

GENERATION AND VALIDATION OF EMPIRICALLY-DERIVED TCP  
APPLICATION WORKLOADS

Félix Hernández-Campos

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill  
2006

Approved by:  
Advisor: Kevin Jeffay  
Reader: F. Donelson Smith  
Reader: Ketan Mayer-Patel  
Reader: J. Steve Marron  
Reader: Andrew Nobel  
Reader: Jan Prins

© 2006  
Félix Hernández-Campos  
ALL RIGHTS RESERVED

# ABSTRACT

FÉLIX HERNÁNDEZ-CAMPOS: Generation and Validation of  
Empirically-Derived TCP Application Workloads.  
(Under the direction of Kevin Jeffay)

This dissertation proposes and evaluates a new approach for generating realistic traffic in networking experiments. The main problem solved by our approach is generating closed-loop traffic consistent with the behavior of the entire set of applications in modern traffic mixes. Unlike earlier approaches, which described individual applications in terms of the specific semantics of each application, we describe the source behavior driving each connection in a generic manner using the a-b-t model. This model provides an intuitive but detailed way of describing source behavior in terms of connection vectors that capture the sizes and ordering of application data units, the quiet times between them, and whether data exchange is sequential or concurrent. This is consistent with the view of traffic from TCP, which does not concern itself with application semantics.

The a-b-t model also satisfies a crucial property: given a packet header trace collected from an arbitrary Internet link, we can algorithmically infer the source-level behavior driving each connection, and cast it into the notation of the model. The result of packet header processing is a collection of a-b-t connection vectors, which can then be replayed in software simulators and testbed experiments to drive network stacks. Such a replay generates synthetic traffic that fully preserves the feedback loop between the TCP endpoints and the state of the network, which is essential in experiments where network congestion can occur. By construction, this type of traffic generation is fully reproducible, providing a solid foundation for comparative empirical studies.

Our experimental work demonstrates the high quality of the generated traffic, by directly comparing traces from real Internet links and their source-level trace replays for a rich set of metrics. Such comparison requires the careful measurement of network parameters for each connection, and their reproduction together with the corresponding source behavior. Our final contribution consists of two resampling methods for introducing controlled variability in network experiments and for generating closed-loop traffic that accurately matches a target offered load.

## ACKNOWLEDGMENTS

First of all, I must thank Kevin Jeffay and Don Smith for their guidance and encouragement throughout my doctoral program. Their patience and friendship have been invaluable all these years. I also thank them, together with other faculty and student members of the Distributed and Real-Time Systems group (DiRT), for building a phenomenal infrastructure for Internet measurement and experimental networking research. DiRT students have greatly contributed to my doctoral experience, most especially Jay Aikat and David Ott.

My committee members and other collaborators have contributed tremendously to my efforts. I am specially in debt with Steve Marron and Andrew Nobel, who have greatly enriched the statistical side of my work. In this regard, being part of SAMSI's "Network Modeling for the Internet" program and of the inter-disciplinary Internet study group at UNC gave me superb opportunities to widen my understanding of Internet research. I must also thank UNC's Department of Computer Science as whole, including faculty, students and staff, for creating an outstanding research and teaching environment. Overall, my years at UNC were an incredibly positive experience.

I thank the National Science Foundation, IBM, Cisco, Intel, Sun Microsystems and others for supporting this work. I am specially grateful to the Computer Measurement Group (CMG) for their doctoral fellowship.

Finally, I thank my family for their support. Their constant example of hard-work, and their respect for intellectual endeavors has motivated me during my entire life. My wife's help with the editing of this manuscript was invaluable, as was her constant encouragement during my graduate studies. More than anybody else, my parents gave me my passion for knowledge, and it is to them that I dedicate this doctoral dissertation.

# TABLE OF CONTENTS

|   |           |
|---|-----------|
| <b>LIST OF TABLES</b>                                 | <b>ix</b> |
| <b>LIST OF FIGURES</b>                                | <b>x</b>  |
| <b>LIST OF ABBREVIATIONS</b>                          | <b>xx</b> |
| <b>1 Introduction</b>                                 | <b>1</b>  |
| 1.1 Abstract Source-Level Modeling . . . . .          | 3         |
| 1.2 Source-Level Trace Replay . . . . .               | 10        |
| 1.3 Trace Resampling and Load Scaling . . . . .       | 13        |
| 1.4 Thesis Statement . . . . .                        | 15        |
| 1.5 Contributions . . . . .                           | 15        |
| 1.6 Overview . . . . .                                | 17        |
| <b>2 Related Work</b>                                 | <b>19</b> |
| 2.1 Packet-Level Traffic Generation . . . . .         | 20        |
| 2.2 Source-Level Traffic Generation . . . . .         | 23        |
| 2.2.1 Web Traffic Modeling . . . . .                  | 27        |
| 2.2.2 Non-Web Traffic Source-level Modeling . . . . . | 30        |
| 2.2.3 Beyond Single Application Modeling . . . . .    | 32        |
| 2.3 Scaling Offered Load . . . . .                    | 33        |
| 2.4 Implementing Traffic Generation . . . . .         | 35        |
| 2.5 Summary . . . . .                                 | 35        |
| <b>3 Abstract Source-level Modeling</b>               | <b>38</b> |
| 3.1 The Sequential a-b-t Model . . . . .              | 40        |
| 3.1.1 Client/Server Applications . . . . .            | 40        |
| 3.1.2 Beyond Client/Server Applications . . . . .     | 48        |
| 3.2 The Concurrent a-b-t Model . . . . .              | 50        |

|          |  |            |
|----------|--|------------|
| 3.3      | Abstract Source-Level Measurement . . . . .                    | 53         |
| 3.3.1    | From TCP Sequence Numbers to Application Data Units . . . . .  | 53         |
| 3.3.2    | Logical Order of Data Segments . . . . .                       | 56         |
| 3.3.3    | Data Analysis Algorithm . . . . .                              | 59         |
| 3.4      | Validation using Synthetic Applications . . . . .              | 65         |
| 3.5      | Analysis Results . . . . .                                     | 69         |
| 3.5.1    | Variability Across Sites . . . . .                             | 72         |
| 3.5.2    | Time-of-Day Variability and Workload Directionality . . . . .  | 80         |
| 3.6      | Summary . . . . .  | 83         |
| <b>4</b> | <b>Network-Level Parameters and Metrics</b>                    | <b>87</b>  |
| 4.1      | Network-level Parameters . . . . .                             | 88         |
| 4.1.1    | Round-Trip Time . . . . .                                      | 88         |
| 4.1.2    | Receiver Window Size . . . . .                                 | 102        |
| 4.1.3    | Loss Rate . . . . .  | 104        |
| 4.2      | Network-level Metrics . . . . .                                | 108        |
| 4.2.1    | Aggregate Throughput Time Series . . . . .                     | 109        |
| 4.2.2    | Throughput Marginals . . . . .                                 | 118        |
| 4.2.3    | Throughput Self-Similarity and Long-Range Dependence . . . . . | 125        |
| 4.2.4    | Time Series of Active Connections . . . . .                    | 132        |
| 4.3      | Summary . . . . .  | 135        |
| <b>5</b> | <b>Generating Traffic</b>                                      | <b>139</b> |
| 5.1      | Replaying Traces at the Source-Level . . . . .                 | 139        |
| 5.1.1    | Trace Partitioning . . . . .                                   | 141        |
| 5.1.2    | Conducting Experiments . . . . .                               | 142        |
| 5.1.3    | Data Collection . . . . .                                      | 145        |
| 5.2      | Validation of Source-level Trace Replay . . . . .              | 146        |
| 5.2.1    | Leipzig-II . . . . .   | 146        |
| 5.2.2    | UNC 1 PM . . . . .   | 154        |
| 5.2.3    | Abilene-I . . . . .  | 158        |
| 5.3      | Summary . . . . .  | 162        |

|          |   |            |
|----------|---|------------|
| <b>6</b> | <b>Reproducing Traffic</b>                    | <b>164</b> |
| 6.1      | Beyond Comparing Connection Vectors . . . . . | 164        |
| 6.2      | Source-level Replay of Leipzig-II . . . . .   | 168        |
| 6.2.1    | Time Series of Byte Throughput . . . . .      | 168        |
| 6.2.2    | Time Series of Packet Throughput . . . . .    | 171        |
| 6.2.3    | Marginal Distributions . . . . .              | 173        |
| 6.2.4    | Long-Range Dependence . . . . .               | 177        |
| 6.2.5    | Time Series of Active Connections . . . . .   | 182        |
| 6.3      | Source-level Replay of UNC 1 PM . . . . .     | 183        |
| 6.3.1    | Time Series of Byte Throughput . . . . .      | 183        |
| 6.3.2    | Time Series of Packet Throughput . . . . .    | 185        |
| 6.3.3    | Marginal Distributions . . . . .              | 186        |
| 6.3.4    | Long-Range Dependence . . . . .               | 189        |
| 6.3.5    | Time Series of Active Connections . . . . .   | 192        |
| 6.4      | Mid-Chapter Review . . . . .                  | 194        |
| 6.4.1    | Observations on Byte Throughput . . . . .     | 194        |
| 6.4.2    | Observations on Packet Throughput . . . . .   | 196        |
| 6.4.3    | Observations on Active Connections . . . . .  | 197        |
| 6.5      | Source-level Replay of UNC 1 AM . . . . .     | 197        |
| 6.5.1    | Time Series of Byte Throughput . . . . .      | 197        |
| 6.5.2    | Time Series of Packet Throughput . . . . .    | 199        |
| 6.5.3    | Marginal Distributions . . . . .              | 200        |
| 6.5.4    | Long-Range Dependence . . . . .               | 202        |
| 6.5.5    | Time Series of Active Connections . . . . .   | 205        |
| 6.6      | Source-level Replay of UNC 7:30 PM . . . . .  | 205        |
| 6.6.1    | Time Series of Byte Throughput . . . . .      | 205        |
| 6.6.2    | Time Series of Packet Throughput . . . . .    | 207        |
| 6.6.3    | Marginal Distributions . . . . .              | 208        |
| 6.6.4    | Long-Range Dependence . . . . .               | 210        |
| 6.6.5    | Time Series of Active Connections . . . . .   | 213        |
| 6.7      | Source-level Replay of Abilene-I . . . . .    | 214        |
| 6.7.1    | Time Series of Byte Throughput . . . . .      | 214        |

|          |  |            |
|----------|--|------------|
| 6.7.2    | Time Series of Packet Throughput . . . . .                 | 215        |
| 6.7.3    | Marginal Distributions . . . . .                           | 216        |
| 6.7.4    | Long-Range Dependence . . . . .                            | 218        |
| 6.7.5    | Time Series of Active Connections . . . . .                | 218        |
| 6.8      | Summary . . . . .  | 221        |
| <b>7</b> | <b>Trace Resampling and Load Scaling</b>                   | <b>223</b> |
| 7.1      | Poisson Resampling . . . . .                               | 226        |
| 7.1.1    | Basic Poisson Resampling . . . . .                         | 226        |
| 7.1.2    | Byte-Driven Poisson Resampling . . . . .                   | 231        |
| 7.2      | Block Resampling . . . . .                                 | 236        |
| 7.3      | Summary . . . . .  | 245        |
| <b>8</b> | <b>Conclusions and Future Work</b>                         | <b>247</b> |
| 8.1      | Empirical Modeling of Traffic Mixes . . . . .              | 248        |
| 8.2      | Refining and Extending our Modeling . . . . .              | 250        |
| 8.3      | Assessing Realism in Synthetic Traffic . . . . .           | 252        |
| 8.4      | Incorporating Additional Network-Level Parameter . . . . . | 254        |
| 8.5      | Flexible Traffic Generation . . . . .                      | 255        |
|          | <b>BIBLIOGRAPHY</b>  | <b>257</b> |



## LIST OF TABLES

|      |   |     |
|------|---|-----|
| 3.1  | Breakdown of the TCP connections found in five traces. . . . .  | 69  |
| 4.1  | Estimated Hurst parameters and their confidence intervals for the packet throughput time series of five traces. . . . .   | 131 |
| 4.2  | Estimated Hurst parameters and their confidence intervals for the byte throughput time series of five traces. . . . .   | 131 |
| 6.1  | Estimated Hurst parameters and their confidence intervals for the byte throughput time series of Leipzig-II and its four types of source-level trace replay. . . . .    | 178 |
| 6.2  | Estimated Hurst parameters and their confidence intervals for the packet throughput time series of Leipzig-II and its four types of source-level trace replay. . . . .  | 181 |
| 6.3  | Estimated Hurst parameters and their confidence intervals for the byte throughput time series of UNC 1 PM and its four types of source-level trace replay. . . . .      | 190 |
| 6.4  | Estimated Hurst parameters and their confidence intervals for the packet throughput time series of UNC 1 PM and its four types of source-level trace replay. . . . .    | 193 |
| 6.5  | Estimated Hurst parameters and their confidence intervals for the byte throughput time series of UNC 1 AM and its four types of source-level trace replay. . . . .      | 203 |
| 6.6  | Estimated Hurst parameters and their confidence intervals for the packet throughput time series of UNC 1 AM and its four types of source-level trace replay. . . . .    | 204 |
| 6.7  | Estimated Hurst parameters and their confidence intervals for the byte throughput time series of UNC 7:30 PM and its four types of source-level trace replay. . . . .   | 211 |
| 6.8  | Estimated Hurst parameters and their confidence intervals for the packet throughput time series of UNC 7:30 PM and its four types of source-level trace replay. . . . . | 212 |
| 6.9  | Estimated Hurst parameters and their confidence intervals for the byte throughput time series of Abilene-I and its four types of source-level trace replay. . . . .     | 219 |
| 6.10 | Estimated Hurst parameters and their confidence intervals for the packet throughput time series of Abilene-I and its four types of source-level trace replay. . . . .   | 220 |
| 7.1  | Estimated Hurst parameters and their confidence intervals for the connection arrival time series of UNC 1 PM and UNC 1 AM, and their Poisson arrival fits. . . . .      | 235 |
| 7.2  | Estimated Hurst parameters and their confidence intervals for five subsamplings obtained from the connection arrival time series of UNC 1 PM and UNC 1 AM . . . . .     | 243 |

## LIST OF FIGURES

|      |  |    |
|------|--|----|
| 1.1  | Network traffic seen from different levels. . . . .  | 4  |
| 1.2  | An a-b-t diagram illustrating a persistent HTTP connection. . . . .  | 6  |
| 1.3  | A diagram illustrating the interaction between two BitTorrent peers. . . . .   | 9  |
| 1.4  | Overview of Source-level Trace Replay. . . . .   | 10 |
| 3.1  | An a-b-t diagram representing a typical ADU exchange in HTTP version 1.0. . . . .                                    | 40 |
| 3.2  | An a-b-t diagram illustrating a persistent HTTP connection. . . . .  | 41 |
| 3.3  | An a-b-t diagram illustrating an SMTP connection. . . . .  | 44 |
| 3.4  | Three a-b-t diagrams representing three different types of NNTP interactions. . . . .                                | 46 |
| 3.5  | An a-b-t diagram illustrating a server push from a webcam using a persistent HTTP connection. . . . .                | 48 |
| 3.6  | An a-b-t diagram illustrating Icecast audio streaming in a TCP connection. . . . .                                   | 49 |
| 3.7  | Three a-b-t diagrams of connections taking part in the interaction between an FTP client and an FTP server. . . . .  | 49 |
| 3.8  | An a-b-t diagram illustrating an NNTP connection in “stream-mode”, which exhibits data exchange concurrency. . . . . | 51 |
| 3.9  | An a-b-t diagram illustrating the interaction between two BitTorrent peers. . . . .                                  | 51 |
| 3.10 | A first set of TCP segments for the connection vector in Figure 3.1: lossless example. . . .                         | 54 |
| 3.11 | A second set of TCP segments for the connection vector in Figure 3.1: lossy example. . . .                           | 55 |
| 3.12 | Distributions of ADU sizes for the testbed experiments with synthetic applications. . . . .                          | 66 |
| 3.13 | Distributions of quiet time durations for the testbed experiments with synthetic applications. . . . .               | 66 |
| 3.14 | Distributions of ADU sizes for the testbed experiments with synthetic applications. . . . .                          | 68 |
| 3.15 | Distributions of quiet time durations for the testbed experiments with synthetic applications. . . . .               | 68 |
| 3.16 | Bodies of the $A$ and $B$ distributions for Abilene-I, Leipzig-II and UNC 1 PM. . . . .                              | 73 |
| 3.17 | Tails of the $A$ and $B$ distributions for Abilene-I, Leipzig-II and UNC 1 PM. . . . .                               | 73 |
| 3.18 | Bodies of the $A$ and $B$ distributions with per-byte probabilities for Abilene-I, Leipzig-II and UNC 1 PM . . . . . | 74 |
| 3.19 | Bodies of the $E$ distributions for Abilene-I, Leipzig-II and UNC 1 PM. . . . .                                      | 74 |
| 3.20 | Bodies of the $E$ distributions with per-byte probabilities for Abilene-I, Leipzig-II and UNC 1 PM. . . . .          | 75 |
| 3.21 | Tails of the $E$ distributions for Abilene-I, Leipzig-II and UNC 1 PM. . . . .                                       | 75 |

|      |  |    |
|------|--|----|
| 3.22 | Average size of the epochs in each connection vector as a function of the number of epochs for Abilene-I, Leipzig-II and UNC 1 PM. . . . .             | 76 |
| 3.23 | Average of the median size of the ADUs in each connection vector as a function of the number of epochs for Abilene-I, Leipzig-II and UNC 1 PM. . . . . | 76 |
| 3.24 | Average of the median size of the ADUs in each connection vector as a function of the number of epochs, for Leipzig-II. . . . .                        | 77 |
| 3.25 | Average of the median size of the ADUs in each connection vector as a function of the number of epochs for Abilene-I. . . . .                          | 77 |
| 3.26 | Bodies of the <i>TA</i> and <i>TB</i> distributions for Abilene-I, Leipzig-II and UNC 1 PM. . . . .  | 77 |
| 3.27 | Tails of the <i>TA</i> and <i>TB</i> distributions for Abilene-I, Leipzig-II and UNC 1 PM. . . . .   | 77 |
| 3.28 | Distribution of the durations of the quiet times between the final ADU and connection termination. . . . .   | 79 |
| 3.29 | Bodies of the <i>A</i> and <i>B</i> distributions for the concurrent connections in Abilene-I, Leipzig-II and UNC 1 PM. . . . .                        | 79 |
| 3.30 | Tails of the <i>A</i> and <i>B</i> distributions for the concurrent connections in Abilene-I, Leipzig-II and UNC 1 PM. . . . .                         | 79 |
| 3.31 | Bodies of the <i>TA</i> and <i>TB</i> distributions for the concurrent connections in Abilene-I, Leipzig-II and UNC 1 PM. . . . .                      | 80 |
| 3.32 | Tails of the <i>TA</i> and <i>TB</i> distributions for the concurrent connections in Abilene-I, Leipzig-II and UNC 1 PM. . . . .                       | 80 |
| 3.33 | Bodies of the <i>A</i> distributions for UNC 1 AM, UNC 1 PM and UNC 7:30 PM. . . . .   | 81 |
| 3.34 | Bodies of the <i>B</i> distributions for UNC 1 AM, UNC 1 PM and UNC 7:30 PM. . . . .   | 81 |
| 3.35 | Bodies of the <i>TB</i> distributions for UNC 1 AM, UNC 1 PM and UNC 7:30 PM. . . . .  | 82 |
| 3.36 | Tails of the <i>TB</i> distributions for UNC 1 AM, UNC 1 PM and UNC 7:30 PM. . . . .   | 82 |
| 3.37 | Bodies of the <i>TA</i> distributions for three UNC traces. . . . .  | 83 |
| 3.38 | Tails of the <i>TA</i> distributions for three UNC traces. . . . .   | 83 |
| 4.1  | A set of TCP segments illustrating RTT estimation from connection establishment. . . . .   | 91 |
| 4.2  | Two sets of TCP segments illustrating RTT estimation ambiguities in the presence of loss and early retransmission in connection establishment. . . . . | 92 |
| 4.3  | A set of TCP segments illustrating RTT estimation using the sum of two OSTTs. . . . .  | 93 |
| 4.4  | A set of TCP segments illustrating the impact of delayed acknowledgments on OSTTs. . . . .   | 95 |
| 4.5  | Comparison of RTT estimators for a synthetic trace: no loss and enabled delayed acknowledgments. . . . .   | 96 |
| 4.6  | Comparison of RTT estimators for a synthetic trace: no loss and disabled delayed acknowledgments. . . . .  | 96 |

|      |   |     |
|------|---|-----|
| 4.7  | Comparison of RTT estimators for a synthetic trace: fixed loss rate of 1% for all connections. . . . .                                      | 98  |
| 4.8  | Comparison of RTT estimators for a synthetic trace: loss rates uniformly distributed between 0% and 10%. . . . .                            | 98  |
| 4.9  | A set of TCP segments illustrating an invalid OSTT sample due to the interaction between loss and cumulative acknowledgments. . . . .       | 98  |
| 4.10 | Comparison of RTT estimators for a synthetic trace: loss rates uniformly distributed between 0% and 10%. . . . .                            | 99  |
| 4.11 | Comparison of RTT estimators for synthetic traces: fixed loss rate of 1%; real RTTs up to 4 seconds. . . . .                                | 99  |
| 4.12 | Bodies of the RTT distributions for the five traces. . . . .  | 100 |
| 4.13 | Bodies of the RTT distributions with per-byte probabilities for the five traces. . . . .  | 100 |
| 4.14 | Comparison of the sum-of-minima and sum-of-medians RTT estimators for UNC 1 PM. . . . .   | 101 |
| 4.15 | Comparison of the sum-of-minima and sum-of-medians RTT estimators for Leipzig-II. . . . .   | 101 |
| 4.16 | Bodies of the distributions of maximum receiver window sizes for the five traces. . . . .   | 103 |
| 4.17 | Bodies of the distributions of maximum receiver window sizes with per-byte probabilities for the five traces. . . . .                       | 103 |
| 4.18 | Measured loss rates from experiments with 1% loss rates applied only on one direction or on both directions of the TCP connections. . . . . | 106 |
| 4.19 | Bodies of the distributions of loss rates for the five traces. . . . .  | 108 |
| 4.20 | Bodies of the distributions of loss rates with per-byte probabilities for the five traces. . . . .  | 108 |
| 4.21 | Breakdown of the byte throughput time series for Leipzig-II inbound. . . . .  | 110 |
| 4.22 | Breakdown of the packet throughput time series for Leipzig-II inbound. . . . .  | 110 |
| 4.23 | Breakdown of the byte throughput time series for Leipzig-II outbound. . . . .   | 112 |
| 4.24 | Breakdown of the packet throughput time series for Leipzig-II outbound. . . . .   | 112 |
| 4.25 | Breakdown of the byte throughput time series for Leipzig-II outbound. . . . .   | 112 |
| 4.26 | Breakdown of the packet throughput time series for Leipzig-II outbound. . . . .   | 112 |
| 4.27 | Breakdown of the byte throughput time series for Abilene-I Ipls/Clev. . . . .   | 113 |
| 4.28 | Breakdown of the packet throughput time series for Abilene-I Ipls/Clev. . . . .   | 113 |
| 4.29 | Breakdown of the byte throughput time series for Abilene-I Clev/Ipls. . . . .   | 114 |
| 4.30 | Breakdown of the packet throughput time series for Abilene-I Clev/Ipls. . . . .   | 114 |
| 4.31 | Breakdown of the byte throughput time series for UNC 1 PM inbound. . . . .  | 115 |
| 4.32 | Breakdown of the packet throughput time series for UNC 1 PM inbound. . . . .  | 115 |
| 4.33 | Breakdown of the byte throughput time series for UNC 1 PM outbound. . . . .   | 116 |

|      |   |     |
|------|---|-----|
| 4.34 | Breakdown of the packet throughput time series for UNC 1 PM outbound. . . . .   | 116 |
| 4.35 | Breakdown of the byte throughput time series for the three UNC traces. . . . .  | 117 |
| 4.36 | Breakdown of the packet throughput time series for the three UNC traces. . . . .  | 117 |
| 4.37 | Byte throughput marginals of Leipzig-II inbound, its normal distribution fit, the marginal distribution of its Poisson arrival fit, and the normal distribution fit of this Poisson arrival fit. . . . .    | 119 |
| 4.38 | Packet throughput marginals of Leipzig-II inbound, its normal distribution fit, the marginal distribution of its Poisson arrival fit, and the normal distribution fit of this Poisson arrival fit. . . . .  | 119 |
| 4.39 | Byte throughput marginals of UNC 1 PM outbound, its normal distribution fit, the marginal distribution of its Poisson arrival fit, and the normal distribution fit of this Poisson arrival fit. . . . .     | 120 |
| 4.40 | Packet throughput marginals of UNC 1 PM outbound, its normal distribution fit, the marginal distribution of its Poisson arrival fit, and the normal distribution fit of this Poisson arrival fit. . . . .   | 121 |
| 4.41 | Quantile-quantile plots with simulation envelopes for the marginal distribution of Leipzig-II inbound. The top four plots show byte throughput, while the four bottom plots show packet throughput. . . . . | 122 |
| 4.42 | Quantile-quantile plots with simulation envelopes for the marginal distribution of UNC 1 PM outbound. The top four plots show byte throughput, while the four bottom plots show packet throughput. . . . .  | 123 |
| 4.43 | Wavelet spectra of the packet throughput time series for Leipzig-II inbound and its Poisson arrival fit. . . . .  | 129 |
| 4.44 | Wavelet spectra of the byte throughput time series for Leipzig-II inbound and its Poisson arrival fit. . . . .  | 129 |
| 4.45 | Wavelet spectra of the packet throughput time series for Abilene-I. . . . .   | 130 |
| 4.46 | Wavelet spectra of the byte throughput time series for Abilene-I. . . . .   | 130 |
| 4.47 | Wavelet spectra of the packet throughput time series for UNC 1 PM. . . . .  | 131 |
| 4.48 | Wavelet spectra of the byte throughput time series for UNC 1 PM. . . . .  | 131 |
| 4.49 | Breakdown of the active connections time series for Leipzig-II. . . . .   | 132 |
| 4.50 | Impact of the definition of active connection on Leipzig-II. . . . .  | 132 |
| 4.51 | Breakdown of the active connections time series for Abilene-I. . . . .  | 133 |
| 4.52 | Impact of the definition of active connection on Abilene-I. . . . .   | 133 |
| 4.53 | Breakdown of active connections time series for UNC 1 PM using both definitions of active connection. . . . .   | 134 |
| 4.54 | Impact of the time-of-day on the active connections time series for the three UNC traces. . . . .   | 134 |
| 5.1  | Overview of Source-level Trace Replay. . . . .  | 140 |

|      |   |     |
|------|---|-----|
| 5.2  | Diagram of the network testbed where the experiments of this dissertation were conducted.   | 141 |
| 5.3  | End-host architecture of the traffic generation system. . . . .   | 143 |
| 5.4  | Bodies and tails of the <i>A</i> distributions for Leipzig-II and its source-level trace replays. . .   | 146 |
| 5.5  | Bodies and tails of the <i>B</i> distributions for Leipzig-II and its source-level trace replays. . .   | 148 |
| 5.6  | Bodies and tails of the <i>E</i> distributions for Leipzig-II and its source-level trace replays. . .   | 149 |
| 5.7  | Bodies and tails of the <i>TA</i> distributions for Leipzig-II and its source-level trace replays. .  | 149 |
| 5.8  | Bodies and tails of the <i>TB</i> distributions for Leipzig-II and its source-level trace replays. .  | 151 |
| 5.9  | Bodies of the round-trip time and receiver window size distributions for Leipzig-II and its source-level trace replays. . . . .   | 152 |
| 5.10 | Bodies the loss rate distributions for Leipzig-II and its source-level trace replays, with probabilities computed per connection (left) and per byte (right). . . . .   | 152 |
| 5.11 | Bodies and tails of the <i>A</i> distributions for UNC 1 PM and its source-level trace replays. .   | 154 |
| 5.12 | Bodies and tails of the <i>B</i> distributions for UNC 1 PM and its source-level trace replays. .   | 155 |
| 5.13 | Bodies and tails of the <i>E</i> distributions for UNC 1 PM and its source-level trace replays. .   | 155 |
| 5.14 | Bodies and tails of the <i>TA</i> distributions for UNC 1 PM and its source-level trace replays.  | 156 |
| 5.15 | Bodies and tails of the <i>TB</i> distributions for UNC 1 PM and its source-level trace replays.  | 156 |
| 5.16 | Bodies of the round-trip time and receiver window size distributions for UNC 1 PM and its source-level trace replays. . . . .   | 157 |
| 5.17 | Bodies of the loss rate distributions for UNC 1 PM and its source-level trace replays, with probabilities computed per connection (left) and per byte (right). . . . .  | 158 |
| 5.18 | Bodies and tails of the <i>A</i> distributions for Abilene-I and its source-level trace replays. . .  | 159 |
| 5.19 | Bodies and tails of the <i>B</i> distributions for Abilene-I and its source-level trace replays. . .  | 159 |
| 5.20 | Bodies and tails of the <i>E</i> distributions for Abilene-I and its source-level trace replays. . .  | 159 |
| 5.21 | Bodies and tails of the <i>TA</i> distributions for Abilene-I and its source-level trace replays. . .   | 160 |
| 5.22 | Bodies and tails of the <i>TB</i> distributions for Abilene-I and its source-level trace replays. .   | 160 |
| 5.23 | Bodies of the round-trip time and receiver window size distributions for Abilene-I and its source-level trace replays. . . . .  | 161 |
| 5.24 | Bodies of the loss rate distributions for Abilene-I and its source-level trace replays, with probabilities computed per connection (left) and per byte (right). . . . . | 161 |
| 6.1  | Byte throughput time series for Leipzig-II inbound and its four types of source-level trace replay. . . . .   | 169 |
| 6.2  | Byte throughput time series for Leipzig-II outbound and its four types of source-level trace replay. . . . .  | 170 |

|      |   |     |
|------|---|-----|
| 6.3  | Packet throughput time series for Leipzig-II inbound and its four types of source-level trace replay. . . . .                         | 171 |
| 6.4  | Packet throughput time series for Leipzig-II outbound and its four types of source-level trace replay. . . . .                        | 172 |
| 6.5  | Byte throughput marginals for Leipzig-II inbound and its four types of source-level trace replay. . . . .                             | 173 |
| 6.6  | Byte throughput marginals for Leipzig-II outbound and its four types of source-level trace replay. . . . .                            | 175 |
| 6.7  | Packet throughput marginals for Leipzig-II inbound and its four types of source-level trace replay. . . . .                           | 176 |
| 6.8  | Packet throughput marginals for Leipzig-II outbound and its four types of source-level trace replay. . . . .                          | 177 |
| 6.9  | Wavelet spectra of the byte throughput time series for Leipzig-II inbound and its four types of source-level trace replay. . . . .    | 178 |
| 6.10 | Wavelet spectra of the byte throughput time series for Leipzig-II outbound and its four types of source-level trace replay. . . . .   | 178 |
| 6.11 | Wavelet spectra of the packet throughput time series for Leipzig-II inbound and its four types of source-level trace replay. . . . .  | 181 |
| 6.12 | Wavelet spectra of the packet throughput time series for Leipzig-II outbound and its four types of source-level trace replay. . . . . | 181 |
| 6.13 | Active connection time series for Leipzig-II and its four types of source-level trace replay.   | 182 |
| 6.14 | Byte throughput time series for UNC 1 PM inbound and its four types of source-level trace replay. . . . .                             | 183 |
| 6.15 | Byte throughput time series for UNC 1 PM outbound and its four types of source-level trace replay. . . . .                            | 184 |
| 6.16 | Packet throughput time series for UNC 1 PM inbound and its four types of source-level trace replay. . . . .                           | 185 |
| 6.17 | Packet throughput time series for UNC 1 PM outbound and its four types of source-level trace replay. . . . .                          | 186 |
| 6.18 | Byte throughput marginals for UNC 1 PM inbound and its four types of source-level trace replay. . . . .                               | 187 |
| 6.19 | Byte throughput marginals for UNC 1 PM outbound and its four types of source-level trace replay. . . . .                              | 187 |
| 6.20 | Packet throughput marginals for UNC 1 PM inbound and its four types of source-level trace replay. . . . .                             | 188 |
| 6.21 | Packet throughput marginals for UNC 1 PM outbound and its four types of source-level trace replay. . . . .                            | 189 |
| 6.22 | Wavelet spectra of the byte throughput time series for UNC 1 PM inbound and its four types of source-level trace replay. . . . .      | 190 |

|      |   |     |
|------|---|-----|
| 6.23 | Wavelet spectra of the byte throughput time series for UNC 1 PM outbound and its four types of source-level trace replay. . . . .   | 190 |
| 6.24 | Wavelet spectra of the packet throughput time series for UNC 1 PM inbound and its four types of source-level trace replay. . . . .  | 193 |
| 6.25 | Wavelet spectra of the packet throughput time series for UNC 1 PM outbound and its four types of source-level trace replay. . . . . | 193 |
| 6.26 | Active connection time series for UNC 1 PM and its four types of source-level trace replay.   | 194 |
| 6.27 | Byte throughput time series for UNC 1 AM inbound and its four types of source-level trace replay. . . . .                           | 198 |
| 6.28 | Byte throughput time series for UNC 1 AM outbound and its four types of source-level trace replay. . . . .                          | 198 |
| 6.29 | Packet throughput time series for UNC 1 AM inbound and its four types of source-level trace replay. . . . .                         | 199 |
| 6.30 | Packet throughput time series for UNC 1 AM outbound and its four types of source-level trace replay. . . . .                        | 199 |
| 6.31 | Byte throughput marginals for UNC 1 AM inbound and its four types of source-level trace replay. . . . .                             | 200 |
| 6.32 | Byte throughput marginals for UNC 1 AM outbound and its four types of source-level trace replay. . . . .                            | 200 |
| 6.33 | Packet throughput marginals for UNC 1 AM inbound and its four types of source-level trace replay. . . . .                           | 201 |
| 6.34 | Packet throughput marginals for UNC 1 AM outbound and its four types of source-level trace replay. . . . .                          | 201 |
| 6.35 | Wavelet spectra of the byte throughput time series for UNC 1 AM inbound and its four types of source-level trace replay. . . . .    | 203 |
| 6.36 | Wavelet spectra of the byte throughput time series for UNC 1 AM outbound and its four types of source-level trace replay. . . . .   | 203 |
| 6.37 | Wavelet spectra of the packet throughput time series for UNC 1 AM inbound and its four types of source-level trace replay. . . . .  | 204 |
| 6.38 | Wavelet spectra of the packet throughput time series for UNC 1 AM outbound and its four types of source-level trace replay. . . . . | 204 |
| 6.39 | Active connection time series for UNC 1 AM and its four types of source-level trace replay.   | 205 |
| 6.40 | Byte throughput time series for UNC 7:30 PM inbound and its four types of source-level trace replay. . . . .                        | 206 |
| 6.41 | Byte throughput time series for UNC 7:30 PM outbound and its four types of source-level trace replay. . . . .                       | 206 |
| 6.42 | Packet throughput time series for UNC 7:30 PM inbound and its four types of source-level trace replay. . . . .                      | 207 |



|      |  |     |
|------|--|-----|
| 6.43 | Packet throughput time series for UNC 7:30 PM outbound and its four types of source-level trace replay. . . . .                        | 207 |
| 6.44 | Byte throughput marginals for UNC 7:30 PM inbound and its four types of source-level trace replay. . . . .                             | 208 |
| 6.45 | Byte throughput marginals for UNC 7:30 PM outbound and its four types of source-level trace replay. . . . .                            | 208 |
| 6.46 | Packet throughput marginals for UNC 7:30 PM inbound and its four types of source-level trace replay. . . . .                           | 209 |
| 6.47 | Packet throughput marginals for UNC 7:30 PM outbound and its four types of source-level trace replay. . . . .                          | 209 |
| 6.48 | Wavelet spectra of the byte throughput time series for UNC 7:30 PM inbound and its four types of source-level trace replay. . . . .    | 211 |
| 6.49 | Wavelet spectra of the byte throughput time series for UNC 7:30 PM outbound and its four types of source-level trace replay. . . . .   | 211 |
| 6.50 | Wavelet spectra of the packet throughput time series for UNC 7:30 PM inbound and its four types of source-level trace replay. . . . .  | 212 |
| 6.51 | Wavelet spectra of the packet throughput time series for UNC 7:30 PM outbound and its four types of source-level trace replay. . . . . | 212 |
| 6.52 | Active connection time series for UNC 7:30 PM and its four types of source-level trace replay. . . . .                                 | 213 |
| 6.53 | Byte throughput time series for Abilene-I Clev/Ipls and its four types of source-level trace replay. . . . .                           | 214 |
| 6.54 | Byte throughput time series for Abilene-I Ipls/Clev and its four types of source-level trace replay. . . . .                           | 214 |
| 6.55 | Packet throughput time series for Abilene-I Clev/Ipls and its four types of source-level trace replay. . . . .                         | 215 |
| 6.56 | Packet throughput time series for Abilene-I Ipls/Clev and its four types of source-level trace replay. . . . .                         | 215 |
| 6.57 | Byte throughput marginals for Abilene-I Clev/Ipls and its four types of source-level trace replay. . . . .                             | 216 |
| 6.58 | Byte throughput marginals for Abilene-I Ipls/Clev and its four types of source-level trace replay. . . . .                             | 216 |
| 6.59 | Packet throughput marginals for Abilene-I Clev/Ipls and its four types of source-level trace replay. . . . .                           | 217 |
| 6.60 | Packet throughput marginals for Abilene-I Ipls/Clev and its four types of source-level trace replay. . . . .                           | 217 |
| 6.61 | Wavelet spectra of the byte throughput time series for Abilene-I Clev/Ipls and its four types of source-level trace replay. . . . .    | 219 |

|      |   |     |
|------|---|-----|
| 6.62 | Wavelet spectra of the byte throughput time series for Abilene-I Ipls/Clev and its four types of source-level trace replay. . . . .   | 219 |
| 6.63 | Wavelet spectra of the packet throughput time series for Abilene-I Clev/Ipls and its four types of source-level trace replay. . . . .   | 220 |
| 6.64 | Wavelet spectra of the packet throughput time series for Abilene-I Ipls/Clev and its four types of source-level trace replay. . . . .   | 220 |
| 6.65 | Active connection time series for Abilene-I and its four types of source-level trace replay. . . . .  | 221 |
| 7.1  | Bodies of the distributions of connection inter-arrivals for UNC 1 PM and 1 AM, and their exponential fits. . . . .   | 227 |
| 7.2  | Tails of the distributions of connection inter-arrivals for UNC 1 PM and 1 AM, and their exponential fits. . . . .  | 227 |
| 7.3  | Bodies of the distributions of connection inter-arrivals for Abilene-I and Leipzig-II, and their exponential fits. . . . .  | 227 |
| 7.4  | Tails of the distributions of connection inter-arrivals for Abilene-I and Leipzig-II, and their exponential fits. . . . .   | 227 |
| 7.5  | Average offered load <i>vs.</i> number of connections for 1,000 Poisson resamplings of UNC 1 PM. . . . .  | 229 |
| 7.6  | Histogram of the average offered loads in 1,000 Poisson resamplings of UNC 1 PM. . . . .  | 229 |
| 7.7  | Tails of the distributions of connection sizes for UNC 1 PM. . . . .  | 230 |
| 7.8  | Analysis of the accuracy of connection-driven Poisson Resampling from 6,000 resamplings of UNC 1 PM (1,000 for each target offered load). . . . .   | 230 |
| 7.9  | Comparison of average offered load <i>vs.</i> number of connections for 1,000 connection-driven Poisson resamplings and 1,000 byte-driven Poisson resamplings of UNC 1 PM. . . . .  | 232 |
| 7.10 | Histogram of the average offered loads in 1,000 byte-driven Poisson resamplings of UNC 1 PM. . . . .  | 232 |
| 7.11 | Analysis of the accuracy of byte-driven Poisson Resampling from 4,000 resamplings of UNC 1 PM (1,000 for each target offered load). . . . .   | 233 |
| 7.12 | Analysis of the accuracy of byte-driven Poisson Resampling using source-level traces replay: replays of three separate resamplings of UNC 1 PM for each target offered load, illustrating the scaling down of load from the original 177.36 Mbps. . . . . | 233 |
| 7.13 | Analysis of the accuracy of byte-driven Poisson Resampling using testbed experiments: replay of one resampling of UNC 1 AM for each target offered load, illustrating the scaling up of load from the original 91.65 Mbps. . . . .                        | 233 |
| 7.14 | Connection arrival time series for UNC 1 PM (dashed line) and a Poisson arrival process with the same mean (solid line). . . . .  | 235 |
| 7.15 | Connection arrival time series for UNC 1 AM and a Poisson arrivals process with the same mean. . . . .  | 235 |

|      |  |     |
|------|--|-----|
| 7.16 | Wavelet spectra of the connection arrival time series for UNC 1 PM and a Poisson arrival process with the same mean. . . . .   | 235 |
| 7.17 | Wavelet spectra of the connection arrival time series for UNC 1 AM and a Poisson arrival process with the same mean. . . . .   | 235 |
| 7.18 | Block resamplings of UNC 1 PM: impact of different block lengths on the wavelet spectrum of the connection arrival time series. . . . .  | 238 |
| 7.19 | Block resamplings of UNC 1 AM: impact of different block lengths on the wavelet spectrum of the connection arrival time series. . . . .  | 239 |
| 7.20 | Block resamplings of UNC 1 PM: average offered load <i>vs.</i> number of connection vectors (left) and corresponding histograms of average offered loads (right) in 3,000 resamplings. . . . .   | 241 |
| 7.21 | Wavelet spectra of several random subsamplings of the connection vectors in UNC 1 PM (left) and 1 AM (right) . . . . .   | 242 |
| 7.22 | Analysis of the accuracy of byte-driven Block Resampling using source-level trace replay: replays of two separate resamplings of UNC 1 PM for each target offered load, illustrating the scaling down of load from the original 177.36 Mbps. . . . . | 243 |
| 7.23 | Analysis of the accuracy of byte-driven Block Resampling using source-level trace replay: replay of one resampling of UNC 1 AM for each target offered load, illustrating the scaling up of load from the original 91.65 Mbps. . . . .               | 243 |
| 7.24 | Wavelet spectra of the packet arrival time series for UNC 1 PM and the source-level trace replays of two block resamplings of this trace. . . . .  | 245 |
| 7.25 | Wavelet spectra of the packet arrival time series for UNC 1 PM and the source-level trace replays of three Poisson resamplings of this trace. . . . .  | 245 |

## LIST OF ABBREVIATIONS

|             |  |
|-------------|--|
| <b>ACK</b>  | Positive acknowledgment TCP segment                    |
| <b>ADU</b>  | Application Data Unit                                  |
| <b>API</b>  | Application Programming Interface                      |
| <b>AQM</b>  | Active Queue Management                                |
| <b>BGP</b>  | Border Gateway Protocol                                |
| <b>BPF</b>  | Berkeley Packet Filter                                 |
| <b>C.I.</b> | Confidence Interval                                    |
| <b>CCDF</b> | Complementary Cumulative Distribution Function         |
| <b>CDF</b>  | Cumulative Distribution Function                       |
| <b>DAG</b>  | Data Acquisition and Generation                        |
| <b>FIFO</b> | First-In First-Out                                     |
| <b>FIN</b>  | TCP control flag indicating “no more data from sender” |
| <b>FTP</b>  | File Transfer Protocol                                 |
| <b>GB</b>   | Gigabyte   |
| <b>GPS</b>  | Global Positioning System                              |
| <b>HTML</b> | HyperText Markup Language                              |
| <b>HTTP</b> | HyperText Transfer Protocol                            |
| <b>I/O</b>  | Input/Output   |
| <b>ICMP</b> | Internet Control Message Protocol                      |
| <b>IP</b>   | Internet Protocol                                      |
| <b>IRC</b>  | Internet Relay Chat                                    |
| <b>ISP</b>  | Internet Service Provider                              |
| <b>K-S</b>  | Kolmogorov-Smirnov test                                |
| <b>KB</b>   | Kilobyte   |
| <b>Kpps</b> | Kilo packet per second                                 |
| <b>LRD</b>  | Long-Range Dependence                                  |
| <b>MB</b>   | Megabyte   |
| <b>MIME</b> | Multipurpose Internet Mail Extensions                  |
| <b>MSS</b>  | Maximum Segment Size                                   |
| <b>MTU</b>  | Maximum Transmission Unit                              |

|                |  |
|----------------|--|
| <b>Mbps</b>    | Megabit per second                             |
| <b>NNTP</b>    | Network News Transfer Protocol                 |
| <b>OSTT</b>    | One-Side Transit Time                          |
| <b>PMA</b>     | Passive Measurement and Analysis               |
| <b>Q-Q</b>     | Quantile-Quantile                              |
| <b>RED</b>     | Random Early Detection                         |
| <b>RFC</b>     | Request For Comments                           |
| <b>RST</b>     | TCP control flag indicating “connection reset” |
| <b>RTT</b>     | Round-Trip Time                                |
| <b>SMTP</b>    | Simple Mail Transfer Protocol                  |
| <b>SSH</b>     | Secure Shell                                   |
| <b>SYN</b>     | Synchronize TCP control segment                |
| <b>SYN-ACK</b> | Positive acknowledgement of SYN segment        |
| <b>TCP</b>     | Transport Control Protocol                     |
| <b>UDP</b>     | User Datagram Protocol                         |
| <b>UNC</b>     | University of North Carolina at Chapel Hill    |
| <b>URL</b>     | Universal Resource Locator                     |