A Characterization of the Variability of Packet Arrival Processes in Workstation Networks

by

Riccardo Gusella

Copyright © 1990
by
Riccardo Gusella

This thesis was typeset in Avantgarde and Courier on a DECstation 5000 using troff. All figures, except for Figure 4.3 which was prepared with Mathematica, were produced using grap.

Acknowledgments

I have been most fortunate in having Domenico Ferrari as my research advisor. His constant scientific and moral support have been the source of inspiration in all my work. Domenico's guidance during difficult moments, when the solution to the most challenging problems seemed out of reach, steered my work into direction. From him I have learned that in ``Computer Science'' one should stress the second word.

I am deeply indebted to Sam Morgan, who supervised my work when I was at Bell Labs, Murray Hill. Sam has been exceptional; his intellectual rigor has been a source of inspiration. It was Sam who suggested the research topic that led to the development of this thesis. He served on my thesis committee, and provided me with the most detailed assessment of the weaknesses in my thesis draft. His constant support and never-ending encouragement have helped me overcome the most perplexing of obstacles.

My thesis work would not have been possible without the direct involvement of a terrific group of people at Sun Microsystems. Marty Itzkowitz and Mark Liu went out of their way to help me collect the measurement data, and provided invaluable technical support. I am also deeply grateful to Chris Rigatuso and Marilyn Shoemaker for all their assistance.

I am indebted to David Anderson for serving as Quals Committee Chairman and for his comments on the thesis manuscript. I am grateful to David Aldous for serving on my Quals Committee, and to David Brillinger, who read the thesis draft, for his invaluable advice.

I would like to thank William Kahan and Ward Whitt.

I owe much to many friends who, directly or indirectly, have influenced my work. I will not attempt a complete list, for I would surely forget someone. Instead, I collectively thank all the members of the Tenet group, and all my friends in the Computing Science Research group at Bell Labs. But I would like to acknowledge especially Ramon Caceres, Peter Danzig, Kerry Fendick, Kathy Meier-Hellstern, Keshav, Bart Miller, Joe Pasquale, Dave Presotto, Doug Terry, Shin-Yuan Tzou, Dinesh Verma, Stefáno Zatti, Songnian Zhou.

My gratitude goes to Kathryn Crabtree, who assisted me in navigating the Computer Science Division bureaucratic apparatus.

I would like to express my gratitude to Phyllis for having helped me to cope with a thesis that at times looked so frustrating. To her I dedicate this work.

My research has been supported by a number of research grants. I kindly acknowledge the support of AT&T Bell Laboratories; Hitachi Ltd.; Olivetti S.p.A.; the University of California, Berkeley, under a Micro grant; and the International Computer Science Institute.

Abstract

We examine the problem of characterizing the variability of measured packet-arrival processes produced by individual workstations in a local-area network. Commonly referred to as burstiness, variability can be informally described as producing sequences of abrupt transitions from low to high arrival rates and vice versa. Since variability can be related to the relationship among successive arrivals, we adopt a quantitative definition, based on indices of dispersion, from the theory of point processes.

We illustrate the measurement methodology and discuss the error analysis. We then analyze the first- and second-order statistical properties of interarrival-time and packet-count series, which reveal the structure of the underlying point processes. We estimate indices of dispersion for intervals and counts, which express the autocorrelation structure of a point process, and warn about the effect of nonstationary data. Using an artificial example based on the Markov-modulated Poisson process, we show how to incorporate into a mathematical model the second-order stochastic parameters that represent dispersion. Fitting is done so that the index of dispersion for counts of the MMPP model matches closely that of the data, a procedure that produces what we call a ``model of variability''.

Finally, we derive a model of variability whose structure follows the structure of the data: the interarrival times of short and long packets are disjoint; the lengths of sequences of short and long packets form a discrete-time Markov chain; and a generalized two-state semi-Markov process, in which interarrival times in each of the states are autocorrelated, is shown to reproduce with good approximation the correlation structure of the data for time scales up to 500 ms. The approximation requires only estimates of the first- and second-order moments of the interarrival times. To complete the model, which is two-dimensional, we also provide simple characterizations for the lengths of short and long packets. Because of the recursive nature of the model's equations, the model is suited for simulation studies.

Table of Contents

Chapter 1: Modeling Variability

I. Introduction 1

II. Central Problem of Dissertation 2

III. Thesis Roadmap 3

IV. Literature Review 6

A. Network Measurements and Their Analysis 6

B. Arrival Process Modeling 7

Chapter 2: Traffic Measurements

I. Introduction 11

II. Berkeley Measurements 12

III. Sun Measurements 13

IV. The Ethernet Driver 15

V. Packet Traces 16

VI. Data Compression 17

VII. Timing Errors 19

VIII. The Sun Engineering Network 27

IX. Summary 33

Chapter 3: Traffic Patterns of Individual Workstations

I. Introduction 35

II. The Network Load of Individual Workstations 36

III. Busy and Idle Phases 38

IV. Histograms of Interarrival Times 39

V. Stationarity of Packet Arrival Processes 47

VI. The Busy Periods 50

VII. Autocorrelation Coefficient Series 50

VIII. Concurrency on the Network 55

IX. Statistical Interactions Among Workstations 57

X. Summary 62

Chapter 4: Indices of Dispersion
I. Introduction 67
II. Indices of Dispersion 69
A. The Index of Dispersion for Intervals 69
B. The Index of Dispersion for Counts 70
C. IDI for Processes with Hyperexponential Interarrival Times 71
D. IDC for Batch Poisson Processes 72
E. IDC for Markov-Modulated Poisson Processes 73
III. Estimated Indices of Dispersion 75
A. Estimated Indices of Dispersion for Intervals 75
B. Estimated Indices of Dispersion for Counts 78
IV. Fitting a 2-State MMPP to an Arrival Process 81
V. Summary 84

Chapter 5: Models of Variability
I. Introduction 85
II. Models for Packet Sequences 86
III. A Generalized Semi-Markov Process 93
IV. Data Fitting and Analysis 96
V. Short-Interval Approximations 99
VI. Models of Packet Lengths 101
VII. Summary 103

Chapter 6: Conclusions
I. Introduction 105
II. Future Work 106
III. Epilogue 106

Appendix A: Analysis of Berkeley Measurements
I. Introduction 107
II. General Statistics 108
A. Network Utilization 109
B. Packet Length 110
C. Network Protocols 111
D. Intranetwork Traffic and Internetwork Traffic 112
E. Interarrival Time 114
F. Contention 116

III. The Transmission Control Protocol 116
A. Data Size 117
B. Connections 117
C. Higher-Level Protocols 118
D. Source-Destination Patterns 119
E. Interarrival Times 120
IV. The Network Disk Protocol 121
A. Utilization 121
B. Interarrival Time 123
V. The Network File System Protocol 124
A. Utilization 125
B. Packet Length 125
C. NFS Procedure Statistics 126
D. Protocol Timers 127
E. Interarrival Time 128
VI. Conclusions 130

References

References 133