NETWORK TRAFFIC MODELING USING HEAVY TAIL DISTRIBUTIONS

by

RAKESH SINGHAI

Electrical Engineering Department

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the

Department of Electrical Engineering

Indian Institute of Technology, Delhi

August, 2009

Dedicated to

Jyoti, Adwait and Aishwarya

Certificate

This is to certify that the thesis entitled **"Network Traffic Modeling Using Heavy Tail Distributions"** being submitted by Rakesh Singhai to the Department of Electrical Engineering, Indian Institute of Technology, Delhi, for the award of the degree of Doctor of Philosophy is the record of the bona-fide research work carried out by him under our supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other university or institute for the award of any degree or diploma.

(**Dr. Rajendra K. P. Bhatt**)
Professor
Electrical Engineering Department
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi 110016

(**Dr. Shiv Dutt Joshi**)
Professor
Electrical Engineering Department
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi 110016

Acknowledgment

I would like to express my deepest sense of gratitude to my supervisors, Prof. Rajendra K. P. Bhatt and Prof. Shiv Dutt Joshi, for their invaluable help and guidance during the course of project. They continually and convincingly conveyed a spirit of adventure in regard to research. Without their guidance and persistent help this dissertation would not have been possible. I would like to thank my Student Research Committee members, Professor H. M. Gupta, Dr. Shankar Prakriya and Dr Dharmraja, whose support, remarks and directions demonstrated to me to work hard.

I would also like to thank my colleagues Dr V. V. Mani, Mr. M. D. Selvaraj, Mr. D. Ravindra, Mr Mayur Vegad, Dr. Sanjay Nalbalwar, Dr. K. K. Sharma and Dr. Lalit Jiwani for their lively company and the suggestion in our technical discussions. Special thanks are due to Mr. M. K. Kaushik, Technical Superintendent, and Mr. Mahesh Kumar, Secretary BSTTM, for providing a conducive environment in the Wireless Communication laboratory, where I conducted most of my research work. I want to put on record my appreciations and thanks to all the faculty members, staff and friends of Electrical Engineering Department.

In addition, a big thank to my *first teachers* who introduced me with this beautiful world of knowledge.

My sincere thanks to my old parents whose eyes always wished me to acquire the highest degree from a world class institution like IIT Delhi. Their love and affection always inspired me. I am thankful to my father-in-law (late) Prof. J. K. Jain for his initial support, guidance and encouragement. My sincere thanks to my mother-in-law Mrs. Manju Jain for supporting and taking care of my family in my absence.

Words are not enough to express my gratitude to my wife, Dr. Jyoti Singhai, for her constant inspiration and encouragement to take up this assignment at the fullest and to my children, Adwait and Aishwarya, to whom I want to dedicate this work for much they suffered during my involvement in this project. This work would not have been possible without their constant support, understanding and patience.

I wish to thank my friend Mr. Basant Chourasia for his help during the whole course of my doctoral research work at IIT Delhi.

I thank the authorities of Government of Madhya Pradesh and Government of India for sponsoring me to pursue my Ph.D. from IIT Delhi. I thank to authorities of IIT Delhi to provide me financial support, quality education and excellent educational facility.

Finally, I would like to thank all those who helped me directly or indirectly in completion of this research work.

Rakesh Singhai

August 2009

Indian Institute of Technology, Delhi

Abstract

Packet arrival processes, for data networks, are not consistent with Poisson processes. It has been established that network data arrival process is self-similar and the number of TCP connections, initiated by a user session, is heavy-tailed. The heavy tail distributions yield better model for data packet arrival models than exponential models. In particular it is observed that the Weibull and Pareto distributions give a good fit for the interarrival-times for Web applications. The evidences, as reported by many authors, that packet data arrive in bursts and exhibit long tail distribution, have prompted many researchers to investigate the whole class of heavy tailed distributions in general.

The entire development of the thesis is based on the observation that modeling of interarrival time, in data packet network, as exponential is not appropriate. This is because this distribution assigns maximum probability of next arrival to the time $t = 0$, when a packet has just arrived, which is an unrealistic assumption. A more realistic modeling would be the one whose peak of distribution is shifted away from zero. In this setting we have proposed a new family of distributions (a Generalized Family of Distributions) which has Weibull and Pareto distributions as subfamilies. In this work it is shown, if packet interarrival times are modeled according to a Generalized Family of Distributions (aGFD), than the corresponding counting process is not a homogeneous Poisson process. We have introduced a specific time dependent arrival rate of packets which can model flexible arrivals of packets from very small interarrival time to large interarrival time. We show that the corresponding counting process for this time dependent arrival rate is non-homogeneous Poisson process (NHPP). It is also shown that the counting

process for Weibull arrivals is a special case of this NHPP and first interarrival time in this process is Weibull distributed and other interarrival times are Weibull-like distributed. We also propose a new time dependent arrival rate for TCP and Telnet packet interarrival times. We show that this time dependent arrival rate gives a novel counting counting process. We also show that this counting process has Pareto distributed first interarrival time and conditional distribution of other interarrivals is Pareto-like. This counting process results when a new time dependent arrival rate, which is suited for modeling LRD data packets, is considered. We can see that this process is NHPP. We also propose an approximate formula for the packet arrival rate for Lognormal distributed interarrival times. This time dependent arrival rate also results in NHPP. Thus the mean packet arrival rate, for the class of heavy-tail distributed interarrival-times, is not constant but rather time-varying and constitute non-homogenous Poisson process. Despite the fact that packet arrival process for data network are not consistent with Poisson process, it is interesting to note that for heavy-tail interarrivals the NHPP traffic model exists. It shows that Poisson nature of data packet arrivals can't be ruled out and the mean arrival rates are now time varying.

Since the proposed time dependent arrival rates, mentioned above, lead to results which are consistent with practically observed traffic, it is pertinent to carry out the corresponding queueing analysis. Heavy tail distributions do not lend themselves to easy queuing analysis since their Laplace transforms are not explicitly available in general. Thus most of the queuing results of heavy tail distributions are obtained for asymptotic regions. We take heavy-tailed, time-varying models for the queueing analysis. We compute blocking probability for $M_t/M/1/c$ time-varying queueing model, by using Fixed Point

Algorithm (FPA). We also extend this result for finding the loss probability in $M_t/M/c/c$ queueing model for Weibull and Pareto distributed time-varying interarrival times. We also determine expressions for offered-load in $M_t/G/\infty$ time-varying queueing models for Weibull and Pareto interarrivals.

Contents

1 Introduction **1**

1.1 Literature Survey 3

1.1.1 Non Self-Similar Traffic Model 4

1.1.2 Self-Similarity & the Associated Traffic Models 7

1.1.3 Time-Varying Queueing Analysis 11

1.2 Motivation And Problem Statement 12

1.2.1 Problem Statement 14

1.3 Major Contribution of the Thesis 14

1.4 Organization of the Thesis 16

2 Mathematical Preliminaries **21**

2.1 Introduction 21

2.2 Poisson process 22

2.3 Fractal Processes 25

2.3.1 Self-similar processes 25

2.3.1.1 Continuous time self-similarity 26

2.3.1.2 Discrete time self-similarity 29

2.3.1.3 Impact of self-similarity 31

2.3.2 Long-Range and Short-Range Dependence 31

2.3.3 $1/f$ Process . . . 33
2.4 Multifractals . . . 35
2.4.1 Infinitely Divisible Cascades . . . 36
2.5 ON-OFF Process . . . 37
2.6 A Brief Overview of Some Important Heavy-Tail Distributions 40
2.6.1 Weibull Distribution . . . 40
2.6.2 Pareto Distribution . . . 40
2.6.3 Lognormal Distribution . . . 43
2.7 A Brief Introduction To Queueing Theory . . . 43
2.7.1 Birth And Death Processes . . . 45
2.7.1.1 The $M/M/1/\infty$ Queue . . . 47
2.7.1.2 System with Limited Waiting Space: The $\mathbf{M/M/1/c}$ Model . . . 47
2.8 Emergence of a New Trend in Teletraffic Theory . . . 49
2.8.1 $M/G/1$ Queue With Heavy-Tailed Service Time . . . 50
2.8.1.1 Pareto Mixture of Exponential Distribution [1] 50
2.8.1.2 Gamma Mixture of Pareto Distribution [13] . 51
2.9 Packets and TCP/IP . . . 51
2.9.1 Internet-Type Queues with Power-Tailed Interarrival Times . . . 54
2.9.2 Analysis of $M_t/M/1/c$ Finite Storage Queueing Model 54
2.10 Conclusions . . . 58
3 aGFD Arrivals **59**
3.1 Introduction . . . 59
3.2 Nonhomogeneous Poisson Process Traffic Model . . . 61

3.3 Heavy Tail Arrivals: A Generalization 62
3.3.1 Heavy Tail Distributions 63
3.3.2 A New Family of Distributions 64
3.4 What Rate Function Makes Heavy-Tail? 65
3.5 Counting Process For aGFD Arrivals 68
3.6 Discussion on NHPP Arrival Packets 71
3.6.1 Motivation 72
3.7 Counting Process for Weibull Arrival Points 73
3.7.1 Distribution of First Interarrival Time 74
3.7.2 Validation 75
3.7.3 Distribution of 'Waiting Time Until The Sum of nth Event' 76
3.7.4 Conditional Distribution of Interarrival Times Given First Arrival Weibull Distributed 77
3.8 Conclusions 79

4 Analysis of Pareto and Lognormal Interarrivals **81**
4.1 Introduction 81
4.2 Motivation 83
4.3 NHPP Network Traffic Model 84
4.3.1 Distribution of First Interarrival Time 91
4.3.2 Validation 92
4.3.3 Distribution of 'Waiting Time Until The Sum of nth Event' 93
4.3.4 Conditional Distribution of Interarrival Times Given First Arrival Pareto Distributed 94

4.4 Counting Process for Lognormal Interarrivals 97

4.5 Conclusions 101

5 Queueing Analysis for Weibull And Pareto Interarrivals 103

5.1 Introduction 104

5.2 Queueing Theory: General Concepts 105

5.3 Traffic Model 106

5.3.1 Nonhomogeneous Poisson Process Traffic Model 106

5.4 Busy Server model 107

5.4.1 The Analysis of $M_t/G/\infty$ Queue 110

5.4.2 Weibull Arrival Rate Function 113

5.4.3 Pareto Arrival Rate Function 115

5.5 Loss Queue Model Using FPA 118

5.5.0.1 FPA Algorithm for $M_t/M/1/c$ Queue 120

5.5.0.2 Results of FPA Algorithm For $M_t/M/1/c$ Queue 123

5.5.1 The Analysis of $M_t/M/c/c$ Queue 124

5.5.1.1 FPA Algorithm for $M_t/M/c/c$ Queue 132

5.5.1.2 Results of FPA Algorithm For $M_t/M/c/c$ Queue 134

5.6 Conclusions 134

6 Conclusions and Future Work 141

6.1 Salient Conclusions 141

6.2 Scope of Future Work 143

A Hazard Function 145

B Counting Process for Weibull Arrival Points 147