2 : \mathbb{R} \to \mathbb{C}$ such that

$$\begin{aligned} H(0,t) &= f_1(t) \\ H(1,t) &= f_2(t). \end{aligned}$$

Consider the sum $Z = X + Y$ of two identically distributed random variables with common distribution F. If X and Y are independent, then the distribution of Z

$$F_Z(x) = F^{*2}(x).$$

Here $*$ denotes convolution of distributions.

If X and Y are strongly related, which implies $Z = 2X$, then

$$F_Z(x) = F(\frac{x}{2}).$$

We assume that there is a continuous deformation linking the distributions $F^{*2}(x)$

and $F(\frac{x}{2})$. As distribution is too restrictive, we turn to characteristic function. In mathematics a valid characteristic function of a random variable is positive definite, and in practice only first and second statistics are considered, so our task boils down to construction of a continuous function linking dependence and independence (homotopy) without losing positive definiteness.

1.3 Outline

This dissertation is organized as follows.

In chapter 2 we present some basic knowledge of networks traffic modeling.

In chapter 3 we study some mathematical tools, such random sum and convolution invariant parameter, in details and reveal a relation between convolution invariant parameter and random sum.

Random sum and convolution are strongly related. Here we reveal an important property of random sum and failure rate, that is, failure rate limit does not vary under convolution, and consequently under random sum. Based on this property, failure rate limit can be considered as a parameter invariant under convolution and random sum.

In chapter 4 we present a single-path random sum model of inter delay time and throughput traffic.

We start with inter delay times. The distribution of inter delay times can be expressed by a power-law function (for example, convex combination of Erlang distributions or Gamma distributions). Second, we derive a duality relation between inter delay times and throughput traffic, and then, by applying this relation, we study the probabilistic characteristics of throughput traffic.

In chapter 6 we present a homotopic approach to model network traffic.

We start with investigation of dependence of random variables and demonstrate that it is reasonable to characterize dependence by applying a homotopic method. We present two approaches to construct homotopy with positive definiteness preserved.

Chapter 2

Fundamentals

2.1 Introduction of Network Traffic Modeling

In general, the objective of network traffic modeling is to provide simple but accurate methods to analyze network for a variety of purposes, such as network design, network management, evaluation of novel services and improvement of protocols. In mathematical terms, the aim of traffic modeling is to find a time series measured at a single location to describe the behavior of network and study the traffic volume. However, due to the extremely dynamic behavior of modern traffic and huge amount of data flows, traditional analysis methods, such as collecting, storing and using the immense amount of information on all end-to-end traffic streams between all sources and destinations, do not apply. A wide range of problems -such as traffic engineering, traffic matrix estimation, anomaly detection, attack detection and capacity planning-demand a more sophisticated analysis. Moreover, traffic information is increasing continuously with exponential growth of internet usage. In this situation, for the efficient utilization of network resources, network engineers must focus on a number of manageable traffic parameters without paying attention to more traffic information than is necessary [1, 27].

Network modeling involves three steps: model selection, parameter estimation and statistical test.

Model Selection

Model selection determines the basic principle that explains the observations. A good stochastic model of network traffic is an effective tool for estimating probability distributions of traffic variables, it balances simplicity and accuracy. A more complex model fits better the data but additional parameters might not provide useful information. In [1] Timothy Neame pointed out that a useful model of network traffic should satisfy the following criteria:

1. It is defined by a small number of parameters.
2. If these parameters are fitted using measurable statistics of an actual traffic stream the following will be achieved:
   (a) the first and second order statistics including the auto covariance function of the stochastic process (the model) will match those of the actual traffic stream, and
   (b) if fed through a single server queue (SSQ), performance results for the model will accurately predict those of the real traffic stream fed into an identical SSQ. This must be true for a wide range of buffer sizes as well as for a wide range of service rates.
3. It is amenable to analysis.

If the process is suited to the character of the traffic that is being modeled, this will give maximum confidence in its usefulness.

Parameter Estimation

Parameter estimation is critical in accurately describing behavior of network traffic through stochastic models. Parameters are estimated based on a set of statistics that are measured or calculated from observed data. The most commonly used statistics are the mean and variance.

Statistical Test

A statistical hypothesis test provides a method of statistical inference using data from observations to predict the behavior of network traffic.

In brief, a good model should be simple but captures the most relevant statistics.

2.2 Measurement of Network Traffic

The measuring and analyzing of real network traffic provide us with a very important knowledge about computer network states. In analyzing process, statistical mathematical tools are crucial for accuracy of a derived mathematical model, described by stochastic parameters for packet size and inter-arrival time [28]. Using this simulation model, we want to acquire information about telecommunication network's performances for:

- improvement of the current network,
- bottleneck searching,

- building and development of new network devices and protocols,
- and for ensuring quality of service (QoS) for real-time streaming multimedia applications.

The simplest tools that measure and capture the packets of network traffic are packet sniffers. Packet sniffers, also known as protocol or network analyzers, are tools that monitor and capture network traffic with all content of network traffic. We can use sniffers to obtain the main information about network traffic, such as packet size, inter-arrival time and the type and structure of IP protocol. Sniffers have become very important and indispensable tools for network administrators [29].

An analytical description of network traffic does not exist, because we cannot predict the size and arrival time of the next packet. Therefore, we can only describe network traffic as a stochastic process. Hence, we have tried to describe these two stochastic processes, arrival time and packet size [29].

2.3 History of Network Traffic Modeling

> "Die Geschichte soll nicht das Gedï¿œchtnis beschweren, sondern den Verstand erleuchten." (The history should not burden the memory, but enlighten the brain)
>
> Gotthold Ephraim Lessing

In this section we review the history of network traffic modeling.

2.3.1 Markovian Models

Markovian model was the first proposed model of traffic process [30, 31]. It is demonstrated that point processes with certain "bursty" features can be qualitatively modeled by the Markovian arrival process in [32]. The Markovian arrival process is proposed to model superposed ATM cell streams [33]. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance has been studied in [5]. The Markov-Modulated Poisson Process (MMPP) and the Markov-Modulated Fluid (MMF) process are developed as simple and accurate analytical techniques to determine the loss probability at an access node to an Asynchronous Transfer Mode (ATM) network [34]. A class of versatile discrete-time Markovian arrival processes is introduced to model b-ISDN traffic [35].

A Markovian model is a mathematical model for the time between job arrivals to a system. The simplest case is a Poisson model where the time between each arrival is exponentially distributed.

Markovian models are simple to calculate. However, they are not adequate for modeling of network traffic. In particular, Markovian models cannot exhibit strong correlations structure in an arrival process [1]. Several techniques have been proposed to establish models which are in some sense Markovian, but which can capture the more complex properties of real traffic [36–38].

2.3.2 Self-Similarity and Long Range Dependence Models

Self-similar models are the mostly common used models of aggregated traffic process.

Traditional traffic modeling has been based on the assumption of independence between the random variables that describe arrivals to a network. The fundamental reason for this assumption has been the analytical tractability. However, high-speed network traffic is characterized by a high "burstiness" and a strong positive correlation [26] and non-Markovian effects, such as high variance and long range dependence, have been observed in a wide variety of traffic sources [7–11, 13–19, 39–43], self-similarity is therefore widely perceived as a statistical feature of network traffic. Self similar model is regarded as a rigorous model of Ethernet local area network (LAN) traffic and has serious implications for the design, control, and analysis of high-speed, cell-based networks. In the following the common used definition of self-similarity is given [9, 12, 13].

A **self-similar process** is the one for which aggregation has no impact on the nature of the process and is invariant in distribution at different degrees of magnification, or different scales on a dimension. If we take a stationary process X_t and denote by $X_t^{(m)}$ a new series obtained by averaging the original one over the non overlapping blocks, so called aggregated process

$$X_t^{(m)} = \frac{1}{m}\left(X_{tm-m+1} + \ldots + X_{tm}\right),$$

then X_t is **strictly self-similar** with parameter H if X_t and $\frac{1}{m^{1-H}} X_t^{(m)}$ are identically distributed. Here H is called **Hurst parameter**.

A less strict definition is that of **second-order self-similarity**. The process $\{X_t\}$ is exactly **second-order self-similar** if

$$r^{(m)}(k) = r(k)m^{2(H-1)}, 0.5 < H < 1,$$

where

$$r(k) = \mathrm{Cov}\left(X_t, X_{t+k}\right) = \mathrm{E}\left[\left(X_t - EX_t\right)\mathrm{E}\left(X_{t+k} - EX_{t+k}\right)\right]$$

and

$$r^{(m)}(k) = \mathrm{Cov}\left(X_t^{(m)}, X_{t+k}^{(m)}\right) = \mathrm{E}\left[\left(X_t^{(m)} - EX_t^{(m)}\right)\mathrm{E}\left(X_{t+k}^{(m)} - EX_{t+k}^{(m)}\right)\right].$$

Here E denotes the expectation and Car the covariance

A process X_t is **asymptotically second-order self-similar** if

$$\mathrm{Var}\left[X_t^{(m)}\right] \sim cm^{2(H-1)} \tag{2.1}$$

where Var denotes the variance.

Here (2.1) is the working definition of high variance.

A sufficient condition for second-order self-similarity is that the auto-correlation function (ACF) is

$$r(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right].$$

If $H = \frac{1}{2}$, this process is uncorrelated $r(k) = 0$.

A process $\{X_t\}_t$ is **long range dependent** (LRD) if its autocorrelation function $r(k)$ decays hyperbolically

$$r(k) \sim k^{2(H-1)}. \tag{2.2}$$

In the self-similar model, changes of the time scale do not affect the distributions of the observed process. However, Internet traffic presents burstiness in a statistical sense only over several time scales [3], and hence self-similar model can not capture the essential nature of network traffic.

2.3.3 Autoregressive Integrated Moving Average Models

Autoregressive Integrated Moving Average (ARIMA) process is used to model traffic process.

In statistics and econometrics, and in particular in time series analysis, Autoregressive Integrated Moving Average (ARIMA) models of order $(p; d; q)$ are generalization of the ARMA$(p; q)$ models and the most general class of models for forecasting a time series. These models apply to some cases where data show evidence of non-stationarity and can be made to be "stationary" by differencing. An ARIMA process is such a stochastic process $\{X_t\}_{t \in \mathbb{N}}$ such that

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

where L denotes a lag operator $LX_t = X_{t-1}$, ϵ_t denote the white noises, and ϕ_i, θ_i denote the parameters.

Factorizing the operator-valued polynomial $1 - \sum_{i=1}^{p} \phi_i L^i$ we obtain

$$\prod_{i=1}^{p} (1 - \lambda_i L) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

where λ_i are the roots of the polynomial

$$1 - \sum_{i=1}^{p} \phi_i \lambda^i = 0.$$

Finally

$$(1-L)^{\,d} X_t = \prod_{i=1}^{p} (1-\lambda_i L)^{-1} \left(1 + \sum_{i=1}^{q} \theta_i L^i \right) \varepsilon_t \quad .$$

Let us consider, as a particular case, the model ARIMA $(1,1,1)$

$$(1-\phi L)\,(1-L) X_t = (1-\theta L)\,\varepsilon_t \quad .$$

then

$$(1-L) X_t = (1-\phi L)^{-1} \varepsilon_t - \theta\,(1-\phi L)^{-1} \varepsilon_{t-1}$$

Let

$$Y_t = (1-\phi L)^{-1} \varepsilon_t,$$

then

$$Y_t - \phi Y_{t-1} = \varepsilon_t \tag{2.3}$$

and

$$X_t - X_{t-1} = Y_t - \theta Y_{t-1}. \tag{2.4}$$

from (2.3)

$$Y_t = \sum_{k=1}^{t} \varepsilon_k \phi^{t-k} \tag{2.5}$$

Combining (2.4) and (2.5) we obtain finally

$$X_t = X_1 + \sum_{i=2}^{t} \left(Y_i - \theta Y_{i-1}\right).$$

The paper [44] presents a theoretical basis for modeling univariate traffic condition data streams as seasonal ARIMA processes. The thesis [45] contributes a specific application of time series outlier modeling theory to vehicular traffic flow data. This outlier detection and modeling procedure uncovered a common ARIMA model form among the seasonally stationary series.

ARIMA model is short range dependent model if d takes only integer values. However, it can be extended to a long range dependent form by considering a fractional value of d which forms the FARIMA model.

2.3.4 Fractional Autoregressive Integrated Moving Average Models

The FARIMA process is selected as a model for inter arrival times instead of traffic process [26].

FARIMA processes are the natural generalizations of standard ARIMA (p,d,q) processes when the degree of differencing d is allowed to take non-integer values. The process is defined as:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

with $0 < d < 0.5$.

A FARIMA(0,d,0) process is a stationary process with autocorrelation function [46]

$$r(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}.$$

If the Hurst parameter $H = d + 0.5$, this is a long range dependent process.

The advantage of FARIMA models is that it can capture both long range dependent and short range dependent correlation structure. However, generation of FARIMA process is in general slower than for other long range dependent source type [1]. A faster generation of FARIMA process is proposed in [26]. A variety of FARIMA process were used to model VBR video traffic [10], to analyze the queueing performance and to provide the input stream to permit the evaluation of different techniques [47–49].

2.3.5 Fractional Brownian Motion

Fractional Brownian motion (FBM) is a generalization of Brownian motion. Unlike classical brownian motion, the increments of fractional brownian motion need not be independent. The fractional brownian motion is a continuous time Gaussian process $B_H(t)$ with zero mean and covariance function of the form

$$\mathrm{E}\left[B_H(t)B_H(s)\right] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)$$

where H is called **Hurst index**.

Definition 2.1. A process $B_H(t)$ is called **Fractional Brownian Motion** with Hurst parameter $0 < H < 1$ if and only if it is stationary increment self-similar Gaussian process.

The value of H determines what kind of process the FBM is:

- if $H = 1/2$ then the process is in fact a Brownian motion or Wiener process;
- if $H > 1/2$ then the increments of the process are positively correlated;
- if $H < 1/2$ then the increments of the process are negatively correlated.

An abstract model based on fractional brownian motion is presented in [50]. This model gives an accurate prediction if a huge amount of ON/OFF sources with heavy tails are superposed [51]. In this same work, the authors also show a relationship between the queueing performance of this limiting, LRD Gaussian process and that of fractional Brownian motion. It was shown in [52] that fractional Brownian motion models the arrivals of network packets better than classical models (at least for some types of traffic), but is parsimonious in the sense that only few parameters describe its statistical behavior. In [53] fractional Brownian motion model of data network traffic is constructed at the application level to formulate some new and relevant research problems about LRD and head towards a further understanding of the users' influence upon the highly complex nature of packet flows in large scale data networks. This FBM process can be considered as a limiting case for many other long range dependent models [1]. Arrival process and ON/OFF source model which allows for long packet trains and long inter-train distances converge to a multi-dimensional reflected fractional Brownian motion [54, 55].

Fractional Brownian motion is the only self-similar Gaussian process. For this reason it cannot explain the essential nature of Internet traffic.

2.3.6 Poisson Pareto Burst Process

Poisson Pareto burst process is an effective model for bursty traffic types [1]. In the PPBP, the burst durations t_i are independent and identically distributed Pareto random variables with distribution function

$$P\left[t_i < x\right] = 1 - \left(\frac{x}{\delta}\right)^{-\gamma}, x > \delta.$$

The burst durations t_i, are independent and identically distributed Pareto random variables, having the same distribution as random variable.

The Poisson Pareto burst process is a simple but accurate model for bursty traffic types. It appears to reflect the basic properties of at least some aggregated data traffic. It can be used to draw some interesting conclusions about the possible evolution of traffic loads on data networks [1].

Poisson Pareto Burst model is a special case of the model proposed in this thesis. In chapter 5 we will explain the hidden mechanism leading to Poisson Pareto Burst model.

2.3.7 log-Normal Models

Log-normal model is a model for inter delay times.

In traffic modeling a simple multiplicative law plays a fundamental role and provides an accurate way to model the underlying structure of network time series. This multiplicative law was introduced in [4, 22, 23]. It leads to log-normal distributions and explains self-similar-like behavior of some traffic variables.

Consider a chain of routers $R_1, R_2, \ldots, R_N$. Through this chain two consecutive packets are sent from a source to a destination, with initial inter delay time τ_0, which can be understood as time intervals between two considered packets. After passing the n-th router, the inter delay time between those two packets changes to τ_n, influenced by the lateral traffic. The transverse traffic on each router can be expressed by the random variables ξ_k. Let τ_n denote the inter delay time between those two test packets after passing n-th router and ξ_n represent the lateral traffic traversing on the n-th router. Then

$$\tau_{n+1} = \tau_0 \prod_{i=1}^{n} (\xi_i + 1) \,.$$

This property is called multiplicative law. Further we can conclude that the inter delay times τ_n are asymptotically log-normally distributed.

This multiplicative law can be considered as a theoretical principle in network modeling, and it is the base of multiple path model of network traffic presented in chapter 5.

2.3.8 LogPH Models

In the recent years, a new class of distributions, namely LogPH distributions, has been introduced for modeling heavy tails in traffic sources for network performance evaluation. It was first proposed in [56] to model Wi-Fi network traffic by using

LogPH distributions. A detailed mathematical treatment of LogPH can be found in [24].

Definition 2.2. A **LogPH random** variable Y is a random variable which can be expressed in the form $Y = e^X$ where X is a random variable with distribution constructed by a convolution of exponential distributions.

The studies of LogPH Network traffic modeling are empirically established, based on two data sets: WWW file size traces by Crovella in 1995 and resent data set of file sizes from the mobile web [25]. By comparison of the LogPH model with various other models like Pareto, Weibull, LogNormal and Log-t model, the authors demonstrate that LogPH fits the best to the real-world network traffic trace and provides more accurate results in network traffic prediction than all these models.

The LogPH model is a special case of the multiple path model presented in chapter 5 and it can be applied to single path routing network. We will provide in chapter 5 a multiple path model to explain the basic mechanism leading to this model. And we will demonstrate that this multiple path model can deal with both single path routing network and multiple path routing network.

Summary

In this chapter we present some basic knowledge of network traffic and review the history of network traffic modeling.

Markovian models are the first proposed models. They are simple to calculate, but they do not capture the strong correlation feature of network traffic.

Self-similarity models seem to provide a better prediction, but the basic mechanism leading to this model is unclear.

Autoregressive Integrated Moving Average models handle only short range dependence.

Fractional Autoregressive Integrated Moving Average models can capture both short range dependence and long range dependence. However, generation of FARIMA process is in general slower than for other long range dependent process.

The Fractional Brownian Motion can be considered as a limiting case for many other long range dependent models.

The Poisson Pareto Burst Process model is a rigorous model and is a special case of the model proposed in this thesis.

The multiplicative law, which leads to log-normal model, can be considered as a theoretical principle in network modeling, and it is the base of the multiple path model of network traffic in this thesis.

The LogPH model is experimentally proved to be a better model than the other heavy-tailed models. However, it lacks the mathematical base.

Part I

Chapter 3

A Convolution Invariant Parameter

Establishment of a stochastic model of network traffic can not get away from settings of parameters, and therefore the existence and rationality of parameters are essential. For most parameters in stochastic models, such as mean and variance, the existence and rationality are ensured by integrability or summability.

In this chapter we present a convolution invariant parameter, the existence and rationality of which, unlike the traditional parameters, are not ensured by integrability or summability, but by the existence of failure rate limit.

Using this parameter we will show that the distribution of random sum is

$$F(x) = 1 - e^{-\lambda x} f(x)$$

where f is such a function that

$$\lim_{x \to \infty} \frac{f(x)}{e^{sx}} = 0$$

for all $s > 0$.

This property is essential to modeling of network traffic.

3.1 A Convolution Invariant and Failure Rate Limit

A Convolution Invariant Parameter

In a sense, scientific research consists in discovery of invariant in a state of flux. Once we discover invariant, we capture the true nature of phenomena.

In mathematical terms, an invariant is a quantity related to a class of objects that remains unchanged when a certain class of modifications are applied to the objects. Invariants are an important means in classification problems: Objects with diverse invariants are essentially distinct; the converse of this statement also applies, that is, objects with equal invariants partake of the same nature.

Convolution is one of the most important operations in probability theory. In [57] it is revealed that convolution can be expressed as an infinite mixture of gamma distribution and in [58] it is demonstrated that limiting failure rate of a convolution behaves like the limiting failure rate of the strongest component. The proof in [58] does not cover the case when the strongest component has an unbounded failure rate, or in other words, the strongest component is heavy-tailed. Our result is similar but the proof is based on a different approach. Firstly we introduce a convolution invariant parameter.

A convolution invariant, as the name suggests, is an under convolution unaltered parameter or quantity. Firstly we consider for a given distribution F the parameter

$$C = C(F) := \sup\left\{t \in \mathbb{R}_{\geq 0} : \limsup_{x\to+\infty} \overline{F}(x)e^{xt} = 0\right\} \tag{3.1}$$

where $\overline{F}$ denotes the tail function of distribution F. In what follows, we prove step by step that this parameter is a convolution invariant, in a mathematical way.

In trying to deal with convolution, as it appears in (3.2), we need the equality [59]

$$\overline{F^{*2}}(x) = \overline{F}(x) + \int_0^x \overline{F}(x-y)dF(y) \tag{3.2}$$

as well as its more general form

$$\overline{F^{*(k+1)}}(x) = \overline{F}(x) + \int_0^x \overline{F^{*k}}(x-y)dF(y), \forall k \in \mathbb{N}. \tag{3.3}$$

The basic idea to prove that C in (3.1) is convolution invariant is to show

$$\left\{s \in \mathbb{R}_+ : \limsup_{x\to+\infty} \overline{F}(x)e^{xs} = 0\right\} = \left\{s \in \mathbb{R}_+ : \limsup_{x\to+\infty} \overline{F^{*k}}(x)e^{xs} = 0\right\}$$

for a given distribution F. In order to do this we need the following three propositions.

Proposition 3.1. *If* $\limsup_{x\to+\infty} \overline{F}(x)e^{xs} = 0$*, then* $\limsup_{x\to+\infty} \overline{F^{*2}}(x)e^{xs} = 0$.

In order to prove this proposition we need the following lemmata.

Lemma 3.2. *If* $\limsup_{x\to+\infty} \overline{F}(x)e^{xs} = 0$ *for some positive* s*, then there exists positive* $s^* > s$ *such that*

$$\overline{F}(x) \leq e^{-xs^*}$$

as long as $x > M$ for some natural number M.

Proof. Given a positive ϵ, there is a positive number M such that $x > M$ implies

$$0 < \overline{F}(x)e^{xs} < \epsilon.$$

We fix an $\epsilon < 1$ and choose an s^* such that $s^* = s + \frac{\ln \frac{1}{\epsilon}}{M}$, then

$$\overline{F}(x) \leq \epsilon e^{-xs} \leq \epsilon^{\frac{x}{M}} e^{-xs} = e^{-x\left(s + \frac{\ln(1/\epsilon)}{M}\right)} \leq e^{-xs^*}.$$

$\square$

Lemma 3.3. *If $\limsup_{x\to+\infty} \overline{F}(x)e^{xs} = 0$, then there exists positive $s^* > s$ such that*

$$\overline{F^{*2}}(x) \leq \overline{F}(x) + \frac{([x] - 2M)\, e^{2s^*} + (2M+1)\, e^{s^*(M+1)}}{e^{s^* x}}$$

as long as $x > 2M$ for some natural number M.

Proof. We use the same s^* and M in lemma 3.2, then let $x > 2M$

$$\begin{aligned}
\int_0^x \overline{F}(x-y)dF(y) &\leq \sum_{k=0}^{[x]} \int_k^{k+1} \overline{F}(x-y)\, dF(y) \\
&= \sum_{k=0}^{[x]} \overline{F}(x-(k+1)) \left(\overline{F}(k) - \overline{F}(k+1)\right) \\
&= \left(\sum_{k=0}^{M-1} + \sum_{k=M}^{[x]-M-1} + \sum_{k=[x]-M}^{[x]} \right) \overline{F}(x-k-1)\, \overline{F}(k) \left(1 - \frac{\overline{F}(k+1)}{\overline{F}(k)}\right) \\
&\leq \sum_{k=0}^{M-1} e^{-(x-k-1)s^*} + \sum_{k=M}^{[x]-M-1} e^{-s^*(x-k-1)} e^{s^*k} + \sum_{k=[x]-M}^{[x]} e^{-s^*k} \\
&\leq M e^{-s^*(x-M)} + ([x]-2M)\, e^{-s^*(x-1)} + (M+1)\, e^{-s^*([x]-M)} \\
&\leq (x-2M)\, e^{-s^*(x-2)} + (2M+1)\, e^{-s^*(x-1-M)} \\
&= \frac{([x]-2M)\, e^{2s^*} + (2M+1)\, e^{s^*(M+1)}}{e^{s^*x}}
\end{aligned}$$

From 3.2

$$\overline{F^{*2}}(x) \leq \overline{F}(x) + \frac{([x]-2M)\, e^{2s^*} + (2M+1)\, e^{s^*(M+1)}}{e^{s^*x}}.$$

□

Now we are able to prove 3.1.

Proof. Applying lemma 3.3

$$\limsup_{x\to+\infty} e^{sx}\overline{F^{*2}}(x) \leq \limsup_{x\to+\infty} e^{xs}\overline{F}(x) + \limsup_{x\to+\infty} \frac{([x]-2M)\, e^{2s^*} + (2M+1)\, e^{s^*(M+1)}}{e^{(s^*-s)x}} = 0.$$

This proposition is proved. □

In the following we prove a more general proposition.

Proposition 3.4. *If* $\limsup_{x\to+\infty} \overline{F}(x)e^{xs} = 0$ *for some* $s > 0$*, then* $\limsup_{x\to+\infty} \overline{F^{*k}}(x)e^{xs} = 0$ *for any natural number* k*.*

Proof. We prove it by mathematical induction. Assume $\limsup_{x\to+\infty} \overline{F^{*k}}(x)e^{xs} = 0$ for a k, we need to prove by (3.3)

$$\lim_{x\to+\infty} e^{xs}\left[\sum_{y=0}^{x-1} \overline{F^{*k}}\left(x-y\right)\left(\overline{F}\left(y\right)-\overline{F}\left(y+1\right)\right)\right] = 0$$

which implies $\overline{F^{*(k+1)}}(x)e^{xs} \to 0$. Applying 3.2 to both F and F^{*k}: there exist positive numbers s_1, s_2 and M_1, M_2 such that

$$\overline{F}\left(x\right) < e^{-xs_1^*}$$

$$\overline{F^{*k}}\left(x\right) < e^{-xs_2^*}$$

as long as $x > M_1$ and $x > M_2$.

Let $M = \max\{M_1, M_2\}$ and $s^* = \min\{s_1, s_2\}$. Let $x > 2M$

$$\begin{aligned}
\int_0^x \overline{F^{*k}}(x-y)dF(y) &\leq \sum_{k=0}^{[x]}\int_k^{k+1} \overline{F^{*k}}\left(x-y\right)dF\left(y\right) \\
&= \sum_{k=0}^{[x]} \overline{F^{*k}}\left(x-(k+1)\right)\left(\overline{F}\left(k\right)-\overline{F}\left(k+1\right)\right) \\
&= \left(\sum_{k=0}^{M-1} + \sum_{k=M}^{[x]-M-1} + \sum_{k=[x]-M}^{[x]}\right)\overline{F^{*k}}\left(x-k-1\right)\overline{F}\left(k\right) \\
&\leq \sum_{k=0}^{M-1} e^{-(x-k-1)s^*} + \sum_{k=M}^{[x]-M-1} e^{-s^*(x-k-1)}e^{s^*k} + \sum_{k=[x]-M}^{[x]} e^{-s^*k} \\
&\leq \frac{\left([x]-2M\right)e^{2s^*} + \left(2M+1\right)e^{s^*(M+1)}}{e^{s^*x}}
\end{aligned}$$

Hence $\limsup_{x \to +\infty} \overline{F^{*(k+1)}}\,(x)\, e^{xs} = 0$. ◻

Proposition 3.5. *If $\overline{F}(x)e^{xs}$ tends to a positive constant c for some s, then $\overline{F^{*2}}(x)e^{xs}$ tends to $+\infty$ for all $k \geq 2$.*

Proof. What we need to demonstrate is that

$$\lim_{x \to \infty} \frac{\sum_{y=0}^{x-1} \overline{F}(x-y)\,[F(y) - F(y+1)]}{\overline{F}(x)} = +\infty.$$

Let $\overline{F}(x)e^{xs}$ converges to a positive constant C . We have two inequalities. For $\forall \epsilon > 0, \exists X > 0, \forall x > X$ the inequality

$$C - \epsilon < e^{sx}\overline{F}(x) < C + \epsilon$$

the inequality

$$\overline{F}(x)e^{xs} \geq \min \left\{ \overline{F}(0), \overline{F}(1)e^{1s}, ..., \overline{F}(X)e^{Xs}, C - \epsilon \right\}.$$

Let

$$m := \min \left\{ \overline{F}(0), \overline{F}(1)e^{1s}, ..., \overline{F}(X)e^{Xs}, C - \epsilon \right\}$$

Then we estimate

$$\begin{aligned}
\frac{\sum_{y=0}^{x-1} \overline{F}(x-y)\,[F(y) - F(y+1)]}{\overline{F}(x)} &\geq \frac{\sum_{y=X+1}^{x-1} \overline{F}(x-y)\,[F(y) - F(y+1)]}{\overline{F}(x)} \\
&\geq \frac{\sum_{y=0}^{x-1} \overline{F}(x-y)e^{s(x-y)} \left[F(y)e^{sy} - \frac{F(y+1)e^{s(y+1)}}{e^s} \right]}{\overline{F}(x)e^{xs}} \\
&\geq \sum_{y=X+1}^{x-1} \frac{m\left[\left(1 - \frac{1}{e^s}\right) C - \epsilon \left(1 + \frac{1}{e^s}\right) \right]}{C + \epsilon}
\end{aligned}$$

The last term tends to infinity as $x \to +\infty$. The proposition is proved. □

Combining propositions 3.1, 3.4 and 3.5, we see for F and its convolutions F^{*k}

$$\left\{ s \in \mathbb{R}_+ : \limsup_{x \to +\infty} \overline{F}\,(x)\, e^{xs} = 0 \right\} = \left\{ s \in \mathbb{R}_+ : \limsup_{x \to +\infty} \overline{F^{*k}}\,(x)\, e^{xs} = 0 \right\}. \tag{3.4}$$

for any k. By taking the supremum of (3.4) we obtain this important theorem

Theorem 3.6. *Parameter $C(F)$ defined through*

$$C\,(F) := \sup \left\{ t \in \mathbb{R}_+ : \lim_{x \to +\infty} \overline{F}(x) e^{xt} = 0 \right\}$$

is invariant under convolution, i.e.

$$C\,(F) = C\left(F^{*k}\right)$$

for all natural number k.

Example. We consider exponential distribution

$$F(x) = 1 - e^{-\lambda x}$$

and its convolution

$$F^{*2}(x) = 1 - e^{-\lambda x}\,(1 + \lambda x)\,.$$

Obviously

$$C\,(F) = C\left(F^{*2}\right) = \lambda.$$

We apply this theorem to compound distribution and obtain the main conclusion in this chapter.

So far we have proved that C is convolution invariant.

Failure Rate Limit

In reliability theory, failure rate function is a powerful tool. It is commonly used to analyze the reliability of the equipments or the components and to quantify the risk in terms of probability that an entity fails in a specific time intervals after having correctly functioned before time t [60] .

Failure rate function $m_F(x)$ is defined by

$$m_F(x) = \frac{f(x)}{1 - F(x)}$$

where F denotes a given distribution function and f denotes its density function. Here $\overline{F} = 1 - F$ is called **tail** or **survival function** of F. And obviously $\hat{m}_F(x)\,\Delta x$ can be interpreted as an approximation of conditional probability of a failure in $(x, x + \Delta x]$.

In the context of inter delay times, we can also introduce the concept of failure rate. It approximates the conditional probability that the text packet arrives in a small time interval $(x, x + \Delta x]$ given that it does not arrive before x.

Example 3.7. Infinite failure rate limit: The failure rate limit of standard normal distribution is infinite

$$\lim_{x\to\infty} m_F(x) = \lim_{x\to+\infty} \frac{\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}}{1-\Phi(x)} = \lim_{x\to+\infty} \frac{\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}x}{\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}} = \lim_{x\to+\infty} x = +\infty.$$

Here Φ denotes the standard normal distribution.

Then $m_F(x)\,\Delta x$ is very huge for large x. However, probability can not be greater than 1, and hence normal distribution is not appropriate for extreme event modeling.

Example 3.8. Zero failure rate limit: The failure rate limit of log-normal distribution is 0.

$$\begin{aligned}\lim_{x\to\infty} m_F(x) = \lim_{x\to+\infty}\frac{F'(x)}{1-F(x)} &= \lim_{x\to+\infty}\frac{F''(x)}{-F'(x)} = \lim_{x\to+\infty}\frac{\frac{d}{dx}\left(\frac{1}{x\sqrt{2\pi}\sigma}\ e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\right)}{-\frac{1}{x\sqrt{2\pi}\sigma}\ e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}}\\ &= \lim_{x\to+\infty}\frac{\frac{1}{x^2\sqrt{2\pi}\sigma}\ e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} + \frac{1}{x\sqrt{2\pi}\sigma}\ e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\ \frac{2(\ln x-\mu)}{x}}{-\frac{1}{x\sqrt{2\pi}\sigma}\ e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}}\\ &= 0\end{aligned}$$

So $\hat{m}_F(x)\,\Delta x \sim 0$ for large x. For this reason log-normal distribution, in general, heavy tail distribution is not sensitive to extreme event.

Example 3.9. Positive failure rate limit: The failure rate limit of exponential distribution is finite

$$\lim_{x\to\infty} m_F(x) = \lim_{x\to+\infty}\frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

Distribution with positive failure rate limit is suitable for extreme events and sensitive to extreme values. In practice a distribution of the form

$$F(x) = 1 - e^{-\lambda x} P(x)$$

is preferable.

We consider only distribution with positive failure rate limit.

Notice due to l'Hï¿œpital's rule

$$\lim_{x\to\infty} \frac{-\ln \overline{F}(x)}{x} = \lim_{x\to\infty} m_F(x).$$

In this section we prove failure rate limit

$$C(F) = \lim_{x\to\infty} m_F(x).$$

where $C(F)$ is defined in (3.1), and therefore is convolution invariant.

The following theorem demonstrates that applying convolution on distribution does not change the limit of rate function.

Theorem 3.10. *Failure rate limit is convolution invariant, i.e.*

$$\lim_{x\to\infty} m_F(x) = \lim_{x\to\infty} m_{F^{*k}}(x)$$

Lemma 3.11. *If $\overline{F}(x) e^{xs}$ converges to 0*

$$\lim_{x\to\infty} \overline{F}(x) e^{xs} = 0$$

for some positive s*, then*

$$s \leq \lim_{x\to\infty} m_F(x).$$

Proof. First if $\overline{F}(x)e^{xs}$ converges to 0 as $x \to +\infty$, then for any ϵ there exists a M such that

$$0 < \overline{F}(x)e^{xs} < \epsilon$$

as long as $x > M$. Thereby

$$-\frac{\ln \overline{F}(x)}{x} > -\frac{\ln \epsilon}{x} + s$$

Let x tend to infinity

$$s \leq \lim_{x\to\infty} -\frac{\ln \overline{F}(x)}{x} = \hat{m}_F(x).$$

□

Lemma 3.12. *If*

$$\limsup_{x\to+\infty} \overline{F}(x)e^{xs} < +\infty$$

and

$$\liminf_{x\to+\infty} \overline{F}(x)e^{xs} > 0$$

for some positive s*, then*

$$s = \lim_{x\to\infty} m(x).$$

Proof. Let C_1 and C_2 denote the limit superior and limit inferior, respectively

$$C_1 = \lim_{x\to\infty} \inf \overline{F}(x)e^{xs}$$

$$C_2 = \lim_{x\to\infty} \sup \overline{F}(x)e^{xs}$$

then given ϵ there exists a X such that

$$C_1 - \epsilon < \overline{F}(x) e^{xs} < C_2 + \epsilon$$

as long as $x > X$. Therefore

$$e^{-xs}(C_1 - \epsilon) < \overline{F}(x) < e^{-xs}(C_2 + \epsilon)$$

and further

$$-\frac{\ln(C_1 - \epsilon)}{x} + s > -\frac{\ln \overline{F}(x)}{x} > -\frac{\ln(C_2 + \epsilon)}{x} + s$$

Let x tend to infinity

$$\lim_{x \to \infty} -\frac{\ln \overline{F}(x)}{x} = s.$$

□

Lemma 3.13. *If there is a positive number s satisfying*

$$\begin{aligned} \lim_{x \to \infty} \inf \overline{F}(x) e^{xs} > 0 \\ \lim_{x \to \infty} \sup \overline{F}(x) e^{xs} < \infty \end{aligned} \tag{3.5}$$

for a given distribution F, then it is the unique one.

Proof. Let s be a positive number fulfilling (3.5)

$$\begin{aligned} C_1 &= \lim_{x \to \infty} \inf \overline{F}(x) e^{xs} \\ C_2 &= \lim_{x \to \infty} \sup \overline{F}(x) e^{xs}. \end{aligned}$$

We consider any positive number differing form s. Let s_2 be a larger positive number $s < s_2$, then

$$\overline{F}(x) e^{xs_2} = \overline{F}(x) e^{xs} e^{x(s_2 - s)} \geq (C_1 - \epsilon) e^{x(s_2 - s)}.$$

and then

$$\lim_{x \to \infty} \sup \overline{F}(x) e^{xs_2} \geq \lim_{x \to \infty} \sup (C_1 - \epsilon) e^{x(s_2 - s)} = \infty.$$

Similarly if $s > s_1$, then

$$\overline{F}(x) e^{xs_1} = \overline{F}(x) e^{xs} e^{x(s_1 - s)} \leq (C_2 + \epsilon) \frac{1}{e^{x(s - s_1)}}.$$

hence

$$\lim_{x \to \infty} \sup \overline{F}(x) e^{xs_1} \leq \lim_{x \to \infty} (C_2 + \epsilon) \frac{1}{e^{x(s - s_1)}} = 0$$

So s is the unique real number satisfying (3.5). □

Before proving our main theorem, we need a lemma.

Lemma 3.14. *Given a failure rate limit s, i.e.*

$$\lim_{x \to \infty} m_F(x) = s > 0$$

Then s cuts $\mathbb{R}_+$ into two parts $(0, s)$ and $(s, +\infty)$. For $s_1 \in (0, s)$

$$\lim_{x \to \infty} \overline{F}(x) e^{xs_1} = 0$$

and for $s_2 \in (s, +\infty)$

$$\lim_{x \to \infty} \overline{F}(x) e^{xs_2} = \infty$$

Proof. First we know

$$s - \epsilon < \frac{-\ln \overline{F}(x)}{x} < s + \epsilon$$

then

$$e^{-x(s-\epsilon)} > \overline{F}(x) > e^{-x(s+\epsilon)}.$$

And further

$$e^{-x(s-s_1-\epsilon)} > \overline{F}(x)\, e^{xs_1} > e^{-x(s-s_1+\epsilon)}.$$

Finally

$$\lim_{x\to+\infty} \overline{F}(x)\, e^{xs_1} = 0.$$

Analogy

$$e^{x(s_2-s+\epsilon)} > \overline{F}(x)\, e^{xs_2} > e^{x(s_2-s-\epsilon)}$$

and

$$\lim_{x\to+\infty} \overline{F}(x)\, e^{xs_2} = \infty.$$

□

Next we prove the main theorem 3.10.

Proof. We prove this theorem by contradiction. Let

$$s_1 = \lim_{x\to\infty} \frac{f(x)}{\overline{F}(x)}$$
$$s_2 = \lim_{x\to\infty} \frac{f^{*k}(x)}{\overline{F^{*k}}(x)}.$$

We assume $s_1 \neq s_2$ and take an s between s_1, s_2.

If $s_1 < s_2$, then $s_1 < s < s_2$. From lemma 3.14

$$\lim_{x \to +\infty} \overline{F}(x) e^{xs} = +\infty$$

and

$$\lim_{x \to +\infty} \overline{F^{*k}}(x) e^{xs} = 0$$

which is a contradiction to the equation 3.3.

If $s_2 < s_1$, then $s_2 < s < s_1$, then

$$\lim_{x \to +\infty} \overline{F}(x) e^{xs} = 0$$

and

$$\lim_{x \to +\infty} \overline{F^{*k}}(x) e^{xs} = +\infty$$

which is a contradiction to the proposition 3.4.

So

$$s_1 = s_2$$

In conclusion

$$\lim_{x \to \infty} m_F(x) = \lim_{x \to \infty} m_{F^{*k}}(x)$$

□

Example. Exponential distribution

$$F(x) = 1 - e^{-\lambda x}$$

the limit of its failure rate function

$$\lim_{x\to\infty} \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda$$

The two fold convolution of exponential distribution is

$$F^{*2}(x) = 1 - e^{-\lambda x}(1 + \lambda x)$$

whose failure rate limit is

$$\lim_{x\to\infty} \frac{\lambda^2 x e^{-\lambda x}}{(1+\lambda x) e^{-\lambda x}} = \lambda$$

Example. Combination of Erlang distributions

$$G = \alpha F_k + \beta F_j, \alpha + \beta = 1, 0 \le \alpha, \beta \le 1$$

then

$$G^{*2} = (\alpha F_k + \beta F_j)^{*2} = \alpha^2 F_{2k} + 2\alpha\beta F_{k+j} + \beta F_{2j}$$

Further

$$\lim_{x\to+\infty} \frac{g(x)}{1 - G(x)} = \lim_{x\to+\infty} \frac{g^{*2}(x)}{\overline{G^{*2}}(x)} = \lambda$$

The following theorem enables us to link convolution invariant parameter and failure rate limit.

Theorem 3.15. *$C(F)$ defined in (3.1) is the failure rate limit of distribution F^{*k}, i.e.*

$$C(F) = \lim_{x\to\infty} m_{F^{*k}}(x)$$

for any $k \in \mathbb{N}$.

Proof. It suffices to show

$$C(F) = \lim_{x \to \infty} m_F(x)$$

by the theorem 3.10.

Assume $C(F) < s < \lim_{x \to \infty} \frac{f(x)}{\overline{F}(x)}$. Then $s > C(F)$ implies

$$\lim_{x \to \infty} \overline{F}(x) e^{xs} = +\infty$$

from definition of $C(F)$ while $s < \lim_{x \to \infty} \frac{f(x)}{\overline{F}(x)}$ implies by 3.14

$$\lim_{x \to \infty} \overline{F}(x) e^{xs} = 0.$$

Contradiction.

Assume $C(F) > s > \lim_{x \to \infty} \frac{f(x)}{\overline{F}(x)}$. Then $s < C(F)$ implies

$$\lim_{x \to \infty} \overline{F}(x) e^{xs} = 0$$

However, $s > \lim_{x \to \infty} \frac{f(x)}{\overline{F}(x)}$ implies by 3.14

$$\lim_{x \to \infty} \overline{F}(x) e^{xs} = +\infty$$

Contradiction. □

Fact 3.16. *Finally*

$$C(F) = \lim_{x \to \infty} m_F(x)$$

Remark. This theorem finds such an application. Consider a system consisting of N independent random components, each of which can be described by a random variable X_n with distribution F. If we know the lifetime Y of the system depends only on the sum of X_n,

$$Y = g^{-1}\left(\sum_{n=1}^{N} X_n\right)$$

Here we assume that g is bijective, monotonically increasing and differentiable. Now we calculate the asymptotic failure rate of the system.

Let $S = \sum_{n=1}^{N} X_n$, then

$$m_S(t) \approx m_X(t)$$

if t is big enough. Notice

$$F_Y(t) = P[Y < t] = P[S < g(t)] = F_S(g(t))$$

then

$$m_Y(t) = \frac{F_S'(g(t))}{1 - F_S(g(t))}\frac{dg(t)}{dt} \approx \frac{F_X'(g(t))}{1 - F_X(g(t))} g'(t)$$

In conclusion, if we know the failure rate of the component, we can calculate the failure rate of the whole system.

3.2 Random Sum

Convolution invariant parameter has a strong relation to random sum.

Random sum originated from the famous Cramï¿œr-Lundberg model and is of great significance in insurance Mathematics [59, 61, 62].

Let's review the famous Cramï¿œr-Lundberg model. Suppose that we want to evaluate the total payment over a period from a portfolio or estimate inter delay time between two observed packets sent through a chain of routers, either using the individual or the collective model. Let N be a nonnegative integer-valued random variable and $X_1, X_2, ...$ a sequence of nonnegative random variables.

Definition 3.17. The random variable Y is said to have a **compound distribution** if it is of the form

$$Y = \sum_{i=1}^{N} X_i$$

where

1. N is a counting process,
2. random variables X_i are identically distributed and independent
3. X_i are independent of N

In this case random variable Y is called **random sum** of $\{X_i\}$ and X_i are called **components** of random sum. We review some basic properties of random sum and compound distribution, detailed proofs of which can be found in [61].

1. Expected random sum is exactly the product of expected counting process and expected components, i.e.

$$\mathrm{E}[Y] = \mathrm{E}[N]\,\mathrm{E}[X] \tag{3.1}$$

2. Compound distribution and its survival function can be expressed in form of convex combination
$$F_Y = \sum_{k\geq 1} p_k F_X^{*k} \tag{3.2}$$
Here $p_k = P[N = k]$.

3. Variance is of the form
$$\mathrm{Var}[Y] = \mathrm{E}[N]\,\mathrm{Var}[X] + \mathrm{Var}[N]\,(\mathrm{E}[X])^2 .$$

Example 3.18. Compound distribution of exponentially distributed random variables is of the form
$$F_Y(x) = \sum_{k=1}^{\infty} p_k E_k(x)$$
where $p_k = P[N = k]$ and E_k denote Erlang distributions
$$E_k(x) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n .$$

Proof. Trivial. □

In the following we reveal a relation between random sum and failure rate limit

Theorem 3.19. *Let G denote compound distribution of a given distribution F, i.e.*
$$F[x] = \sum_{k=1}^{+\infty} P[N = k]\, F^{*k}(x)$$

where N denotes a counting process, then

$$C(G) = C(F).$$

Proof. From [59] we have

$$\overline{G}(x) = \sum_{k=1}^{+\infty} p_k \overline{F^{*k}}(x)$$

then

$$\overline{G}(x)e^{xs} = \sum_{k=1}^{+\infty} p_k \overline{F^{*k}}(x)e^{xs}$$

If $\overline{F}(x)e^{xs}$ converges to 0 for some $s > 0$ and therefore all $\overline{F^{*k}}(x)e^{xs}$ converge to 0. Then

$$0 \leq \limsup_{x\to+\infty} \overline{G}(x)e^{xs} \leq \sum_{k=1}^{+\infty} p_k \limsup_{x\to+\infty} \overline{F^{*k}}(x)e^{xs} = 0$$

$\overline{G}(x)e^{xs}$ converges to 0. So

$$\left\{s \in \mathbb{R}_+ : \lim_{x\to+\infty} \overline{F}(x)\, e^{xs} = 0\right\} = \left\{s \in \mathbb{R}_+ : \lim_{x\to+\infty} \overline{G}(x)\, e^{xs} = 0\right\}$$

We take supremum of the above two sets and then

$$C(F) = C(G) \tag{3.3}$$

□

Corollary 3.20. *For a compound distribution G with components F*

$$G = \sum_{k\geq 1} p_k F^{*k}$$

then

$$\lim_{x\to\infty} m_G(x) = \lim_{x\to\infty} m_F(x).$$

Proof. We obtain by (3.3)

$$\lim_{x\to\infty} m_G(x) = C(G) = C(F) = \lim_{x\to\infty} m_F(x).$$

□

3.3 Distribution of Random Sum

Theorem 3.21. *Given a random sum*

$$Y = \sum_{i=1}^{N} X_i.$$

Then the distribution F_Y is of the form

$$F_Y(x) = 1 - e^{-\lambda x} f(x)$$

where f is an exponentially bounded function, i.e.

$$\limsup_{x\to+\infty} \frac{f(x)}{e^{sx}} = 0, \quad \forall s > 0$$

and λ is the failure rate limit of X_i.

Proof. Directly from corollary 3.20. □

In practice, power-law functions are preferable.

Definition 3.22. A function is said to be **power-law** if

$$\lim_{x\to\infty} \frac{p(x)}{x^\alpha} = C > 0$$

for some $\alpha > 0$.

Then the distribution of random sum is

$$F(x) = 1 - e^{-\lambda x} p(x) \tag{3.1}$$

where p denotes a power-law function.

Proposition 3.23. *Given a power-law function f. Then*

$$\lim_{x\to+\infty} \frac{\ln p(tx)}{\ln p(x)} = 1$$

for any $t > 0$.

Proof. We notice a power-law function can be expressed in the form

$$p(x) = x^\alpha O(x)$$

for some real number $\alpha > 0$. Here $O(x)$ denotes a function such that

$$\lim_{x\to+\infty} O(x) = C$$

Then

$$
\begin{aligned}
\lim_{x \to +\infty} \frac{\ln p(tx)}{\ln p(x)} &= \lim_{x \to +\infty} \frac{\ln x^{\alpha} + \ln t^{\alpha} + \ln O(tx)}{\ln x^{n} + \ln O(x)} \\
&= \lim_{x \to +\infty} \frac{\ln x^{\alpha}}{\ln x^{\alpha} + \ln O(x)} + \lim_{x \to +\infty} \frac{\ln t^{\alpha}}{\ln x^{\alpha} + \ln O(x)} + \lim_{x \to +\infty} \frac{\ln O(tx)}{\ln x^{\alpha} + \ln O(x)} \\
&= 1
\end{aligned}
$$

□

In conclusion, for random sum

$$Y = \sum_{i=1}^{N} X_i$$

Then the distribution is of the form (3.1), where λ is exactly the failure rate limit of random components X_i.

Summary

This chapter focuses on two mathematical tools: random sum and convolution invariant parameter.

Failure rate limit is a convolution invariant parameter, consequently it is a parameter of random sum.

This parameter plays a central role in our modeling of network traffic.

Chapter 4

A Single Path Model

In this chapter we present a single path model of inter delay time and throughput traffic.

4.1 A Single Path Model of Inter Delay Times

Packet inter delay time is the time between consecutive packets and it is of great importance in traffic management, monitoring, and control tasks in networks.

The packet inter delay time can be applied in many ways [63]. For example, estimated values of packet inter delay time are used to measure the traffic rate for the QoS-enabled Internet [64]. The rate estimation is an essential part of call admission [65], link-sharing [66], and fair scheduling algorithm [67].

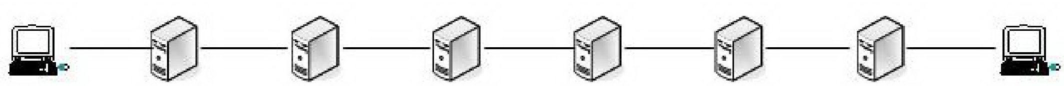
In the network security improvement, the packet inter delay time can be also used for investigating unsolicited internet traffic or for identifying the abnormal or unexpected network activity. The packet inter delay time patterns can be used to identify attacks or network phenomena. Estimation of the end-to-end performance and its improvement are important for web transactions [63, 68].

Moreover, the packet inter delay time plays a significant role in management of network devices and helps reduce consumption of network devices such as LAN Switches. An incoming packet is buffered and wakes up the switch interface. The decision on whether to sleep is based on the estimation of next packet inter delay time. If the inter arrival time is estimated to be long enough, interface goes to sleep, otherwise it stays awake [63, 69].

In the modeling of inter delay time, self-similarity and heavy tail are claimed to be observed [12, 14, 40, 43, 68].

Model Formulation

We consider such a one path model.



A sequence of packets are sent through a fixed chain of routers. We assume that all packets travel with the same speed and require the same service time on each router. We send two packets through this channel with some initial time delay τ_0 between them, which will be called the initial inter-delay time. We denote by τ_n the inter delay times between two observed packets after passing n^{th} router. If the considered packets are the only ones in the channel then the inter-delay remains the same when they reach the destination. In general, however, the inter-delay time can change on channel's routers. These changes are caused by the lateral traffic on the routers and the corresponding additional service times. The inter-delay times can both increase and decrease. The reason of the increase is obvious. The decrease of inter-delay times can happen when some packets from the transversal traffic leave the channel and, at the same time, there are also packets waiting for the service (we will call this situation buffering). The decrease cannot be, however, smaller than the service time of a single packet. The lateral traffic can influence packets delays in the following way. If an additional packet enters in between two observed packets then it increases the delay about the amount of time needed for its service.

The lateral traffic on each router will be represented by random variables $\xi_i^{(n)}, n = 1, 2, ...$, i.e. the difference between the number of packet that enter and leave during i^{th} time interval. In order to simplify the notation, let us also assume that each random variable $\xi_i^{(n)}$ contains the service time of the packet that enters a router. In other words, the equality $\xi_i^{(n)}$ means that no packets come into the chain from lateral traffic at time i. Now after passing $n+1$-th router the inter delay time will be

$$\tau_{n+1} = \sum_{i=1}^{\tau_n} \xi_i^{(n)} \tag{4.1}$$

Consider, if no packets enters and leaves the chain on the n-th router, all ξ_i^n take value 1 then $\tau_{n+1} = \tau_n$. If only one packet leaves the chain, then $\tau_{n+1} = \tau_n - 1$. So if we set $\xi_i^n \geq 0$, then both "entering" and "leaving" cases are included.

Furthermore we assume

1. inter delay times τ_n take only natural numbers,

2. lateral traffic $\xi_i^{(n)}$ are identically distributed and independent,
3. lateral traffic $\xi_i^{(n)}$ are independent of precedent inter del time τ_n.

In the following we will study this model in details.

Based on these three assumptions and applying the basic properties of random sum we obtain recursive expressions of the mean and the variance of inter delay time τ_{n+1}:

1. Expected inter delay time τ_{n+1} is exactly the product of expected lateral traffic and expected precedent inter delay time, i.e.

$$\mathrm{E}\left[\tau_{n+1}\right] = \mathrm{E}\left[\tau_n\right]\mathrm{E}\left[\xi\right]. \tag{4.2}$$

2. Variance is of the form

$$\mathrm{Var}\left[\tau_{n+1}\right] = \mathrm{E}\left[\tau_n\right]\mathrm{Var}\left[\xi\right] + \mathrm{Var}\left[\tau_n\right]\left(\mathrm{E}\left[\xi\right]\right)^2. \tag{4.3}$$

The following theorem provides general expressions of mean and variance of τ_{n+1}.

Theorem 4.1. *The means and variances are of the form*

1. $\mathrm{E}\left[\tau_{n+1}\right] = \tau_0\left(\mathrm{E}\left[\xi\right]\right)^{n+1}$
2. $\mathrm{Var}\left[\tau_{n+1}\right] = \begin{cases} \tau_0 \mathrm{Var}\left[\xi\right]\left(\frac{E^{2n+1}[\xi] - E^n[\xi]}{E[\xi]-1}\right), & \mathrm{E}\left[\xi\right] \neq 1 \\ \tau_0\left(n+1\right)\mathrm{Var}\left[\xi\right] & \mathrm{E}\left[\xi\right] = 1 \end{cases}$

Proof. We see trivially

$$\mathrm{E}\left[\tau_{n+1}\right] = \mathrm{E}\left[\tau_n\right]\mathrm{E}\left[\xi\right] = \mathrm{E}\left[\tau_0\right]\left(\mathrm{E}\left[\xi\right]\right)^{n+1} = \tau_0\left(\mathrm{E}\left[\xi\right]\right)^{n+1}.$$

The variance is given by

$$\mathrm{Var}\left[\tau_{n+1}\right] = \mathrm{E}\left[\tau_0\right]\left(\mathrm{E}\left[\xi\right]\right)^n \mathrm{Var}\left[\xi\right] + \mathrm{Var}\left[\tau_n\right]\left(\mathrm{E}\left[\xi\right]\right)^2.$$

Divided by $E^{2n+2}\left[\xi\right]$

$$\frac{\mathrm{Var}\left[\tau_{n+1}\right]}{\left(\mathrm{E}\left[\xi\right]\right)^{2n+2}} = \frac{\mathrm{E}\left[\tau_0\right]\mathrm{Var}\left[\xi\right]}{\left(\mathrm{E}\left[\xi\right]\right)^2}\left(\frac{1}{\mathrm{E}\left[\xi\right]}\right)^n + \frac{\mathrm{Var}\left[\tau_n\right]}{\left(\mathrm{E}\left[\xi\right]\right)^{2n}}.$$

If $E\left[\xi\right] \neq 1$

$$\frac{\mathrm{Var}\left[\tau_{n+1}\right]}{\left(\mathrm{E}\left[\xi\right]\right)^{2n+2}} = \frac{\mathrm{E}\left[\tau_0\right] Var\left[\xi\right]}{\left(\mathrm{E}\left[\xi\right]\right)^2}\frac{\left(1 - \left(\frac{1}{\mathrm{E}[\xi]}\right)^{n+1}\right)}{1 - \left(\frac{1}{\mathrm{E}[\xi]}\right)}$$

then

$$\mathrm{Var}\left[\tau_{n+1}\right] = \tau_0 \mathrm{Var}\left[\xi\right]\left(\frac{\left(\mathrm{E}\left[\xi\right]\right)^{n+1} - 1}{\mathrm{E}\left[\xi\right] - 1}\right)\left(\mathrm{E}\left[\xi\right]\right)^n.$$

If $E\left[\xi\right] = 1$

$$\mathrm{Var}\left[\tau_{n+1}\right] = \tau_0 \mathrm{Var}\left[\xi\right]\left(n+1\right).$$

There is a simplified model of this model, i.e.

$$\tau_{n+1} = \tau_0 \prod_{i=1}^{n}\left(\xi_i^{'} + 1\right)$$

instead of 4.1. For this model we found the limit distribution is log-normal [3]. This random sum model is more difficult to study. We have already evaluated the basic parameters, like expectation and variance. However, the study of tail and invariant is more interesting. In the following we will discuss the limit distribution. □

This stochastic process is obviously not stationary because mean and variance change with index n. However, failure rate limit is a convolution invariant parameter [**?**]

$$C(F) := \sup\left\{t \in \mathbb{R} : \lim_{x \to +\infty} \overline{F}(x)e^{xt} = 0\right\}$$

does not vary in this process.

Proposition 4.2. *Let F_n denote the distribution of inter delay time τ_n of the process of inter delay times $\{\tau_n\}_n$ remains unchanged. Moreover*

$$C(F_k) = C(F_1)$$

for all k.

Inter Delay Time τ_{n+1}

The inter delay time τ_{n+1} can be expressed as

$$\tau_{n+1} = \sum_{i=1}^{\tau_n} \xi_i^{(n)},$$

and therefore it is of the form

$$F(x) = 1 - e^{-\lambda x} p(x) \tag{4.4}$$

where $p(x)$ denotes a power-law function and λ denotes failure rate limit of lateral traffic ξ.

However, not each function of the form (4.4) is a valid distribution function. To construct a distribution of the form (4.4) we consider convex combination of Erlang distributions

$$F(x) = \sum_{i=1}^{n} \alpha_i E_i(x)$$

where E_i are Erlang distributions of the form

$$E_i(x) = 1 - e^{-\lambda x} \left(\sum_{k=0}^{i-1} \frac{(\lambda x)^k}{k!} \right)$$

and

$$\sum_{i=1}^{n} \alpha_i = 1, 0 \leq \alpha_i \leq 1.$$

For example, we consider a convex combination of Erlang distributions

$$F(x) = \alpha \left(1 - e^{-\lambda x}\right) + \beta \left(1 - e^{-\lambda x} \left(\sum_{i=0}^{k-1} \frac{(\lambda x)^i}{i!} \right)\right)$$

where $\alpha + \beta = 1$ and $0 \leq \alpha, \beta \leq 1$.

In the following we want to show that this distribution exhibits heavier tail than exponential distribution. First we evaluate the parameters mean and variance.

$$\mathrm{E}[X] = \alpha \frac{1}{\lambda} + \beta \frac{k}{\lambda} = \frac{\alpha + \beta k}{\lambda}$$

$$\mathrm{E}\left[X^2\right] = \alpha \frac{2}{\lambda^2} + \beta \frac{k^2 + k}{\lambda^2}$$

Variance

$$\mathrm{Var}\left[X\right] = \frac{\alpha\left(1-\alpha\right)k^2 + \left(1-\alpha\right)\left(1-2\alpha\right)k + \alpha\left(2-\alpha\right)}{\lambda^2}. \qquad \textbf{(4.5)}$$

In (4.5) k^2 term appears for $0 < \alpha < 1$. However, variance of poisson model increases linearly. Therefore, this model has a higher variance than Poisson model.

4.2 Single Path Model of Throughput Traffic

In communication networks, another one of the most frequently measured and discussed traffic variables is throughput traffic. Though the meaning of throughput traffic differs in different publications, the basic idea behind this concept is the volume of traffic in a regular time intervals [4].

Throughput traffic is a key quality-of-service metric in data networks. It describes the average "speed" of data transfer during a typical data connection. It is usually defined as the ratio of the average number of packets sent (or received) per data request to the average duration of the data transfer. Throughput traffic depends inherently on the requested data traffic and network architecture (positioning of the base stations) and in fact may significantly vary across different network [70].

It has been experimentally observed that the statistic characteristics of traffic volume differs significantly in different length of time intervals Δt. To be more precise, it is more shaped in long time intervals while usually undetermined for short time intervals. Different hypotheses concerning the observed distributions, which can vary with Δt, have been proposed. Here we discuss now a rigorous model for throughput traffic [4].

Throughput traffic is strongly related to inter delay times. Or, more accurately, there exists a strong statistical duality between these two traffic variables. Here we derive the duality between throughput traffic and inter delay times.

Let R_t denote the total number of test packets received up to time t. Here the stochastic process $\{R_t\}_t$ is called the **transmission process**. S_t denote the throughput traffic. Assume that the throughput traffic is, in average, proportional to the number of test packets received up to time t with integer constant σ

$$S_t = R_t \sigma.$$

Let $\tau^{(1)}, \tau^{(2)}, \ldots$ denote intervals between consecutive test packets. We assume that inter delay times are independent and identically distributed. Then we have

$$\begin{aligned}
R_t = 0 &\iff t < \tau^{(1)} \\
R_t = 1 &\iff \tau^{(1)} < t < \tau^{(1)} + \tau^{(2)} \\
R_t = n &\iff \sum_{i=1}^{n} \tau^{(i)} < t < \sum_{i=1}^{n+1} \tau^{(i)}.
\end{aligned}$$

Therefore

$$P\left[R_t < n\right] = P\left[\sum_{i=1}^{n} \tau^{(i)} > t\right] = 1 - P\left[\sum_{i=1}^{n} \tau^{(i)} \le t\right] = 1 - F^{*n}(t) \qquad (4.1)$$

where $F(x)$ is the distribution.

If inter delay time can be expressed by

$$F(t) = \sum_{i=0}^{k} \alpha_i E_i(t), \sum_{i=0}^{k} \alpha_i = 0, 0 \le \alpha_i \le 1$$

where E_k denotes the Erlang distribution and then

$$\begin{aligned} F^{*n}(t) &= \left(\sum_{i=1}^{k} \alpha_i E_i(t) \right)^{*n} \\ &= \sum_{j=n}^{nk} \beta_j E_j(t) \end{aligned}$$

where a convex combination $\sum_{j=n}^{\sigma n} \beta_j = 1, 0 \le \beta_j \le 1$.

So the distribution of throughput traffic is

$$P\left[S_t < \sigma n\right] = P\left[R_t < n\right] = 1 - \sum_{j=n}^{\sigma n} \beta_j E_j(t) = \sum_{j=n}^{\sigma n} \beta_j \left(1 - E_j(t)\right).$$

And the tail of S_t is of the form

$$P\left[S_t \ge \sigma n\right] = \sum_{j=n}^{nk} \beta_j E_j(t)$$

Tail of Poisson model

$$1 - \sum_{i=0}^{j-1} \frac{e^{-\lambda t} \left(\lambda t\right)^i}{i!} = E_j(t)$$

As

$$E_j(t) > E_{j+1}(t)$$

Let $N = nk$ and the tail of R_t is of the form

$$\sum_{j=n}^{nk} \beta_j E_j(t) > E_{nk}(t)$$

The tail is heavier than Poisson model.

The design of robust and reliable networks and network services has become an increasingly challenging task in the Internet world. To achieve this goal, understanding the characteristics of the Internet traffic plays a more and more critical role. Empirical studies of measured high-variance traffic traces have led to the wide recognition of self-similarity in network traffic [39]. Here we provide an alternative explanation of this high-variance traffic.

Summary

In this chapter we study in detail a random sum model or a single path model, which was proposed in [4].

This random sum model is a generalization of log-normal model. By using the convolution invariant parameter, we prove that the distribution of inter delay time can be expressed in the form of convex combination of Erlang distributions, and therefore it has a heavier tail than Poisson model.

Next we turn to the throughput traffic. We derive the distribution of throughput traffic by using a dual relation between inter delay time and throughput traffic.

Chapter 5

A Multiple Path Model

A more practical and natural model is multiple path model.

The original idea of the Internet is to construct a reliable global system of interconnected computer networks for the packet radio system work that could maintain effective communication in the face of jamming and other radio interference, or withstand intermittent blackout such as caused by being in a tunnel or blocked by the local terrain. For this purpose, B. Kahn, the inventor of Transmission Control Protocol (TCP) and the Internet Protocol (IP) posed four ground rules:

1. Each distinct network would have to stand on its own and no internal changes could be required to any such network to connect it to the Internet.

2. Communications would be on a best effort basis. If a packet didn't make it to the final destination, it would shortly be retransmitted from the source.

3. Black boxes would be used to connect the networks; these would later be called gateways and routers. There would be no information retained by the gateways

about the individual flows of packets passing through them, thereby keeping them simple and avoiding complicated adaptation and recovery from various failure modes.

4. There would be no global control at the operations level.

All these four rules make a multiple path model of internet traffic necessary.

In the packet switched networks, all transmitted data, regardless of content, type or structure, are divided into small packets for transmission. These packets are sent out from the source computer, travel along the most efficient path around the network to to the destination computer and finally reassembled into their proper sequence.

This does not necessarily mean that it is along the shortest path they travel. In fact, the path the packets traverse is conditioned by routing algorithms and network state.

Network State

A packet-switched computer network can reach a state that little or no useful communication is happening due to congestion, which is called congestive collapse. Congestion collapse generally occurs when the total incoming traffic to a node exceeds the outgoing bandwidth. When a network is in such a condition, packets may choose an alternative path to arrive their destination.

Routing Algorithm

Most currently deployed routing algorithms select at a time only a single path for the traffic between each source-destination pair. However, a division of traffic over multiple path is more flexible [71–77]. It could offer many benefits, including:

1. Customization of application performance requirements

2. Improvement of end-to-end reliability

3. Avoidance of congested paths

Though multi path routing has not been widely applied, due to scalability and economic challenges, "the economic incentives for providing value-added services will likely grow in the future and hopefully motivate the creation of new inter network business models that enable Internet-wide multi path routing" [77].

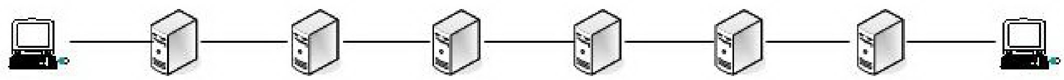
Based on these two points, a multi path model concerning network topological structure is more practical. In this model we assume that each packet may make a journey of its own and introduce a random variable N to represent the number of routers which the packets traverse. In fact, this model applies as well to analysis of single path communication if we specify the number of router by an integer n instead of a random variable N.

5.1 A Multiple Path Model of Inter Delay Times

In this section a multiple path model based on the multiplicative model [4] is present.

First we review the multiplicative model.

In multiplicative model we confine our discussion to a chain of routers $R_1, R_2, ..., R_N$, through which we send a sequence of packets. Between each two consecutive packets there is initially some initial inter delay time, denoted by τ_0.



Due to influence of lateral traffic, this inter delay time after having passed through n-th router is

$$\tau_{n+1} = \tau_n (1 + \xi_n)$$

where the random variables $\xi_k, k = 1, 2, ...$ represent lateral traffic. We assume that ξ_k are non negative integers. In such case ξ_k can be interpreted as the quantitative change of the unit gap on the k-th router.

Suppose that we observe two packets traveling from a source to a destination. We will denote by τ_j the inter delay time between these packets after having passed through the j-th router R_j, then

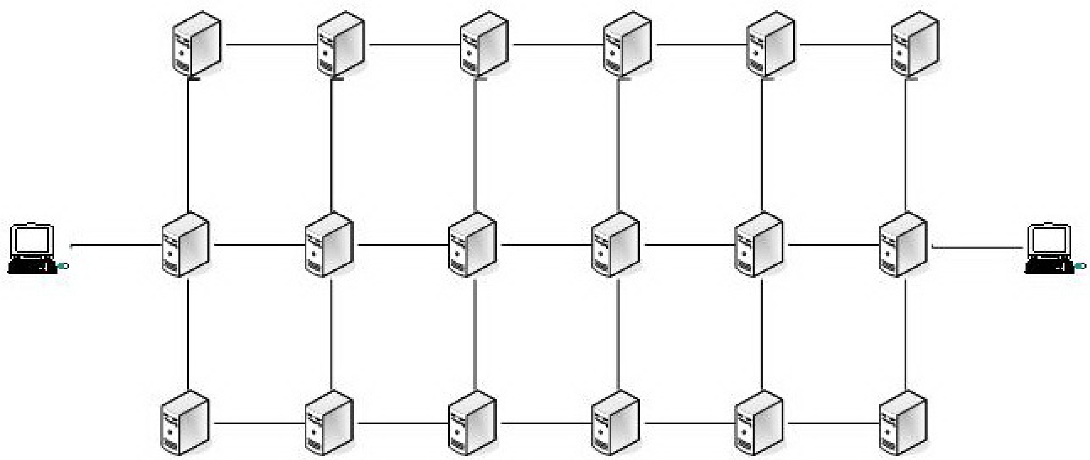
$$\tau_{n+1} = \tau_n (1 + \xi_n) .$$

Recursively we obtain

$$\tau_{n+1} = \tau_0 \prod_{k=1}^{n+1} (1 + \xi_k) .$$

Now we generalize this multiplicative model to net-like model.

In this model, there are more than one paths between source and destination, so the model can reflect the circumstances of real-world Internet traffic.



Two consecutive packets are sent from a source to a destination through a multiple path network, with initial inter delay time τ_0. This inter delay time can be changed due to disturb traffic on each router. The disturb traffic on the k-th router that the two packets actually pass can be expressed by the random variables ξ_k. Here we assume that ξ_k are identically distributed and independent.

Along each possible path from source to destination, inter delay time obeys a multiplicative law. As the number of routers that these packets actually pass is uncertain, we use a random variable N to represent the number of routers between source and destination. Then the inter delay time at the destination r_* is

$$\tau_* = \tau_0 \prod_{k=1}^{N} (1 + \xi_k) \, .$$

Then we evaluate the basic parameters of this inter delay time. Here we assume $\tau_0 = 1$, without loss of generality.

Proposition 5.1. *Inter delay time τ_* has the k-th moment of the form*

$$\mathrm{E}\left[\tau_*^k\right] = \mathrm{E}\left[\left[\mathrm{E}\left[\xi^k\right]\right]^N\right].$$

Proof. We calculate

$$\mathrm{E}\left[\tau_*^k\right] = \mathrm{E}\left[\mathrm{E}\left[\tau_*^k \mid N\right]\right] = \mathrm{E}\left[\mathrm{E}\left[\left(\prod_{i=1}^{N} \xi_i\right)^k \mid N\right]\right] = \mathrm{E}\left[\prod_{i=1}^{N} \mathrm{E}\left[\xi_i^k \mid N\right]\right]$$
$$= \mathrm{E}\left[\left[\mathrm{E}\left[\xi^k\right]\right]^N\right].$$

□

Corollary 5.2. *The moment generating function M_Y of random product can be expressed in the form*

$$M_{\tau_*}(t) = \sum_{k \geq 0} \frac{\mathbb{E}\left[\mathbb{E}^N\left[\xi^k\right]\right]}{k!} t^k$$

and characteristic function

$$\varphi_{\tau_*}(t) = \sum_{k \geq 0} \frac{\mathrm{E}\left[\mathrm{E}^N\left[\xi^k\right]\right]}{k!} (it)^k .$$

Theorem 5.3. *The distribution of inter delay time τ_* is*

$$F_{\tau_*}(x) = 1 - x^{-\lambda} p(\ln x).$$

Proof. We define a random variable

$$h = \ln \tau_*$$

then

$$h = \sum_{k=1}^{N} \ln\left(1 + \xi_k\right).$$

Apparently, h is compoundly distributed and the distribution is

$$F_h(x) = 1 - e^{-\lambda x} p(x)$$

where P denotes a power-law function. Then the distribution of τ_*

$$P\left[\tau_* < x\right] = P\left[\ln \tau_* < \ln x\right] = 1 - e^{-\lambda \ln x} p(\ln x) = 1 - x^{-\lambda} p(\ln x). \quad (5.1)$$

□

In [18, 20, 39–42] it has been demonstrated that the distribution of the files transferred across the Internet is indeed heavy-tailed. Heavy-tailed distributions are those probability distributions whose tails are not exponentially bounded, that is, the tails functions tend toward zero more slowly than any exponential ones. A working definition for a heavy-tailed distribution is given as follows. Consider a random variable with tail distribution function $\overline{F}(x) = 1 - F(x)$. If

$$\overline{F}(x) \sim x^{-\alpha} L(x), 0 \leq \alpha \leq 2 \quad (5.2)$$

where $L(x)$ is a slowly varying function

$$\lim_{x \to +\infty} \frac{L(tx)}{L(x)} = 1, \forall t > 0,$$

then this distribution is heavy tailed.

This multi path model gives a sophisticated interpretation of the heavy-tailed inter delay time.

5.2 Multiple Path Model of Throughput Traffic

We have already derived a dual relation between inter delay time and throughput traffic in section 4.2

$$P\left[S_t < \sigma n\right] = P\left[R_t < n\right] = 1 - P\left[\sum_{i=1}^{n} \tau_*^{(i)} \leq t\right] = 1 - F^{*n}(t)$$

where $F(x)$ is the distribution of inter delay times and R_t denote the test packets received up to time t and

$$\tau_*^{(i)} = \prod_{j=1}^{N_i}\left(1 + \xi_j^{(i)}\right)$$

Notice that $\prod_{i=1}^{n}\prod_{k=1}^{N_i}\left(1 + \xi_k^{(i)}\right)$ is a random product and it can be rewritten in the form

$$P\left[\prod_{i=1}^{n}\prod_{k=1}^{N_i}\left(1 + \xi_k^{(i)}\right) < x\right] = \sum_{j=1}^{m} \alpha_j E_j(\ln x).$$

Here m is a natural number.

Then

$$
\begin{aligned}
P\left[\sum_{i=1}^{n}\tau_*^{(i)} \le t\right] &= P\left[\sum_{i=1}^{n}\prod_{j=1}^{N_i}\left(1+\xi_j^{(i)}\right) < t\right] \le P\left[n\left(\prod_{i=1}^{n}\prod_{j=1}^{N_i}\left(1+\xi_i^{(i)}\right)\right)^{\frac{1}{n}} < t\right] \\
&= P\left[\prod_{i=1}^{n}\prod_{k=1}^{N_i}\left(1+\xi_i^{(i)}\right) < \left(\frac{t}{n}\right)^n\right] \\
&= \sum_{j=1}^{m}\alpha_j E_j(n\ln\frac{t}{n}).
\end{aligned}
$$

Here we use the inequality of arithmetic and geometric means

$$
\sum_{i=1}^{n}\frac{x_i}{n} \ge \left(\prod_{i=1}^{n}x_i\right)^{\frac{1}{n}}, \quad x_i > 0.
$$

Next we notice

$$
\left(1+\xi_1^{(1)}\right)\left(1+\xi_2^{(1)}\right)+\left(1+\xi_1^{(2)}\right)\left(1+\xi_2^{(2)}\right) \le 1+\left(1+\xi_1^{(1)}\right)\left(1+\xi_2^{(1)}\right)\left(1+\xi_1^{(2)}\right)\left(1+\xi_2^{(2)}\right)
$$

so

$$
\sum_{i=1}^{n}\tau_*^{(i)} = \sum_{i=1}^{n}\prod_{j=1}^{N_i}\left(1+\xi_j^{(i)}\right) \le n-1+\prod_{i=1}^{n}\prod_{j=1}^{N_i}\left(1+\xi_j^{(i)}\right)
$$

and again we use the inequality of arithmetic and geometric means

$$
\begin{aligned}
P\left[\sum_{i=1}^{n}\tau_*^{(i)} < t\right] &= P\left[\sum_{i=1}^{n}\prod_{j=1}^{N_i}\left(1+\xi_j^{(i)}\right) < t\right] \ge P\left[n-1+\prod_{i=1}^{n}\prod_{k=1}^{N_i}\left(1+\xi_i^{(i)}\right) < t\right] \\
&= P\left[\prod_{i=1}^{n}\prod_{k=1}^{N_i}\left(1+\xi_i^{(i)}\right) < t-n+1\right] = \sum_{j=1}^{m}\alpha_j E_j(\ln(t-n+1)).
\end{aligned}
$$

Finally we find the upper bound and lower bound for throughput process S_t

$$1-\sum_{j=1}^{m}\alpha_j E_j\left(\ln\left(\frac{t}{n}\right)^n\right)\le P\left[S_t<\sigma n\right]\le 1-\sum_{j=1}^{m}\alpha_j E_j\left(\ln\left(t-n+1\right)\right).$$

Let $\gamma = t-n$ and $n\to+\infty$

$$\ln\left(\frac{t}{n}\right)^n=\ln\left(1+\frac{t-n}{n}\right)^n\to t-n=\gamma$$

and

$$\ln\left(t-n+1\right)\approx t-n=\gamma.$$

Finally we can derive for throughput traffic

$$P\left[S_t<\sigma n\right]=1-\sum_{j=1}^{m}\alpha_j E_j\left(\gamma\right)$$

for some γ.

This explains why the Internet traffic presents high burstiness at small time scales and becomes smooth over a large range of aggregation levels.

Summary

In most of network traffic model, at least one random source is assumed to be heavy tail. In this chapter we contribute a new model. In this model, the inter delay time can still exhibit heavy tail even if all random source are light-tailed.

We introduce a random product model and prove that the inter delay time is heavy tailed by using convolution invariant parameter. Finally we derive the upper bound and lower bound of throughput traffic.

Part II

Chapter 6

A Homotopic Approach in Modeling of Network Traffic

In statistics, dependence is a statistical relationship between two events or two sets of data. Two events are said to be independent if the occurrence of one event makes it neither more nor less probable that the other occurs [78].

Calculation of the sum of distributed random variables is one of the most important problems in probability theory. In most of stochastic models, independence between objects is assumed. However, dependence occurs everywhere in reality. Therefore it is wise and necessary to introduce some dependence into our model.

We recall that the distribution sum $Z = \sum_{i=1}^{n} X_i$ of two identically distributed and independent random variables is of form

$$F_Z(x) = F^{*n}(x)$$

while if X and Y are strongly related, which implies $Z = nX$, then

$$F_Z(x) = F(\frac{x}{n}).$$

We assume that there is a continuous deformation linking the distributions $F^{*n}(x)$ to $F(\frac{x}{n})$. However, distribution is too restrictive and not appropriate for calculation. An equivalent but more convenient tool for specifying random variable X, characteristic function φ_X

$$\varphi_X(x) = E\left[e^{ixX}\right]$$

steps forward.

The homotopic approach was chosen because it allows to control positive definiteness of functions. This means that we always deal with characteristic functions instead of distribution functions.

In the following we present a homotopic approach to handle dependence in stochastic models.

6.1 A homotopic Approach

Homotopy is one of the fundamental concepts in algebraic topology. In algebraic topology, two continuous functions are called **homotopic** if one can be "continuous deformed" into another one, such a deformation being called a **homotopy** between the two functions [79]. In precise mathematical language, a **homotopy** is a continuous function $H : [0,1] \times \mathbb{R} \to \mathbb{C}$ of two given function $f_1, f_2 : \mathbb{R} \to \mathbb{C}$ such than

$$\begin{aligned} H(0,t) &= f_1(t) \\ H(1,t) &= f_2(t). \end{aligned}$$

Here we do not pursue in depth the field of algebraic topology in depth but we focus on the application of homotopy in stochastic modeling of network traffic.

We consider firstly the sum $Z = X_1 +, ..., +X_n$ of n identically distributed random variables X_n whose common characteristic function is denoted by φ. If they are independent, then the distribution of Z

$$\varphi_Z(x) = \varphi^n(x).$$

And they are strongly related, which implies $Z = nX$, then

$$\varphi_Z(x) = \varphi(nx).$$

We want to construct a deformation of characteristic functions with the following four requirements fulfilled:

Continuity

We assume that dependence can increase continuously, in other words, there exists a continuous process from independence to complete dependence.

Single Parametrization

In practice only first and second statistics are considered and the most familiar and often used measure of dependence between two random variables X and Y is the correlation coefficient, which is obtained by dividing the covariance of the two variables by the product of their standard deviations

$$\rho_{X,Y} = \frac{\mathrm{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}.$$

Under this premise, only one parameter suffices to represent dependence.

Invariable First Derivative at the Origin

The expected value operator is linear

$$\mathrm{E}\left[Y\right] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathrm{E}\left[X\right]$$

for all identically distributed random variables, dependent or independent. Therefore

$$\frac{d\varphi_Y(x)}{dx}|_{x=0} = in\mathrm{E}\left[X\right]$$

is free from dependence of X.

Positive Definiteness

By Bochner theorem A, a valid characteristic function of random variable must be positive definite.

Above all, our task boils down to construction of a one parametric continuous function (homotopy) connecting $\varphi^n(x)$ and $\varphi(nx)$ with positive definiteness preserved. To be more precise, we are to find a homotopy $H : [0, 1] \times \mathbb{R} \to \mathbb{C}$ such that

1. it connects $\varphi^n(x)$ and $\varphi(nx)$

$$
\begin{aligned}
H(0,x) &= \varphi^n(x) \\
H(1,x) &= \varphi(nx).
\end{aligned}
$$

2. $H(\alpha,x)$ is a positive definite function with respect to x for all $0 \le \alpha \le 1$

$$\sum_{i,j=1}^{n} H\left(\alpha, x_i - x_j\right) x_i x_j > 0, \forall x_i, x_j \in \mathbb{R}, \forall n \in \mathbb{N}.$$

3. the first partial derivative at the origin $\frac{\partial H(\alpha,x)}{\partial x}|_{x=0}$ is independent of α

$$\frac{\partial H\left(\alpha, x\right)}{\partial x}|_{x=0} = n \frac{\partial \varphi\left(x\right)}{\partial x}|_{x=0}.$$

   Here a homotopy can be comprehended as a characteristic function with a single unknown parameter.

In the following, we present two approaches of constructing such a homotopy. First we study a subclass of distribution, infinitely divisible distribution, and demonstrate that φ^α is a valid characteristic function. Second, we use convex combination in which positive definiteness is automatically preserved.

Construction of Homotopy for Infinitely Divisible Distribution

We recall the definition of infinitely divisible distribution.

Definition 6.1. A distribution F is **infinitely divisible** if for every natural number

n there exists a distribution F_n such that

$$F = F_n^{*n}$$

In other words, F is **infinite divisible** if and only if for each n it can be represented as the distribution of the sum of n independent random variables with a common distribution F_n.

Proposition 6.2. *Let φ is a characteristic function of an infinitely divisibly distributed random variable, then φ^α is also a valid characteristic function of some random variable for all positive α.*

Proof. it suffice to prove φ^α is a positive definite function for all real numbers.

Firstly we know from definition that $\varphi^{\frac{1}{n}}$ is a valid characteristic function and consequently $\varphi^{\frac{m}{n}}$ for all natural numbers m, n. Hence φ^α is positive definite for all rational numbers.

Secondly we assume φ^α is not positive definite function for some irrational number α, i.e. there exists a vector $(x_1, x_2, ..., x_n)$ such that

$$\sum_{i,j=1}^{n} x_i x_j \varphi^\alpha (x_i - x_j) < 0.$$

Let $(\alpha_n)_n$ be a sequence of rational numbers converging to α, so

$$\sum_{i,j=1}^{n} x_i x_j \varphi^{\alpha_n} (x_i - x_j) \geq 0$$

for all n and

$$\lim_{n\to\infty}\sum_{i,j=1}^{n} x_i x_j \varphi^{\alpha_n}(x_i - x_j) = \sum_{i,j=1}^{n} x_i x_j \varphi^{\alpha}(x_i - x_j).$$

A sequence of positive numbers converges to a negative number, which leads to a contradiction. The proposition is proved

We have proved that φ^α is well-defined characteristic function for all positive numbers, then we construct a homotopy

$$H(\alpha, x) = \varphi^{n-\alpha(n-1)}\left(\frac{nx}{n-\alpha(n-1)}\right)$$

□

Remark 6.3. Infinitely divisible distribution is strongly related to Levy process. Let X denote an infinitely divisibly distributed random variable and $X_{\frac{1}{n}}$ denote a random variable distributed by F_n, then $\{X_t\}_t$ is a Levy process.

Construction of Homotopy by Convex Combination

An another approach to construct such a homotopy is to use convex combination

$$H(\alpha, x) = \alpha\varphi(nx) + (1-\alpha)\varphi^n(x).$$

Obviously such H is continuous, one-parametric, with positive definite preserved and with invariable first derivative.

Parameter Determination

From probability theory, if random variable has second moments, then the characteristic function φ is twice continuously differentiable on the entire real line

$$\varphi_Y^{(2)}(0) = i^2 E Y^2.$$

As was mentioned above, a homotopy $H(\alpha, x)$ can be comprehended as a characteristic function with a single undetermined parameter α

$$-\frac{\partial^2}{\partial x^2} H(\alpha, x)|_{x=0} = m_Y^2$$

where m_Y^2 denotes second moment of random variable Y. In this equation all terms except α are known and parameter α can be determined by solving this equation. Finally the density function can be computed through inverse Fourier transform

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-isx} H(\alpha, s) ds.$$

Example. We consider sum $Y = \sum_{i=1}^{n} X_i$ of n exponential distributed random variables X_i. We construct a homotopy

$$H(\alpha, x) = \left(1 - \frac{inx}{\lambda\left(n - \alpha\left(n-1\right)\right)}\right)^{-(n-\alpha(n-1))}, 0 \leq \alpha_n \leq n-1$$

then

$$\begin{aligned}
\frac{\partial}{\partial x}H(\alpha,x)|_{x=0} &= -\left(n-\alpha\left(n-1\right)\right)\left(1-\frac{inx}{\lambda\left(n-\alpha\left(n-1\right)\right)}\right)^{-(n-\alpha(n-1))-1}\frac{-in}{\lambda\left(n-\alpha\left(n-1\right)\right)} \\
&= \left(1-\frac{inx}{\lambda\left(n-\alpha\left(n-1\right)\right)}\right)^{-(n-\alpha(n-1))-1}\frac{in}{\lambda}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial^2}{\partial x^2}H_n(\alpha_n,x)|_{x=0} \\
= & -\left(n-\alpha\left(n-1\right)+1\right)\left(1-\frac{inx}{\lambda\left(n-\alpha\left(n-1\right)\right)}\right)^{-(n-\alpha(n-1))-2}\frac{-n^2}{\lambda^2\left(n-\alpha\left(n-1\right)\right)}|_{x=0} \\
= & \frac{-\left(n-\alpha\left(n-1\right)+1\right)n^2}{\lambda^2\left(n-\alpha\left(n-1\right)\right)}.
\end{aligned}$$

Finally the density function of Y is

$$f_Y(x)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{-isx}\left(1-\frac{inx}{\lambda\left(n-\alpha\left(n-1\right)\right)}\right)^{-(n-\alpha(n-1))}ds$$

where

$$\frac{\left(n-\alpha\left(n-1\right)+1\right)n^2}{\lambda^2\left(n-\alpha\left(n-1\right)\right)}=m_Y^2.$$

6.2 Homotopic Approach in Network Traffic Modeling

A few studies have advocated the existence of upper bound in the time-scale associated with the correlation structure of TCP traffic and beyond which correlation becomes negligible [80]. Network traffic exhibits a correlation structure only over a finite range of time-scales [81], which is clearly not consistent with the concept of self-similarity. To mathematically describe this time scale related correlation structure, we need to introduce a new parameter in our model which represents this time scale dependent correlation structure. For this reason homotopic approach, which allows to describe continuous deformation between functions, is a useful tool in modeling network traffic.

We choose a time unit and denote by S_t the number of packets received up to time t. As traffic is time scale related, we use t to measure time scale. The stochastic process $\{S_t\}$ is called **transmission process**. We denote the mean

$$\mathrm{E}\left[S_t\right] = \mu_t$$

and variance

$$\sigma_t^2 = \mathrm{E}\left[S_t^2\right] - \mu_t^2.$$

We consider first the amount of packets received in a time unit, which can be represented by a random variable S_1 and let φ denote the given characteristic function of S_1. Now we want to calculate the characteristic functions of S_t, which represents the number of packets in time t.

For this propose we construct a homotopy $H_t(\alpha_t, x)$ such that

$$\begin{aligned} H_t(0, x) &= \varphi^t(x) \\ H_t(1, x) &= \varphi(tx) \end{aligned}$$

under some necessary assumptions.

Here the value of α_t depends uniquely on the dependence among packets received in time t and can be obtained by solving the equation

$$\frac{\partial^2 H_t(\alpha_t, x)}{i^2 \partial x^2}|_{x=0} = \mathrm{E}\left[S_t^2\right] = \sigma_t^2 + \mu_t^2$$

For each time scale t we can solve a parameter α_t.

In conclusion, the transmission process $\{S_t\}$ is characterized by the density functions

$$f_t(x) = \frac{1}{2\pi} \int_{\mathbb{R}} H_t(\alpha_t, s) e^{-ixs} ds$$

Example 6.4. We consider the characteristic function of poisson distribution

$$\varphi(x) = \exp(\lambda(e^{ix} - 1))$$

and construct a homotopy

$$H_t(\alpha_t, x) = (1 - \alpha_t) \exp(t\lambda(e^{ix} - 1)) + \alpha_t \exp(\lambda(e^{itx} - 1))$$

then

$$\frac{\partial^2}{\partial x^2} H_t(\alpha_t, x)|_{x=0} = (1 - \alpha_t)\left(\lambda^2 t^2 + \lambda t\right) + \alpha_t t^2 \left(\lambda^2 + \lambda\right) = \mu_t^2 + \sigma_t^2.$$

Finally the density function is

$$f_t(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (1 - \alpha_t) \exp(t\lambda(e^{iu} - 1)) + \alpha_t \exp(\lambda(e^{itu} - 1)) e^{-ixu} du$$

where

$$\alpha_t = \frac{\mu_t^2 + \sigma_t^2 - \lambda^2 t^2 - \lambda t}{\lambda t \left(t-1\right)}.$$

Summary

In modeling of network traffic, dependence is a factor that cannot be ignored as real network traffic exhibits strong positive correlation. In order to inject dependence into our models we present a homotopic approach.

We need a homotopy linking strong dependent case and independent case. A valid homotopy, which satisfies this condition, must be positive definite due to the Bochner theorem. We present two ways of constructing such homotopy.

Chapter 7

Conclusions and Perspectives

Summary of Contributions

This thesis began by arguing that new mathematical tools for internet traffic are required due to the failure of traditional mathematical models in network traffic. This dissertation proposed two mathematical tools: random sum and homotopy.

First we generalize from a multiplicative model to two random sum related model models: single path model and multiple path model, and give an explanation of heavy tail and long range dependence effects in network traffic.

In addition, we propose a new approach, homotopic approach, to model network traffic.

The one path model and multiple model concern more about network topology, such as source, router, destination such factors while homotopic model considers only dependence and correlation. For this reason, the first two models, in particular the multiple path model are more favorable in practice.

The most important contributions of this thesis can be summarized as follows:

- **Random Sum and Failure Rate Limit:** Our research began with a powerful mathematical tool: random sum. Most of the problems are highly random and complex, and for this reason random sum is a useful tool to analyze network traffic. In chapter 3, we proved that failure rate limit is invariant under convolution, and therefore can be taken as an invariant parameter under random summation. Furthermore we demonstrate by using this parameter that random sum can be approximated by a distribution equipped with a polynomial.

- **A Single Path Random Sum Model:** In 4 we employed random sum to derive a single path model of inter delay times and throughput traffic. We also showed that inter delay times under this model can be approximated by a convex combination of Erlang distributions. Finally, we deduced a model of throughput traffic by using a duality relation between inter delay times and throughput traffic.

- **A Multiple Path Random Product Model:** In packets switched network, packets are routed individually, sometimes resulting in different paths. In 5 we present a multi path model by using random product. Under this model, we can explain the mechanism behind the heavy tail and long range dependence of Internet traffic.

- **A Homotopic Approach in Modeling of Network Traffic:** In modeling of Internet traffic, dependence is a factor that cannot ignore. We introduce in Chapter 6a homotopic approach to analyze dependence of internet traffic. We constructed a homotopy linking strongly dependent case to independent case, with positive definiteness preserved. We obtained then the distribution by calculating the Fourier Transform of this homotopy.

Perspectives

- **Construction of an appropriate homotopy:** We present a homotopic approach to model network traffic. However, it is difficult to check positive definiteness of a given function. This makes construction of an effective homotopy with positive definiteness preserved a challenging task in practice.

- **New financial model instead of Black-Scholes model:** Normal distribution is inappropriate for extreme event modeling and this is probably the reason why it is claimed "the Black-Scholes equation was the mathematical justification for the trading that plunged the world's banks into catastrophe". However, a model based on a distribution of the form

  $$F(x) = 1 - e^{-\lambda x} f(x)$$

  seems to be a more reasonable model. Firstly it can handle heavy tail, secondly it is related to Levy process, and hence it can be taken as a valid integrand.

- **Relation between power-law distributions and exponential distributions:** Power-law distributions occur in many scientific fields and have significant consequences for understanding the nature of natural and man-made phenomena. There seems to be a relation between power-law distributions and exponential distributions. Consider a power-law distributed random variable X

$$P\left[X > x\right] \sim x^{-\lambda}$$

and let $X' = \ln X, x' = \ln x$. Then

$$P\left[X' > x'\right] \sim e^{-\lambda x'}.$$

In this way, many power-law related phenomena might be described by simple but hidden exponentially distributed random variables.

Appendix A

Probability theory

In this section we will outline some of the basics of probability theory needed in this dissertation.

Probability theory is the branch of mathematics concerned with probability, the analysis of random phenomena. The central objects of probability theory are random variables, stochastic processes, and events. As a mathematical foundation for statistics, probability theory is essential to science and engineering that involve quantitative analysis of large sets of data. Methods of probability theory also apply to descriptions of complex systems given only partial knowledge of their state, as in statistical mechanics.

Probability and Distribution

Definition A.1. A **probability space** $(\Omega, \mathcal{F}, P)$ consists of a basic space Ω, a σ-algebra $\mathcal{F}$ and a probability measure P. A **random variable** X is a measurable function on $(\Omega, \mathcal{F}, P)$. A **probability distribution** F_X of a random variable X on the real line is determined by the probability of a scalar random variable X in a half-open interval $(-\infty, x]$:

$$F_X(x) = P[X \leq x]$$

The **expected value** of random variable X is defined as

$$\mathrm{E}[X] = \int_\Omega X dP$$

the **variance**

$$\mathrm{Var}[X] = \int_\Omega (X - \mathbb{E}(X))^2 \, dP$$

the n^{th} **moment**

$$\mathrm{E}[X^n] = \int_\Omega X^n dP$$

Example A.2. In probability theory and statistics, the exponential distribution is the probability distribution of the amount of time until some event occurs. For instance, the amount of time (starting from now) until an earthquake occurs, or until a new war breaks out, or until a telephone call you receive turns out to be a wrong number are all random variables that tend in practice to have exponential distributions.

A continuous random variable X whose probability distribution is given, for some $\lambda > 0$, by

$$F_X(x) = 1 - e^{-\lambda x}$$

is called **exponential** random variable. The exponential distribution often arises, in practice, as being the distribution of the amount of time until some specific event occurs. For instance, the amount of time (starting from now) until an earthquake occurs, or until a new war breaks out, or until a telephone call you receive turns out to be a wrong number are all random variables that tend in practice to have exponential distributions.

Example A.3. The sum of n i.i.d exponentially distributed random variables with parameter λ is **Erlang distribution**

$$F^{*k}(x) = 1 - \sum_{j=0}^{k-1} \frac{e^{-\lambda x}(\lambda x)^j}{j!} \tag{A.1}$$

Proof. Define an integral operator $\mathcal{L}$ through

$$\mathcal{L}\Pi(x) = \int_0^x \lambda e^{-\lambda y}\Pi(x-y)dy$$

Let $\Pi(x) = e^{-\lambda x}$, then

$$\mathcal{L}^k\Pi(x) = \frac{e^{-\lambda x}(\lambda x)^k}{k!}$$

We prove (A.1) by mathematical induction. Assume this equality is satisfied, then

$$\begin{aligned}\mathcal{L}^{k+1}\Pi(x) &= \mathcal{L}\mathcal{L}^k\Pi(x) = \int_0^x \frac{\lambda e^{-\lambda y}e^{-\lambda(x-y)}(\lambda(x-y))^k}{k!}dy \\ &= \frac{e^{-\lambda x}(\lambda x)^{k+1}}{(k+1)!}\end{aligned}$$

Let F denote exponential distribution function and define an integral operator $\mathcal{L}$ through

$$\mathcal{L}\Pi(x) = \int_0^x \Pi(x-y)dF(y)$$

Then

$$\begin{aligned}\overline{F^{*k}}(x) &= \overline{F}(x) + \int_0^x \overline{F^{*(k-1)}}(x-y)dF(y) \\ &= \overline{F}(x) + \mathcal{L}\overline{F^{*(k-1)}}(x) = \sum_{j=0}^{k-1} \mathcal{L}^j \overline{F}(x) \\ &= \sum_{j=0}^{k-1} \frac{e^{-\lambda x}(\lambda x)^j}{j!}\end{aligned}$$

Then

$$F^{*k}(x) = 1 - \sum_{j=0}^{k-1} \frac{e^{-\lambda x}(\lambda x)^j}{j!}$$

□

Characteristic Function

An equivalent specification of random variable is the characteristic function of a random variable. It was originally developed as a tool for the solution of problems in probability theory and admit many important applications in the branch of mathematics as well as in Mathematical Statistics. In particular, it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions [82].

Definition A.4. Characteristic function φ_X of a random variable X is defined by

$$\varphi_X(t) = \mathrm{E}\left[e^{itX}\right], t \in \mathbb{R}$$

Theorem A.5. [Uniqueness Theorem] *Two distribution functions F_1 and F_2 are identical if and only if their characteristic functions φ_1 and φ_2 are identical.*

Proof. See [82]. □

Theorem. [Bochner Theorem] An arbitrary function $\varphi : \mathbb{R} \to \mathbb{C}$ is the characteristic function of some random variable if and only if φ is positive definite, continuous at the origin, and if $\varphi(0) = 1$.

Radon-Nikodym Theorem

In probability theory, the Radon-Nikodym theorem is very important in extending the ideas of probability theory from probability functions and density functions defined over real numbers to probability measures defined over arbitrary sets. It provides a unified description for studying continuous and discrete systems.

Definition A.6. A measure ν is **absolutely continuous** with respect to μ, i.e. $\nu \ll \mu$ if $\mu(A) = 0$ implies that $\nu(A) = 0$.

Theorem A.7. *Let μ and ν be σ finite measures on space $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, there is a function $f : \Omega \to \mathbb{R}_{\geq 0}$ so that for all $A \in \mathcal{F}$*

$$\nu[A] = \int_A f d\mu$$

This Function f is usually denoted $\frac{d\nu}{d\mu}$ and called the **Radon-Nikodym derivative**.

Both probability function and density function are Radom-Nikodym derivative with respect to corresponding probability measure.

Independence, Sub independence and Conditional Independence

In probability theory, independence of two random variables is a fundamental concept, which means that the realization of one random variable does not affect the probability distribution of the other and extends to dealing with collections of more than two events or random variables [78, 83–85].

Definition A.8. Two random variables X and Y are said to be **independent** if

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$$

Sub independence is a weak form of independence [85].

Two random variables X and Y are said to be **sub-independent** if

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$$

This is a generalization of the concept of independence of random variables. Independence implies sub independence, but not conversely.

Proposition A.9. *If X and Y are independent, then they are sub-independent.*

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$$

Proof. Trivial □

Proposition A.10. *If X, Y, Z are independent, then X, Y are conditionally indepen-*

dent given Z.

Proof. Trivially

$$\mathrm{E}[XY|Z] = \mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[X|Z]\,\mathrm{E}[Y|Z]$$

$\square$

Appendix B

Heavy Tail

B.1 Heavy Tail in Probability Theory

In probability theory, heavy-tailed distributions are probability distributions whose tails are not exponentially bounded: that is, they have heavier tails than the exponential distribution [59, 61, 62, 86]. In practice three important class of heavy tail distribution are, the fat-tailed distributions, long-tailed distributions and sub exponential distributions, among which sub-exponential distributions are most commonly used.

Definition B.1. The distribution of a random variable X with distribution function F is said to have a **heavy tail** if

$$\limsup_{x \to +\infty} \overline{F}(x) e^{xs} = +\infty$$

for all $s > 0$.

Remark B.2. In practice, a working definition of heavy tail is, a heavy-tailed distribution is one that has a tail function of the form

$$\overline{F}(x) \sim x^{-\alpha}L(x), 0 \leq \alpha \leq 2$$

where $L(x)$ is a slowly varying function

$$\lim_{x\to+\infty}\frac{L(tx)}{L(x)} = 1, \forall t > 0$$

It has been shown in [18–20,39–43] that the distribution of the files transferred across the Internet is indeed heavy-tailed.

Proposition B.3. *X is* **heavy-tailed** *distributed if and only if $E\left[e^{sX}\right] = \infty$ for all $s > 0$.*

Proof. Firstly we assume if $\limsup_{x\to+\infty}\overline{F}(x)e^{xs} = +\infty$, then

$$E\left[e^{sX}\right] = \int_0^\infty e^{sx}dF(x) = \int_t^\infty e^{sx}dF(x) \geq e^{st}\int_t^\infty 1dF(x) = e^{st}\overline{F}(t)$$

Therefore

$$E\left[e^{sX}\right] \geq \limsup_{t\to+\infty} e^{st}\overline{F}(t) = +\infty$$

Secondly if $E\left[e^{sX}\right] = \infty$, then F is not bounded by exponential distribution, i.e. for any $a, b > 0$

$$\overline{F}(x) > ae^{-bx}$$

for some x. We choose $b < s$, then

$$\limsup_{x \to +\infty} \overline{F}(x) e^{sx} > \limsup_{x \to +\infty} a e^{(s-b)x} = +\infty$$

□

Example B.4. A **log-normal distribution** is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then $Y = \log(X)$ has a normal distribution. We consider the moment generating function.

$$\mathrm{E}\left[e^{tX}\right] = \sum_{n=0}^{\infty} \frac{e^{n\mu + \frac{\sigma^2}{2}}}{n!} t^n$$

Log-normal distribution is heavy tailed from proposition B.3.

Definition B.5. The distribution of a random variable X is said to have a **fat tail** if

$$f_X(x) = ax^{-(1+\alpha)}, \alpha > 0$$

Fat-tailed distribution is heavy-tailed as

$$\mathrm{E}\left[e^{tX}\right] = a \int_0^{+\infty} e^{tx} x^{-(1+\alpha)} dx = \infty$$

for all t.

B.2 Heavy Tail and Extreme Events Modeling

Extreme event modeling [59,62] concerns events that occur with relatively small probability but have a significant influence on the behavior of the whole model. For example, we consider extremal claims in insurance mathematics. We denote by $X_1, X_2, ..., X_n$ independent and identically distributed claims all with claim size distribution F and then we introduce a new random variable $\max_{1\leq i\leq n} X_i$. We may ask about the distribution of $\sum_{1\leq i\leq n} X_i$. We consider a class of large claim distributions. Their defining property is

$$\lim_{x\to+\infty} \frac{P\left[\sum_{1\leq i\leq n} X_i\right]}{P\left[\max_{1\leq i\leq n} X_i\right]} = 1$$

for every $n \geq 2$. Thus the tails of the distribution of the sum and of the maximum of the claims are asymptotically of the same order, which clearly indicates the strong influence of the largest claim on the total claim amount and it gives a natural description of extreme events.

Appendix C

Infinitely Divisible Distribution and Levy Process

Example C.1. In this section we introduce a class of distributions, namely **infinitely divisible distributions**. A detail discussion can be found in [87].

Definition C.2. A distribution F is **infinitely divisible** if for every natural number n there exists a distribution F_n such that

$$F = F_n^{*n}$$

In other words, F is infinite divisible if and only if for each n it can be represented as the distribution of the sum of n independent random variables with a common distribution F_n.

Example C.3. *Here are some examples of infinitely divisible distributions*

1. *The degenerated distribution with characteristic function*

$$\varphi(t) = e^{iat}$$

2. *Gamma distributions including exponential distribution with characteristic function*

$$\varphi(t) = \frac{1}{(1 - it\beta)^{\alpha}}$$

3. *Poisson distribution with characteristic function*

$$\varphi(t) = e^{\lambda(e^{it}-1)}$$

4. *Cauchy Distribution*

$$\varphi(t) = e^{it\mu - \theta|t|}$$

5. *The negative binomial distribution*

$$\varphi(t) = \left(\frac{1-p}{1-pe^{it}}\right)^{r}$$

6. *Normal distribution*

$$\varphi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Infinitely divisible distributions are intimately connected with Lï¿œvy process. In probability theory, a Lï¿œvy process, named after the French mathematician Paul Lï¿œvy, is a stochastic process with independent, stationary increments.

Definition C.4. A stochastic process $X = \{X_t : t \geq 0\}$ is said to be a **Lï¿œvy process** if it satisfies the following properties:

1. $X_0 = 0$ almost surely,

2. Independence of increments: For any $0 \leq t_1 < t_2 < \cdots < t_n < \infty, X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent,

3. Stationary increments: For any $s < t, X_t - X_s$, is equal in distribution to X_{t-s},

4. Continuity in probability: For any $\epsilon > 0$ and $t \geq 0$ it holds that

$$\lim_{h \to 0} P(|X_{t+h} - X_t| > \epsilon) = 0$$

Remark. The distribution of a Lï¿œvy process has the property of infinite divisibility: for any given natural number n

$$X_1 = \sum_{i=1}^{n} \left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \right)$$

Conversely, for any infinitely divisible probability distribution F, we can construct a Lï¿œvy process X such that the law of X_1 is given by F.

Appendix D

Application of Convolution Invariant in Finance

In economy and finance, it is widely acknowledged that heavy tail is one of the main properties of observed prices which is not verified by the Black & Scholes models. Empirical studies showed that prices and returns of assets obey to power laws [88,89]

$$1 - F(x) \sim kx^{-\alpha}$$

In [90] it is asked, "Can we fully explain the power law distribution of financial variables, particularly returns and trading volume?" In the following we propose an answer to this question.

Let S_t denote the prices of a financial asset where time is measured on days units. Then

$$S_t = S_0 \prod_{i=1}^{t} (1 + r_i)$$

Here r_i represents the interest rate at time i. We define

$$h_t = \ln S_t$$

then h_t can be expressed by sum of random variables, and therefore from (3.1)

$$P\left[h_t < x\right] = 1 - e^{-\lambda x} p_t(x)$$

So

$$P\left[S_t < x\right] = P\left[\ln S_t < \ln x\right] = P\left[h_t < \ln x\right] = 1 - x^{-\lambda} p_t(\ln x)$$

where p_t is a power-law function depending on t.

Bibliography

[1] T. Neame, *Characterisation and model of Internet traffic streams*. PhD thesis, The University of Melbourne, 2003.

[2] A. S. Tanenbaum, *Computer Networks*. Prentice Hall, 5 ed., October 7 2010.

[3] F. Melakessou, *Route Diversity on the Internet: A naatural WAy to improve the Network Traffic*. PhD thesis, University of Luxembourg, 2008.

[4] F. Melakessou, U. Sorger, and Z. Suchanecki, "A multiplicative law of network traffic and its consequences," *Acta Physica Polonica B.*, vol. 40, no. 5, 2009.

[5] H. Heffes and D. M. Lucantoni, "A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer perfornmance," *Journal on Selected Areas in Communications*, pp. SAC–4(6):856–867, September 1986.

[6] M. Zukerman and I. Rubin, "On multi channel queueing systems on multi channel queueing systems with fluctuating parameters," *Proceedings of IEEE Infocom*, pp. '86, '86 pages 600–608,, April 1986.

[7] D. Duffy, A. A. McIntosh, M. Rosenstein, and W. Willinger, “Statistical analysis of ccsn/ss/ traffic data from working ccs subnetworks,” *Journal on Selected Areas in Communications,*, pp. 12(13):544–551, April 1994.

[8] C. Huang, M. Devetsikiotis, I. Lambardis, and A. R. Kaye., “Modeling and simulation of self-similar variable bit rate compressed video,” *Computer Communications Review*, pp. 25:114–125, August 1995.

[9] M. Garrett and W. Willinger, “Logarithmic asymptotics for steady-state tail probabilities in a single server,” *Journal of Applied Probability*, pp. 31:131 – 156, 1994.

[10] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, “Long-range-dependence in variable-bit-rate video traffic,” *IEEE Transations on Communications*, pp. 43(2/3/4):1566–1579, February/March/April 1995.

[11] P. R. Jelenkovic, A. A. Lazar, and N. Semret., “The effect of multiple time scales and subexponentiallity in mpeg video streams on queueing behavior,” *IEEE Journay on Selected Areas in Communications*, pp. 15(6):1052–1071, August 1997.

[12] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, “On the self-similar nature of ethernet traffic,” *IEEE/ACM Transactions on Networking (TON)*, vol. 2, pp. 1–15, Feb. 1994.

[13] W. Willinger, M. S. Taqqu, and A. Erramilli., “Abibliographical guide to self-similar traffic and performance modeling for modern high-speed networks,” *Stochastic Networks: Theory and Applications*, pp. 339–366, 1996.

[14] E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, 1997.

[15] J. L. Jerkins and J. L. Wang, "A measurement analysis of atm cell-level aggregate traffic," *Proceedings of IEEE Globecom '97*, pp. 1589–1595, November 1997.

[16] R. G. Addie, M. Zukerman, and T. D. Neame, "Fractal traffic: Measurements modelling and performance evalutation," *IEEE Infocom '95*, pp. 977–984, April 1995.

[17] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling.," *IEEE/ACM Transactions on Networking*, pp. 3(3):226–244,, 1995.

[18] E. E. Cohen, A. A. Fredericks, and C. D. Pack, "The internet and the public switched telephone network – a troubled marriage," *Proceedings of ITC 16*, pp. 23–32, 1999.

[19] V. A. Bolotin, Y. Levy, and D. Liu, "Characterizing data connection and messages by mixtures of distributions on logarithmic scale," *Proceedings of ITC 16*, pp. 887–894, June 1999.

[20] A. Downey, "Lognormal and pareto distributions in the internet," *Proceeding of Computer Communications*, pp. 790–801, 2005.

[21] T. Karagiannis, M. Molle, T. Karagiannis, M. Molle, and M. Faloutsos, "Long-range dependence-ten years of internet traffic modeling," tech. rep., October 2004.

[22] U. Sorger and Z. Suchanecki, "Network traffic modeling, i: Packets delay models.".

[23] U. Sorger and Z. Suchanecki, "Delays of interconnected flows," *International Journal of Theoretical Physics*, vol. 50, pp. 2242–2251, July 2011.

[24] S. Ahn, J. H. Kim, and V. Ramaswami, "A new class of models for heavy tailed distributions in finance and insurance risk," *In Insurance: Mathematics and Economics*, vol. 51, pp. 43–52, 2012.

[25] V. Ramaswami, K. Jain, and V. A. R. Jana, "Modeling heavy-tails in traffic sources for network performance evaluation," vol. 246, pp. 23–44, Springer India, 2014.

[26] J. C. Lopez-Ardao, C. Lopez-Garcia, A. Suarez-Gonzalez, M. Fernandez-Veiga, and R. Rodriguez-Rubio, "On the use of selfsimilar processes in network simulation," *ACM Transactions on Modeling and Computer Simulation*, vol. 10, no. 2, pp. 125–151.

[27] A. Lakhina, *Network-Wide Traffic Analysis: Methodes and Applications*. PhD thesis, 2007.

[28] M. Fras, J. Mohorko, and Ž. Čucej, "Estimating the parameters of measured self similar traffic for modeling," in *OPNET, IWSSIP*, (Maribor, Slovenia), pp. 27–30, June 2007.

[29] *Matjaž Fras, Jože Mohorko and Žarko Čučej (2012). Modeling and Simulating the Self-Similar Network Traffic in Simulation Tool, Telecommunications Networks - Current Status and Future Trends, Dr. Jesús Ortiz (Ed.),*

ISBN: 978-953-51-0341-7, InTech, DOI: 10.5772/37471. Available from: http://www.intechopen.com/books/telecommunications-networks-current-status-and-future-trends/modeling-and-simulating-the-self-similar-network-traffic-in-simulation-tool.

[30] J. Roberts, U. Mocci, and J. Virtamo, eds., *Broadband Network Teletraffic, Final Report of Action COST 242*. Springer, 1996.

[31] M. Schwartz, *Broadband Integrated Networks*. Prentice-Hall, 1996.

[32] M. F. Neuts, “Models based on the markovian arrival process,” *IEICE Transactions on Communications*, pp. 1255–1265, Dec 1992.

[33] S. H. Kang, Y. H. Kim, D. K. Sung, and B. D. Choi, “An application of markovian arrival process (map) to modeling superposed atm cell streams,” *IEEE Transactions on Communications*, vol. 50, pp. 633–642, April 2002.

[34] N. B. Shroff and M. Schwartz, “Improved loss calculations at an atm multiplexer,” *IEEE/ACM Transactions on Networking,*, vol. 6, pp. 411–421, August 1999.

[35] C. Blondia, “A discrete-time batch markovian arrival process as bisdn traffic model,” *Belgian Journal of Operation Research, Statistics and Computer Science*, vol. 32, no. 3,4, pp. 3–23, 1993.

[36] C. Herrmann, “Analysis of a discrete-time smp/d/1/s finite buffer queue with applications in atm,” *Proceedings of IEEE Infocom*, March 1993.

[37] O. Rose, "Discrete-time analysis of a finite buffer with vbr mpeg video traffic input," *Proceedings of ITC 15*, pp. 413–422, 1997.

[38] E.-J. Ang and J. Barria, "The markov modulated regulated brownian motion: A second-order fluid flow model of a finite buffer," *Queueing Systems*, pp. 263–287, 2000.

[39] M. E. Crovella, M. S. Taqqu, and A. Bestavros, "Heavy-tailed probability distributions in the world wide web transferred across the internet is indeed heavy-tailed.," *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, pp. 3–25, 1998.

[40] J. Judge, H. W. P. Beadle, and J. Chicharo, "Modeling world-wide web request traffic," *Proceedings of International Conference on Telecommunications*, pp. 257–262, April 1997.

[41] J. Judge, J. Chicharo, and H. W. P. Beadle, "The size of http response packets and calculation of www traffic volumes," *Proceeding of IS and T SPIE Multimedia Computing and Networking*, Feb 1997.

[42] K. Park, G. Kim, and M. Crovella., "On the effect of traffic self similarity on network performance," *Proceedings of the 1997 SPIE International Conference in Performance and Control of Net work Systems*, 1997.

[43] K. Park, G. Kim, and M. Crovella, "On the relationship between file sizes, transport protocols, and self-similar network traffic.," *Proceedings of the Fourth International Conference on Network Protocols*, October 1996.

[44] B. M. Williams, M.ASCE1, and F. Lester A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results," *JOURNAL OF TRANSPORTATION ENGINEERING*, pp. 664–672, 2003.

[45] B. M. Williams and L. A. Hoel, *MODELING AND FORECASTING VEHICULAR TRAFFIC FLOW AS A SEASONAL STOCHASTIC TIME SERIES PROCESS*. PhD thesis, University of Virginia, Charlottesville, 1999.

[46] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, no. 1, pp. 165–176, 1981.

[47] F. Xue, Y. S. J. Liu, L. Zhang, and O. Yang, "Traffic modeling based on farima models," *Proceedings of IEEE 1999 Canadian Conference on Electrical and Computer Engineering*, May 1999.

[48] F. Xue and L. Trajkovic, "Performance analysis of a wavelet-based hurst parameter estimator for self-similar traffic," *Proceedings of SPECTS 2K*, July 2000.

[49] M. S. Taqqu, S. Teverovsky, and W. Willinger, "Estimators for longrange dependence: An empirical study," *Fractals*, vol. 3, no. 4, pp. 785–798, 1995.

[50] I. Norros, "Lrd of fractional brownian motion and application in data network," *LRD of Fractional Brownian Motion and Application in Data Network*, vol. 13, pp. 953–962, 1995.

[51] F. Brichet, J. W. Roberts, A. Simonian, and D. Veitch, "Heavy traf- fic analysis of a storage model with long range dependent on/off sources.," *Queueing Systems*, vol. 23, pp. 197–216, 1996.

[52] T. Dieker, *Simulation of fractional Brownian motion*. PhD thesis, University of Twente Department of Mathematical Sciences, P.O. Box 94079 1090 GB Amsterdam The Netherlands, 2004.

[53] B. Yang, “Lrd of fractional brownian motion and application in data network.” 2008.

[54] K. MAJEWSKI, “Fractional brownian heavy traffic approximations of multiclass feedforward queueing networks,” *Queueing Systems*, vol. 50, pp. 199–230, 2005.

[55] J. Kim, “Weak convergence of various models to fractional brownian motion,” *Kangweon-Kyungki Math. Jour*, vol. 15, no. 1, pp. 71–78, 2007.

[56] V. R. J. R. Amitabha Ghosh, Rittwik Jana and N. K. Shankaranarayanan, “Modeling and characterization of large-scale wi-fi traffic in public hot-spots,” *In INFOCOM*, 2011.

[57] H. W. Blocka, Naftali, A. Langberg, and T. H. S. a, “The limiting failure rate for a convolution of gamma distributions,” *Statistics and Probability Letters*, July 9 2014.

[58] H. Block, N. Langberg, and T. Savits, “The limiting failure rate for a convolution of life distributions.” 2014.

[59] T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels, *Stochastic Processes for Insurance and Finance*. Wiley Series in Probability and Statistics, Wiley, Sept 2009.

[60] M. Finkelstein, *Failure Rate Modelling for Reliability and Risi*. Springer, November 2008.

[61] R. M. Adelson, “Compound poisson distribution,” *Operational Research Society*, vol. 17, pp. 73–75, Mar. 1966.

[62] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events*. Springer, 1997.

[63] T. Phit and K. Abe, “Packet inter-arrival time estimation using neural network models.”.

[64] F. Agharebparast and V. C. M. Leung, “A new traffic rate estimation and monitoring algorithm for the qos-enabled internet,” *Global TelecommunicationsConference 2003(GLOBECOM ’03.)*, vol. 7, pp. 3883–3887, Dec 2003.

[65] M. Grossglauser and D. Tse, “A framework for robust measurement based admission control,” *IEEE/ACM Trans. Networking*, vol. 7(3), pp. 293–309, June 1999.

[66] S. Floyd and V. Jacobson, “Link-sharing and resource management models for packet networks,” *IEEE/ACM Trans. Networking*, vol. 3, pp. 365–386, Aug 1995.

[67] F. Agharebparast and V. Leung, “Efficient fair queuing with decoupled delay-bandwidth guarantees,” *Proc. IEEE Globecom’01*, pp. 2601–2605, 2001.

[68] T. Kushida and Y. Shibata, “Empirical study of interarrival packet and packet losses,” *Proc. 22nd International Conference on Distributed Computing Systems Workshops(ICDCSW'02),*, p. 223, July 2002.

[69] M. Gupta, S. Glover, and S. Singh, “A feasibility study for power management in lan switches,” *Proc. 12nd International Conference on Network Protocols*, pp. 361–371, 2004.

[70] M. J. B. Błaszczyszyn and M. Karray, “How user throughput depends on the traffic demand in large cellular networks,” in *WiOpt - SpaSWiN*, 2014.

[71] S.-J. Lee and M. Gerla, “Split multipath routing with maximally disjoint paths in ad hoc networks,” vol. 10, pp. 3201–3205, June 2001.

[72] A. Nasipuri, R. Castaneda, and S. R. Das, “Performance of multipath routing for on-demand protocols in mobile ad hoc networks,” *Mobile Networks and Applications*, vol. 6, pp. 339–349, Aug.

[73] M. K. Marina and S. R. Das, “On-demand multi path distance vector routing in ad hoc networks,” *Proc. ICNP 2001*, pp. 14–23, Nov.

[74] “Multipath routing in the presence of frequent topological changes,” *IEEE Communications Magazine*, vol. 39, pp. 132–138, Nov.

[75] H. Lim, K. Xu, and M. Gerla, “Tcp performance over multipath routing in mobile ad hoc networks,” *Proc. ICC 2003*, vol. 2, pp. 1064–1068, May 2003.

[76] A. Tsirigos and Z. J. Haas, “Analysis of multipath routing,” *IEEE Trans. Wireless Communications*, vol. 3, pp. 138–146, Jan 2004.

[77] J. He and J. Rexford, “Towards internet-wide multipath routing.” Towards Internet-wide Multipath Routing.

[78] V. Petrov, *Sums of Independent Random Variables*. Springer, softcover reprint of the original 1st ed. 1975 ed., October 2011.

[79] A. Hatcher, *Algebraic Topology*. December 2001.

[80] D. R. Figueiredo and D. T. B. Liu, V. Misra, “On the autocorrelation structure of tcp traffic,” *Daniel R. Figueiredo, , Benyuan Liu , Vishal Misra , Don Towsley*, vol. 40, pp. 339–361, October 22 2002.

[81] S. Manthorpe, I. Norros, and J. Y. L. Boudec, “The second-order characteristics of tcp,” *Proc. of Performance ’96*, 1996.

[82] E. Lukacs, *Characteristic Functions*. Hodder Arnold, January 1976.

[83] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[84] M. M. Meerschaert and H. Scheffler, *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. Wiley Series in Probability and Statistics, Wiley-Interscience, July 2001.

[85] G. G. Hamedani, “Why independence when all you need is sub-independence,” *Journal of Statistical Theory and Applications*, 2003.

[86] S. Asmussen, *Ruin Probabilities*. Advanced Series on Statistical Science and Applied Probability, World Scientific Publishing Company, September 2010.

[87] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2. 1971.

[88] N. Champagnat, M. Deaconu, A. Lejay, N. Navet, and S. Boukheroua, "An empirical analysis of heavy-tails behavior of financial an empirical analysis of heavy-tails behavior of financial data: The case for power laws." Aug 14 Aug 14.

[89] E. properties of asset returns: stylized facts and statistical issues, "Quantitative finance," pp. 223–236, 2001.

[90] X. Gabaix, "Power laws in economics and finance," *Annual Review of Economics*, pp. 55–93, June 2009.

[91] J. M. Peha, "Retransmisstion mechanisms and self-similar traffic models," *Proc. of IEEE/ACM/SCS Communications Network and Distributed Systems Modeling and Simulation*, pp. 48–52, Jan 1997.

[92] J. Zimmermann, A. Clacrk, and G. Mohay, "The use of packet inter-arrival times for investigating unsolicited internet traffic," *Proc. The first International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05),*, pp. 89–104, Nov 2005.

[93] J. D.Hamilton, *Time Series Analysis*. Princeton University Press, 1994.

[94] B. George and J. Gwilym, *Time Series Analysis: forecasting and control*. 1976.

[95] O. Bradley and M. S. taqqu, "Financial risk and heavy tail," *Heavy-tailed distributions in Finance*, 2002.

[96] R. V. Bjorn Sundt, *Recursions for Convolutions and Compound Distributions with Insurance Applications*. Springer, 2009.

[97] M. Einsiedler, *Ergodic Theory: with a view towards Number Theory*. Graduate Texts in Mathematics, Springer, 1sr ed., September 2010.

[98] P. R. Halmos, *Lectures on Ergodic Theory*. Martino Fine Books, 2013.

[99] K. E. Petersen, *Ergodic Theory*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1989.

[100] E. Zeidler, *Quantum Field Theory 1-3*. Springer, 2009.

[101] N. Dunford and J. T. Schwatz, *Linear Operators 1-3*. Wiley Classics Library, Wiley-Interscience, 1988.

[102] T. Szarek, "The stability of markov operators on polish space," *Studia Math.*, pp. 145–152, 2000.

[103] D. Pollard, *Convergence of Stochastic Processes*. Springer Series in Statistics, October 2011.

[104] P. Billingsley, *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, Wiley-Interscience, 2 ed., 1968.

[105] F. W. Steutel and K. van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*. CRC Press, October 2003.

[106] R. D. Prasad, "Probability distributions of algebraic functions of independent random variables," *SIAM Journal on Applied Mathematics*, vol. 18, pp. 614–626, May 1970.

[107] G. Samoradnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. 1994.

[108] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. No. 97 in Applied Mathematical Sciences, Springer, 1998.