

NISS Affiliates Workshop on Modeling and Analysis of Network Data

March 9–10, 2001
NISS Headquarters
Research Triangle Park, NC

Organizing Committee

K. R. Krishnan, Telcordia Technologies
Diane Lambert, Bell Labs, Lucent Technologies
John P. Lehoczky, Carnegie Mellon University
J. S. Marron, University of North Carolina at Chapel Hill (Chair)
Sidney Resnick, Cornell University

Program

Friday, March 9, 2001

Session I: Network Measurement and Monitoring

- 1:00 PM Welcome and Introductions
- 1:15 Global DNS Nameserver Performance Analysis using NeTraMet
Nevil Brownlee, CAIDA
- 1:45 Discussion
- 1:55 PackMime: An Internet Traffic Generator
Jin Cao, William S. Cleveland, Dong Lin and Don X. Sun, Bell Labs, Lucent Technologies
- 2:45 Discussion
- 2:55 Bandwidth and Congestion Measurement via Remote Monitoring
Sam Weerahandi, Telcordia Technologies
- 3:45 Discussion
- 3:55 Internet Topology
Andre Broido and kc claffey, CAIDA
- 4:45 Discussion
- 4:55 Summary Discussion
K. R. Krishnan, Telcordia Technologies
- 5:20 Adjourn
- 5:45 Shuttle to Governors Inn
- 7:00 Workshop Dinner

Saturday, March 10, 2001

Session II: Network Performance Control

- 8:30 AM Scheduling Your Network Connections
Mor Harchol-Balter, Carnegie Mellon University
- 9:20 Discussion
- 9:30 Duality Model of TCP and Queue Management Algorithms
Steven Low, California Institute of Technology
- 9:20 Discussion
- 10:30 On Dynamic Control of Stochastic Networks: Fluid and Higher Order Approximations
Ruth J. Williams, University of California, San Diego
- 11:20 Discussion
- 11:30 Summary Discussion
John P. Lehoczky, Carnegie Mellon University
- 12:00 N Lunch

Session III: Internet Traffic Models

- 1:15 PM Statistical Resource Sharing in the Presence of Heavy Tails
Predrag Jelenkovic, Columbia University
- 2:05 Discussion
- 2:15 On Understanding Large-Scale TCP/IP Networks: A Lesson in Internet Traffic Modeling
Walter Willinger, AT&T Labs Research
- 3:05 Discussion
- 3:15 Multiplicative Modeling of Network Traffic
Rolf Riedi, Rice University
- 4:05 Discussion
- 4:15 Summary Discussion
Don Sun, Bell Labs, Lucent Technologies
- 4:45 Adjourn

Abstracts

Session I: Network Measurement and Monitoring

Global DNS Nameserver Performance Analysis using NeTraMet

Nevil Brownlee, CAIDA

The Domain Name System can be viewed as having three levels: host, local nameserver, global nameserver. Enterprise networks have local nameservers, which query the global nameservers to resolve subdomains of the top-level domains. Most local nameservers run BIND, which distributes its queries across all of the global nameservers.

We use NetTraMet meters to collect data on DNS requests and we responses. For a chosen set of local nameservers we collect 5-minute samples of request counts, lost request percentage (no matching response), and median response time for DNS queries of all the global nameservers.

Strip charts of these metrics allow one to see performance variations in individual servers (affecting a single server), and in the Internet paths to them (affecting several servers). They provide a useful tool for monitoring a site's global connectivity.

Detailed plots of these metrics, together with total traffic rate on a site's Internet connection(s), are also discussed. These can yield some understanding of the way variations in traffic rate and/or request rate can produce corresponding variations in response time and loss percentage.

PackMime: An Internet Traffic Generator

Jin Cao, William S. Cleveland, Dong Lin, and Don X. Sun

Bell Labs, Lucent Technologies

Overview: PackMime, a closed-loop IP traffic model, generates synthetic packet traffic that mimics the behavior of live aggregate Internet packet traffic. The PackMime architecture consists of two parts: (1) statistical models that stochastically generate the aggregate request traffic for TCP connections for each of a collection of clouds of hosts and each application protocol; and (2) a simulated network carrying packets resulting from the generated connections. The traffic for each cloud enters and exits the network at some node of the network. Actual TCP implementations from the BSD kernel are used to deliver the packets across the simulated network and provide TCP feedback.

Statistical Models: The models generate connection request variables (as time sequences) for each cloud and for each application protocol whose traffic is carried on the simulated network. For HTTP 1.0, the connection variables are request times, client request file sizes, downloaded server file sizes, round-trip times between the clients and the network, round-trip times between the servers and the network, server delays, and bottleneck link speeds (used not to model bottlenecks but rather to create packet jitter). All of these variables exhibit long-range dependence that decreases to independence as the traffic request load increases. The inter-arrival times have a Weibull distribution with an upper tail that is longer than the exponential but that tends to the exponential as the load increases. For each variable, a very simple fractional sum-difference (FSD) time series model-class, together with transformations of the variables, model the long-range dependence. The FSD model parameters and the Weibull shape and scale parameters change with the load, which is measured in new connections/second, and simple functions of load model the parameter changes.

Input: To run PackMime, one needs to specify the network, the clouds, and the application request loads for each cloud.

Empirical Study: The basis of PackMime is extensive analysis of packet header databases collected on live Internet wires. There are two types of analysis. (1) The statistical models have been built and validated based on analyses of the databases. (2) PackMime performance in generating synthetic packet traffic has been evaluated based on analyses of the live packet traffic of the databases. For this latter task we have developed a suite of statistical tools for analyzing packet processes: packet inter-arrival times and packet sizes. The tools are applied to the live header databases and to similar databases of headers from the PackMime synthetic traffic, and the results are compared. In particular, PackMime reproduces our finding that Internet packet arrivals on a link tend to Poisson and packet sizes tend to independence as the request load increases.

S-Net: Empirical analysis is carried out using S-Net, an Internet traffic measurement and analysis system that begins with packet header collection on network wires, and ends with data analysis on a cluster of Linux PCs running S, a language and system for organizing, visualizing, and analyzing data. S-Net allows detailed analysis of packet header information, achieved in part by extensive visualization. We have studied two header databases collected by our Bell Lab team and our Helios team, an NGI project based at the NCNI network (about a terabyte of compressed headers); and we have studied databases collected by a Harvard team and by the NLANR team.

PackMime Applications: PackMime has been used to study performance issues on a single link carrying HTTP 1.0 traffic. A PackMime queueing analysis shows that as the traffic load increases, queueing tends to that of packet traffic with Poisson arrivals and independent service times. We have used PackMime to provision the link speed of the wire that connects our Bell Labs research network to the rest of the Internet through studying the HTTP 1.0 transfer time distribution and packet loss as the link speed changes. This replaces the over-provision method that has been used historically on the link, which seems today unattractive because it has become less acceptable to waste money.

Status: We believe we have amply demonstrated that the PackMime approach succeeds. But (1) we need to expand the scope, and (2) study further certain model features.

- So far, only HTTP 1.0 has been fully modeled, but work is in progress on a number of other application protocols.
- So far, the ASM simulator of Dong Lin has provided the network simulations, but we are now exploring the combination of the PackMime statistical models with other network simulators (such as NS) to expand the potential scope of network simulation.
- PackMime draws the feedback line between packets and requests. We believe, as others have argued strongly, that without TCP feedback, open-loop packet presentation is not sufficient to gain an understanding of Internet performance. But PackMime does not extend the feedback into the user request process, even though users clearly respond to congestion. We believe we can learn much about the network without user feedback, but this issue surely deserves further discussion.
- We were not satisfied with initial results of PackMime. Packet arrivals were too regular and we did not see as much jitter as is present in live traces. We have modeled congestion in the cloud to the extent that the round-trip time variables are based on observed round-trip times, which are affected by congestion; that is, we do not use propagation delays. But the cloud round-trip time is constant during the connection, and we have not modeled buffering and packet drops in the cloud, so there is no added jitter. We do add server delay, because we can measure it, but this does not add enough jitter. Because the technique was quite simple, and because it worked, we created jitter through the bottleneck link speed variable, which generates a bottleneck link speed for each connection. This issue needs further attention.

For more information see cm.bell-labs.com/stat/InternetTraffic.

Bandwidth and Congestion Measurement via Remote Monitoring

Sam Weerahandi, Telcordia Technologies

The paper deals with the problem of estimating the physical bandwidth and the used bandwidth of a link in an IP network, particularly the Internet. We carry out the bandwidth estimation based on active measurements taken from a remote monitor. This type of estimation is useful in situations where one needs to know the Quality of Service (QoS) of a leased network or a third party network that affects the QoS of own customers and it is not possible to obtain passive measurements from devices of such networks. Like the well known tool PathChar, our method also determines the bandwidth between any two nodes in such a network by taking various delay measurements from the remote monitor to the end points of the link. Unlike PathChar, however, we do not require estimation of bandwidth for all hops leading to the link of interest, thus minimizing the traffic injected for the purpose of estimation. Moreover, in addition to the physical bandwidth, we also estimate the delay contribution of the hop and used bandwidth near real-time.

Delay data are measured as the size of the packet vary over a reasonable range. The data obtained in this manner is then statistically analyzed to estimate both the physical bandwidth and the used bandwidth. A model appropriate for bandwidth estimation will be presented along with an overview of derivations. Problem of estimating parameters of bandwidth models will also be discussed. Some results comparing the performance of bandwidth estimation using our methodology and that of PathChar will be presented and additional capabilities of our solution will be discussed.

Internet Topology

Andre Broido and kc claffey, CAIDA

We describe fundamentals of analyzing Internet graphs at various layers, and problems in gathering and analyzing Internet routing and topology data in graph form. Using skitter topology data and Oregon Route Views BGP table data, we present IP graphs, their connected components, and the combinatorial core of Internet topology. We discuss techniques for assessing which portions of the global Internet are characterized by the highest degree of ‘connectivity,’ both in central/backbone and access/delivery components of the topology. We describe several combinatorial approaches, including:

1. Extracting the core component of the Internet whose bidirectional connectivity most readily captured by measurement, even in observation conditions which are far from ideal
2. Ordering ‘node centrality’ by the lengths of shortest paths originating from them;
3. Comparing access points by the size of the “access cone” — the number of nodes/prefixes/ASes and/or the size of address space that depend wholly or in part on this access point for their global connectivity.

Our analysis introduces several new concepts for graph-theoretic routing and topology analysis:

1. The “dual AS graph,” that captures more policy constraints in the infrastructure than conventional (dimension 1) graph. In particular, a graph of AS adjacencies is a poor descriptor for peerings further than one hop due to the influence of policy.

2. The BGP atom, a unit of connectivity analysis that correspond to an equivalence class of IPv4 network prefixes that share the same set of AS paths.
3. Connected subspaces of a prefix, a unit of IP level connectivity that allows us to avoid certain biases arising in straightforward compression of IP graph to prefix representation.

We will also present tables showing the top 50 ‘most connected’ nodes, at various levels of Internet node granularity. We cover distributions of AS path and prefix lengths, use of prepending, and outdegree vs indegree of ASes. Time permitting, some newer results regarding diffusion on Internet graphs and nodes’ geodesic load will be discussed.

Background material is available at:

- www.caida.org/outreach/isma/0012/finalReport.xml;
- www.caida.org/outreach/isma/0012/;
- www.caida.org/broido/bgp/atoms.html;
- www.caida.org/broido/bgp/asoutdeg.html; and
- www.caida.org/broido/bgp/stubas.html.

Session II: Network Performance Control

Scheduling Your Network Connections

Mor Harchol-Balter, Carnegie Mellon University

This talk proposes a method for improving the performance of Web servers servicing static HTTP requests. The idea is to give preference to those requests which are quick, or have small remaining processing requirements, in accordance with the SRPT (Shortest-Remaining-Processing-Time) scheduling policy.

The implementation is at the kernel level and involves controlling the order in which socket buffers are drained into the network. Experiments use the Linux operating system and the Apache and Flash web servers.

Empirical results indicate that SRPT-based scheduling of connections yields significant reductions in mean response time, mean slowdown, and variance in response time at the Web server. Most importantly, and counter to intuition, the large requests are not at all (or hardly) penalized by SRPT-based scheduling.

To explain the above results, we present some of our new theorems on SRPT scheduling under workloads with a heavy-tailed property. These theorems motivate our implementation work.

Duality Model of TCP and Queue Management Algorithms

Steven Low, California Institute of Technology

We describe a duality model of congestion control and apply it to TCP and active queue management schemes. Congestion control is the interaction of source rates with certain congestion measures in the network. The basic idea is to regard source rates as primal variables and congestion measures as dual variables,

and congestion control as a Lagrangian method that iterates on source rates and congestion measures to maximize aggregate source utility subject to capacity constraint. In TCP, the primal iteration is carried out by source algorithms such as Reno or Vegas, and the dual iteration is carried out by queue management such as DropTail, RED or REM. We present these algorithms and their generalizations, derive their utility functions, and study their interaction.

On Dynamic Control of Stochastic Networks:

Fluid and Higher Order Approximations

Ruth J. Williams, University of California, San Diego

This talk will give a partial survey and present some recent developments concerning approximate models for stochastic networks, where the networks allow for flexible scheduling of resources through dynamic (state-dependent) alternate routing and sequencing. Fluid and higher order approximations to network dynamic control problems will be discussed. In particular, connections between these levels of approximation will be described. Situations in which the network is congested or heavily loaded will receive particular emphasis. Some examples that illustrate resource pooling effects will be given.

Session III: Internet Traffic Models

Statistical Resource Sharing in the Presence of Heavy Tails

Predrag Jelenkovic, Columbia University

Statistical sharing of network bottleneck resources, e.g., bandwidth, processing power and buffer space, represents a common way of increasing the capacity of network transmission elements. This sharing allows for temporary high demands of a subset of users to be accommodated due to lower short-term needs of the remaining clients. While the benefits of sharing were well understood for exponential traffic models, the impact of heavy tails and long-range dependence has just been uncovered. Our results quantify these benefits in the case of baseline On-Off transmission demands. In addition, the explicit nature of the results explains the tradeoff between bandwidth/processing power and buffer space that might be a useful guideline for future network designs.

On Understanding Large-Scale TCP/IP Networks:

A Lesson in Internet Traffic Modeling

Walter Willinger, AT&T Labs Research

Recent empirical discoveries concerning various scaling properties of the temporal dynamics of Internet traffic or of some of the topological features associated with the physical structure of the Internet have resulted in a number of proposed models or “explanations” of these “emergent” phenomena. In this talk, I will discuss a simple framework for identifying those explanations that are relevant for the networking application at hand and can provide a basis for developing a useful, consistent, and verifiable theory of

large-scale TCP/IP networks. In turn, I will also show why some proposed models and “explanations” are irrelevant, at least as far as the networking context is concerned.

Multiplicative Modeling of Network Traffic

Rolf Riedi, Rice University

Recent studies have made it clear that network traffic is ‘multifractal’ in nature. Using moments of (possibly) all orders multifractal analysis does justice to the positive, non-Gaussian marginals of traffic traces and goes beyond the second order approximations of linear processes.

Multifractal structures are naturally tied to multiplicative frameworks. Indeed, a simple stochastic traffic model based on the Haar wavelet demonstrates the effectiveness of a multiplicative approach and its superiority over the typical additive schemes, both in multifractal statistics capturing ‘burstiness’ as well as in network relevant metrics. Due to the inherent tree structure, wavelet based models are not second order stationary. To overcome this disadvantage this talk explores novel, stationary multifractal processes and the statistical challenges they present.

NISS Affiliates

Corporations

Amgen
AT&T Shannon Laboratory
Avaya Labs
Bell Laboratories, Lucent Technologies
General Motors
GlaxoSmithKline
Mayo Clinic
Merck & Company
MetaMetrics
Pfizer Inc.
Procter & Gamble
Quintiles Transnational
Research Triangle Institute
SAS Institute
Telcordia Technologies
Visual Insights

Government Agencies and National Laboratories

Bureau of Labor Statistics
Bureau of Transportation Statistics
Census Bureau
US Environmental Protection Agency
Los Alamos National Laboratory
National Center for Education Statistics
National Institute of Standards and Technology
National Security Agency
Pacific Northwest National Laboratory

University Departments

American University (Mathematics and Statistics)
Carnegie Mellon University (Statistics)
Duke University (Statistics and Decision Sciences)
University of Iowa (Statistics)
Iowa State University (Statistics)
Johns Hopkins University (Mathematical Sciences)
University of Maryland Baltimore County (Mathematics and Statistics)
University of Minnesota (Biostatistics)
North Carolina State University (Statistics)
University of North Carolina at Chapel Hill (Biostatistics)
University of North Carolina at Chapel Hill (Statistics)
Ohio State University (Statistics)
Pennsylvania State University (Statistics)
University of Pennsylvania (Statistics)
Purdue University (Statistics)
Stanford University (Statistics)
Texas A&M University (Statistics)
Western Michigan University (Statistics)