Final Program Report:

Network Modeling of the Internet

March 11, 2005

1. Introduction, Motivation and Initial Ideas

Because of the size and complexity of the internet, and the nature of the protocols, Internet traffic has proven to be very challenging to model effectively. Yet modeling is critical to improving Quality of Service and efficiency. The main research goal of this program was to address these issues by bringing together researchers from three communities:

- Applied probabilists studying heavy traffic queueing theory and fluid flow models;
- Mainstream Internet traffic measurers/modelers and hardware/software architects;
- Statisticians.

The timing was deemed right for simultaneous interaction among all three communities, because of current trends away from dealing with Quality of Service issues through over-provisioning of equipment. This trend suggested that heavy traffic models would be ideally situated to play a leading role in future modeling of Internet traffic, and in attaining deeper understanding of the complex drivers behind Quality of Service. This SAMSI program aimed at catalyzing this process through building strong bridges among the three communities.

An additional goal was to enhance contact between these research communities and potential industrial partners. The location of SAMSI is ideal for this purpose, because the Research Triangle is becoming a world center for the networking industry.

Program Leadership: Kevin Jeffay (North Carolina), James Landwehr (Avaya Laboratories), John P. Lehoczky (Carnegie Mellon), J. S. Marron (North Carolina/SAMSI, Co-Chair), Ruth Williams (California San Diego, Co-Chair), Walter Willinger (AT\&T Research), Donald Towsley (Massachusetts).

2. Program Goals

The main value of the SAMSI Program on Network Modeling for the Internet came from bringing together statisticians, probabilists and network researchers (computer scientists and electrical engineers). While there had been previous pairwise contacts between these groups, there were great benefits from strengthening these connections.

Specific research emphases, where this new three way interaction was especially fruitful, include:

- Measurement and Modelling
- Heavy Traffic – Congestion Control
- Internet Tomography
- Sensor Networks

3. Workshops

One workshop was originally planned for each of the 4 areas of emphasis listed above. The schedule for the first two was seriously disrupted by Hurricane Isabel. This was handled by turning the Measurement and Modeling Workshop into a smaller meeting of local participants (which Darryl Veitch of Sprint Labs attended as well), because we had critical mass in this area. The Heavy Traffic – Congestion Control Workshop was postponed (because of insufficient local involvement to start activities).

The format of all workshops was aimed at maximizing long term involvement. Plans varied somewhat, depending on the nature of the topic, and on the current state of the desired collaboration. These contacts were weakest in the case of Sensor Networks, so the format was mostly the traditional talk + discussion type, but with non-expert discussants employed as often as possible. For the other workshops, at least some pairwise collaborations had already been established, so activities were deliberately aimed at maximizing engagement of participants. Some of the devices, with discussion, were:

i. Five Minute Madness Introductory Session. A session of 5 minute talks, intended to serve as an introduction. While there is insufficient time for all participants to speak, a cross-section of junior and senior people, from all three backgrounds, gave all participants a clear view of the diversity of people present at the workshop. Some emphasis in selection of people was given towards senior researchers, with the idea that everybody would be interested in hearing a sound bite description of their current work (and there wouldn't be time for all of them to give longer talks).

ii. Theme Problems. These were intended to be the centerpoint of the workshop activities. They were introduced by a very few carefully chosen talks, which were aimed at providing the basis of the following group discussion (as opposed to conventional talks on the speaker's current research). The theme problem for the cancelled Measurement and Modeling Workshop was to be "Causes of Burstiness". The theme problem for the Heavy Traffic Workshop was "Congestion Control, and what can be done in a network laboratory?" The theme problems for the Internet Tomography Workshop were "Validation of Tomographic Methods" and "Spatial Temporal data collection and analysis".

iii. Breakout Discussion Groups. These were the forum in which the main progress on theme problems was made. To maximize the desired three way collaborations, discussion groups were pre-assigned with the goal of deliberate balance. Observation of previous SAMSI discussion groups revealed that when the groups were large, only a relatively few members actually contribute much to the discussion. To create an environment where young people felt comfortable joining the discussion and asking questions (thus feeling that they are also "owners" of the process), groups of size about 10 were used. To ensure that discussion stayed on topic, and to allow useful summarization of the results, each group had a leader (generally a senior scientist), who was responsible for reporting that group's results to the full group. To help leaders organize their thoughts, they were given a transparency and pen for use in the final reporting session.

iv. Pie in the Sky Session on big picture ideas. Here speakers were requested to give brief presentations on problems that they had no idea how to solve (in contrast to what they had recently done, or what they were planning to do next).

Intended lasting benefits, from these workshops, in terms of human resources included raising the awareness of participants of the value of our three-way collaboration, and encouraging their engagement in it. Researchers of all levels were targeted, but because it seemed likely the largest long term impact would be on young researchers (from all three areas, and from both academics and industry), a large share of resources was devoted to supporting their attendance. Because of the organizing committee's concept that this SAMSI program should also draw additional statisticians into the field of network analysis, a number of relatively young researchers (with an emphasis on under-represented groups) with no background in network analysis were invited. Our target group here was people who have been recently promoted, with the idea that they were most likely to be interested in (and perhaps even looking for) new research areas.

Note: Extensive information about these workshops is available in the SAMSI Annual Reports for Years 2003-04 and 2004-05. In particular, these annual reports contain the programs, abstracts, and lists of workshop participants for each of the program workshops.

4. Working Groups

The mini-meeting that replaced the Measurement and Modeling Workshop led to a discussion process aimed at developing ideas on how to organize this diverse group of people into effective working groups. A listing of topics of interest suggested a rough grouping as "Model Based Topics" and "Method Based Topics". A leader was selected for each group, who was charged with organizing and leading meetings, and eventually reporting to the larger group. To keep things straight, SAMSI Postdoc Cheolwoo Park constructed and maintained the program's web page at:

http://www.samsi.info/200304/int/int-project.html

This collaborative research had the largest direct impact in terms of human resources, because these were the people with the chance to most deeply feel the value of our three-way collaboration. As is clear from the group leaders listed below, from the above web page (which include all of senior and junior faculty, through postdocs, to include even advanced graduate students), a wide variety of level of researchers was targeted and asked to join.

Details of these groups (Name, leader, members, group objectives) were:

4.1. Changepoints and Extremes, Richard Smith,

Robert Buche (North Carolina State University)
Fred Godtliebsen (University of Tromso, Norway)
Krishanu Maulik (EURANDOM, The Netherlands)
Cheolwoo Park (SAMSI)
Juhyun Park (University of North Carolina, Chapel Hill)
Haipeng Shen (University of North Carolina, Chapel Hill)
Murad Taqqu (Boston University)
Haakon Tjelmeland (Norwegian University of Science and Technology, Norway)
Zhengyuan Zhu (University of North Carolina, Chapel Hill)

Pictures of bursty internet traffic tend to show a number of characteristics:
(a) periods of bursty behavior interspersed with non-bursty behavior;
(b) seemingly sharp transitions from one to the other (though it is not clear just how sharp these transitions are);
(c) within a bursty period, a much greater than usual frequency of extreme observations, both above and below the overall mean.
The objective of this working group is to explore approaches in which the transitions between bursty and non-bursty behavior are considered changepoints, with models and methods from extreme value theory used to characterize the bursty periods. The changepoints form some version of a point process (the simplest model would assume a homogeneous Poisson process, but others can be considered) while the bursty periods are viewed as independent realizations of some kind of threshold-exceedance process whose parameters are allowed to vary from one bursty period to another in the manner of a hierarchical model. The structure is flexible enough to allow for a variety of alternative specifications, and can be fitted through the Reversible Jump MCMC paradigm. Finally, it was proposed that the fitted model(s) be used to calculate a variety of "indexes of burstiness" (the development of which should be considered part of the research program) and thus used to compare different internet servers and paradigms.

4.2 Formulation of Suite of Models, Zhengyuan Zhu

Jay Aikat (University of North Carolina, Chapel Hill)
Kevin Jeffay (University of North Carolina, Chapel Hill)
Steve Marron (SAMSI/University of North Carolina, Chapel Hill)
Jonathan Mattingly (Duke University)
Krishanu Maulik (EURANDOM, The Netherlands)
Cheolwoo Park (SAMSI)
Juhyun Park (University of North Carolina, Chapel Hill)
Surajit Ray (University of North Carolina, Chapel Hill)
David Rolls (SAMSI)
Haipeng Shen (University of North Carolina, Chapel Hill)

The goal was to develop useful statistical models for Internet traffic flow that are simple to analyze and simulate, and can capture the characteristics of actual traffic data that are important to electronics engineers and computer scientists. To achieve that goal, a top-down approach was combined with a bottom-up approach. The class of statistical models proposed was based on the fact that Internet traffic is an aggregation of individual connections, each with a random starting time, random duration (from heavy-tail distribution) and a random throughput. The bottom-up approach models the starting time, duration, throughput and their dependence structure by analyzing the data derived from the actual traffic flow, and studies the statistical property of the random process of aggregated packet counts thus obtained.
The working group addressed the following problems:

1. Identifying appropriate measures for evaluating how well a model fits the data (i.e., how to determine if model generated data is close to actual traffic data).
2. Finding good models under those measures.
3. Developing methodology to estimate the parameters of the model as well as making statistical inference.
4. Simulating traffic flow efficiently under such model.

4.3 Multifractional Brownian and Stable Motion, Stilian Stoev

Robert Buche (North Carolina State University)
Arka Ghosh (University of North Carolina, Chapel Hill)
Krishanu Maulik (EURANDOM, The Netherlands)
George Michailidis (University of Michigan)
Cheolwoo Park (SAMSI)
David Rolls (SAMSI)
Robert Wolpert

The fractional Brownian motion (FBM) has been a very successful model for traffic in modern telecommunication networks such as Ethernet-LAN and more generally, the Internet. FBM captures two major characteristic features of the network traffic: *time scale invariance* (statistical self-similarity) and *Long-Range Dependence* (LRD). A Gaussian stochastic process $X=\{X(t)\}$, $t>0$ is said to be FBM if it has mean zero, stationary increments and is self-similar, that is, for all a>0, the processes {X(at)}, t>0 and $a^H X(t)$, t>0, have equal finite-dimensional distributions. The parameter H belongs to the range (0,1) and is called the self-similarity parameter of the FBM process X. H is also the Hurst parameter of fractional Gaussian noise time series Y(k):=X(k)-X(k-1), k=1,2,... The FBM process can be also regarded as a physical traffic model. It appears in the limit of the superposition of independent ON/OFF sources with heavy-tailed ON and OFF periods, which mimic the flows in a busy network link. FBM is also the limit process of a physical infinite Poisson source model with heavy-tailed sources. Extensive studies of real traffic data, however, indicate that FBM alone cannot be used to explain the traffic burstiness (see http://www-dirt.cs.unc.edu/net_lrd/). Traffic burstiness appears to be a serious non-stationary effect in data, that cannot be contributed to seasonality or periodicity.

The goal of this group was to study, whether and how can the FBM model be augmented to account for non-stationarity effects (burstiness) in real data. Focus was on the so-called multifractional Brownian motion (MBM) model. The MBM processes Y={Y(t)}, t>0 extend the class of FBM processes by allowing the self-similarity parameter H to change with time, that is, H=H(t). More precisely, Y(t) is defined by replacing the parameter H in an integral representation of the FBM procsess by a function of time H(t), t>0. The resulting MBM process is Gaussian and locally self-similar, that is, Y(t) behaves, locally, like a self-similar process with self-similarity parameter H(t). The research focused on the following themes:

* Extending existing physical models (e.g. the infinite Poisson source model or the ON/OFF model) by obtaining the MBM as a stochastic-process limit.
* Exploring the connections between the infinite Poisson source model and the ON/OFF model.
* Validation of the models under consideration, by:

(a) developing techniques to estimate H(t), locally, by using novel exploratory tools such as Dependent SiZer and wavelet analysis.

(b) Estimating the parameters in the physical model from traffic data; experimenting with synthetic traffic data over a real network.

4.4 Structural Breaks, Vladas Pipiras

Robert Buche (North Carolina State University)
Fred Godtliebsen (University of Tromso, Norway)
Cheolwoo Park (SAMSI)
Stilian Stoev (Boston University)
Murad Taqqu (Boston University)

An increasing number of publications in the time series literature suggest that evidence of long-range dependence is an artifact of structural breaks. Structural breaks (also: level shifts, regime switching, structural instability) is one type of non-stationarity of a time series. The explanation for long-range dependence through structural breaks is particularly prevalent in Econometrics where evidence for long-range dependence was found in the time series of volatility of stock prices, inflation rates and other economic indicators. The goal of this working group was to explore the idea of structural breaks in the context of Internet traffic modeling where evidence of long-range dependence is alsoubiquitous. In particular, the working group focused around the following themes:

1. definition and types of structural breaks, models (some mixture, Markov-switching and other models),
2. structural breaks versus long-range dependence (possibly versus other phenomena),
3. statistical tests to discriminate between structural breaks and long-range dependence
4. applications to Internet traces.

4.5 Semi-experiments - Look & See, Juhyun Park

Ian Dinwoodie (Duke University)
Felix Hernandez Campos (University of North Carolina, Chapel Hill)
Kevin Jeffay (University of North Carolina, Chapel Hill)
Steve Marron (SMASI/University of North Carolina, Chapel Hill)
Jonathan Mattingly (Duke University)
Cheolwoo Park (SAMSI)
Surajit Ray (University of North Carolina, Chapel Hill)
Haipeng Shen (University of North Carolina, Chapel Hill)
Don Smith (University of North Carolina, Chapel Hill)

As a way of characterizing internet traffic, the semi-experimental approach proposed by Darryl Veitch leads to a new direction of thinking. In particular, it provides an interactive tool that helps narrow the gap between modeling and data analysis. With a laboratory at hand, this has a lot of potential to apply to study of traffic modeling. The main activities of this group were to explore various aspects of traffic data by adapting the semi-experimental approach to provide a basis for formal modeling.

4.6 Semi-experiments - Formal Testing, Mike Devetsikiotis

Ian Dinwoodie (Duke University)
Arka Ghosh (University of North Carolina, Chapel Hill)
Steve Marron (SAMSI/University of North Carolina, Chapel Hill)
Jonathan Mattingly (Duke University)
George Michailidis (University of Michigan)
David Rolls (SAMSI)
Zhengyuan Zhu (University of North Carolina, Chapel Hill)

The task of systematically characterizing the multitude of "factors" affecting traffic behavior and network performance, is a difficult but central one. Assessing the "effects" of such factors, ranging from qualitative statements, to full quantification, to (even better) sensitivity analysis, is particularly important. The approach of "semi-experiments" seems to present a refreshingly novel approach to this line of work. The objective of this working group was to engage in the careful and systematic study of such "semi-experiments" and attempt to

(a) formalize them in a statistical sense (i.e., relate them to design of experiments, factorial analysis, statistical confidence, hypothesis testing, etc.);
(b) combine them or extend them utilizing the group's own expertise in statistics, output analysis techniques, and network system design methodologies.

4.7 Testbeds - Lab Experiments, Don Smith,

Jay Aikat (University of North Carolina, Chapel Hill)
Felix Hernandez Campos (University of North Carolina, Chapel Hill)
Steve Marron (SAMSI/University of North Carolina, Chapel Hill)
George Michailidis (University of Michigan)
Cheolwoo Park (SAMSI)
David Rolls (SAMSI)
Haipeng Shen (University of North Carolina, Chapel Hill)

Networking research has long relied on simulation as the primary vehicle for demonstrating the effectiveness of proposed algorithms and mechanisms used in routers or TCP/IP protocols. Typically one constructs a network testbed in a laboratory and conducts experiments with actual network hardware and software (or one simulates network hardware and software in software such as the NS network simulator). In either case experimentation proceeds by simulating the use of the (real or simulated) network by a given population of users running applications such as FTP or web browsers. Traffic generators are used to inject application-level data objects into the network according to a model of how the applications or users behave. A critical aspect of this empirical methodology is ensuring that the resulting synthetic traffic, appearing as packets flowing through the network, preserves the essential characteristics of packet flows in real networks. An especially important property to study is the "burstiness" of packet-level traffic because it has been shown to have strong influences on many of the algorithms and mechanisms (e.g., active queue management) that networking researchers study. This working group considered two important questions:
(1) How should we measure and characterize both real and synthetic packet-level traffic so we can verify that synthetic traffic preserves all the essential properties of real traffic?
(2) Can we design controlled experiments using a testbed network to confirm various hypotheses and findings from other empirical studies about the physical factors that lead to "burstiness" in traffic?

4.8 SiZer and Wavelet, Cheolwoo Park

Fred Godtliebsen (University of Tromso, Norway)
Arka Ghosh (University of North Carolina, Chapel Hill)
Juhyun Park (University of North Carolina, Chapel Hill)
Stilian Stoev (Boston University)
Murad Taqqu (Boston University)

In an analysis of long range dependent time series, a Logscale Diagram using a wavelet method is quite useful. The Logscale Diagram is essentially a log-log plot of variance estimates of the wavelet details at each scale, against scale, complete with confidence intervals about these estimates at each scale. It can be thought of as a spectral estimator where large scale corresponds to low frequency. For example, one can estimate the Hurst Parameter from a Logscale Diagram by applying a weighted least square fit for a certain range of scales. SiZer enables meaningful statistical inference, while doing exploratory data analysis using statistical smoothing methods (e.g. histograms or scatterplot smoothers). It is a new visualization that brings clear and immediate insight into a central scientific issue in exploratory data analysis: Which features observed in a smooth of data are "really there"? This central question is critical in real data analysis, because discovery of a new feature, such as an unexpected "bump" or surprising "regions of decrease/increase", might lead to important new scientific insight. One common factor of these two tools is that they are looking the data at various scales. Combining these two tools to create a new tool for the analysis of long range dependent time series was the goal of this working group.

4.9 Heavy Traffic, Robert Buche

Arka Ghosh (University of North Carolina, Chapel Hill),
Chuan Lin (North Carolina State University)
Steve Marron (SAMSI/University of North Carolina, Chapel Hill)
Cheolwoo Park (SAMSI)
Juhyun Park (University of North Carolina, Chapel Hill)
Vladas Pipiras (University of North Carolina, Chapel Hill)

The Internet Heavy-Traffic working group investigated open questions posed by Ruth Williams, UCSD, in her talk on Modelling for the Internet Workshop and also described in the preprint by herself and Frank Kelly: "Fluid Model for a Network Operating under a Fair Bandwidth-Sharing Policy". In particular, the goal was to derive a workload – reflected diffusion characterizing a part of a network operating at capacity ("heavy-trafffic").

5. Graduate Student Participation

Arka Ghosh, (Statistics & Operations Research, UNC), was supported by SAMSI for the 2003-2004 year. He participated in all program workshops and was an active member of the working groups on Multifractional Brownian and Stable Motion, Semiexperiments – Formal Testing, SiZer and Wavelets, and Heavy Traffic. He presented a Postdoc – Grad Student Seminar.

Felix Hernandez Campos, (Computer Science, UNC), supported by SAMSI for the full 2003-2004 year. Participated in all program workshops. Active member of Comparison of Hurst Parameter Estimators, Semiexperiments – Look and See, Testbeds – Lab

Experiments working groups. Performed critical data base work, providing data for entire program.

Myung Hee Lee, (Statistics & Operations Research, UNC), supported by SAMSI for the Spring of 2004). Performed research on periodicities of flow arrival times within documents, and of round trip times. Started preliminary micro-array work, leading into Computational Biology Program next year.

Chuan Lin, (Operations Research, NCSU), not supported by SAMSI. Participated in Workshop on Heavy Traffic and Congestion Control. Active Member of Heavy Traffic working group.

Juhyun Park, (Statistics & Operations Research, UNC), supported by SAMSI for the full 2003-2004 year. Participated in all program workshops. Active member of Changepoints and Extremes, Suite of Models, Comparison of Hurst Parameter Estimators, SiZer and Wavelets, and Heavy Traffic working groups. Lead the Semiexperiments – Look and See working groups. Presented a Postdoc – Grad Student Seminar. Co-author on 2 papers in progress.

6. SAMSI Courses

Two courses were taught during the Fall Semester of 2003. Both were listed at all 3 Triangle Universities, and were attended by a wide (both in terms of background, and also universities) range of students, both enrolled and auditing. Both courses met one evening per week, in the NISS main Classroom.

6.1: *Data Statistical Analysis and Modelling of Internet Traffic Data*
INSTRUCTOR: J. S. Marron
COURSE DESCRIPTION: The analysis and modelling of internet traffic data represents an important major challenge for engineers, for computer scientists, for statisticians and for probabilists. Really new ideas and models are needed because heavy tailed distributions and long range dependence (both appearing at a number of different points) render standard methods, such as classical queueing theory, unusable. This course considers a variety of mnethods for understanding and modelling internet traffic at a variety of levels, from individual TCP traces, to monitoring traffic on a main link. An important underlying concept is cross scale views of data. Novel graphical views of data play an important role. To reach a broad audience, prerequisites are kept to a minimum, with needed foundational material, including Q-Q plots, time series analysis, long range dependence, and SiZer analysis being introduced as needed.
COURSE WEB PAGE: http://www.samsi.info/200304/int/traffic-course.html

6.2: *Long range Dependence and Heavy Tails*
INSTRUCTOR: Murad S. Taqqu
COURSE DESCRIPTION: This course will focus on long-range dependence and heavy tails, notions which are relevant in computer traffic networks. Long-range dependence

occurs when the covariances of a time series decrease slowly, like a power function. Heavy tails occur when the probability distribution of the time series has infinite variance and behaves like a power function. We will introduce self-similar processes which are idealized models that can encompass long-range dependence and/or heavy tails. We will focus first on fractional Brownian motion and on the related FARIMA time series models. To deal with infinite variance and heavy tails, we will introduce in a systematic fashion, infinite variance stable processes. We will study their properties and describe a number of stable (heavy-tailed) self-similar processes, including the so-called "Telecom model". We will also describe statistical methods for detecting the presence of long-range dependence and for estimating its intensity, focusing on wavelet methods since these are particularly useful in this regard.
COURSE WEB PAGE: http://www.samsi.info/200304/int/dependence-course.html

7. Research Papers

Research papers followed mostly from the working groups. They are not organized in that way here, because a number of them have roots in two or more working groups.

- Chang, H., A.Q. Fu, N.D. Le and J. SIDEC *"Designing Environmental Monitoring Networks for Measuring Extremes"* SAMSI Technical Report 2005-4

- Dinwoodie, I. H., Mosteig, E., Gamundi, E. *"Algebraic Equations for Blocking Probabilities in Asymmetric Networks"* Open Systems and Information Dynamics 12 (2005) 273-288 (with E. Gamundi and E. Mosteig.) SAMSI Technical Report 2004-15

- Gonzalez, B., Hernandez-Campos, F., Marron, J. S., and Park, C. *"Visualization Challenges in Internet Traffic Research"* To appear in Graphics of Large Datasets, Ed: A. Unwin, M. Theus, and H. Hofmann, Augsburg, Germany. (2006)
- Hernandez-Campos, F., Marron, J.S., Resnick, S.I., Park, C., and Jeffay, K. *"Extremal Dependence: Internet Traffic Applications"* Stochastic Models, 21(1), 1-35. (2005)

- Michailidis, Lawrence E., G. and Nair, V.N. *"Flexicast Delay Tomography"* To appear Journal of the Royal Statistical Society, Series B (2005)

- Michailidis, Lawrence E., G. and Nair, V.N. *"Local Area Network Analysis Using End-To-End Delay Tomography"* Performance Evaluation Review, 33, 39-45 (2005)

- Michailidis,Lawrence, E., G., Nair, V.N. and Xi, B. *"Network Tomography: A Review and Recent Developments"* To appear in Festschrift For Peter Bickel, J. Fan and H. Koul (eds), IMS Monograph Series, Hayward, CA (2005)

- Park, C., Godtliebsen, F., Taqqu, M., Stoev, S. and Marron, J. S. *"Visualization and Inference Based on Wavelet Coefficients, SiZer and SiNos"* Submitted to Computational Statistics and Data Analysis. SAMSI Technical Report No. 2004-10.

- Park, C., Hernandez-Campos, F., Marron, J. S., Rolls, D. and Smith, F. D. *"Long Range Dependence in a Changing Internet Traffic Mix"* Computer Networks, 48, 401-422. SAMSI Technical Report No. 2004-9.

- Park, C., Hernandez-Campos, F., Le, L., Marron, J. S., Park, J., Pipiras, V., Smith, F. D., Smith, R. L., Trovero, M., and Zhu, Z. *"Long Range Dependence Analysis of Internet Traffic"* Submitted to Statistical Science. (2004)

- Park, C., Hernandez-Campos, F., Marron, J. S., Smith, F. D., and Jeffay, K. *"Correlations of Size, Duration, and Rate: in TCP Connections: the Case Against"* Submitted to Computer Networks. (2006)

- Park, C., Marron, J. S. and Rondonotti, V. *"Dependent SiZer: Goodness of Fit Tests for Time Series Models"* Journal of Applied Statistics, 31, 999-1017

- Rolls, D., Michailidis, G. and Hernandez Campos. F. *"Queueing Analysis of Network Traffic" Computer Networks, 48, 447-473*

- Rondonotti, V., Marron, J.S., and Park, C. *" SiZer for Time Series: A New Approach to the Analysis of Trends"* Under revision, Submitted to Journal of Time Series Analysis. (2004)

- Stoev, S., Taqqu, M., Park, C. and Marron, J. S. *"LASS: a Tool for the Local Analysis of Self-Similarity"* Computational Statistics and Data Analysis, 50, 2447-2471. SAMSI Technical Report No. 2004-7.

- Stoev, S., Taqqu, M., Park, C. and Marron, J. S. *"Strengths and Limitations of the Wavelet Spectrum Method in the Analysis of Internet Traffic"* SAMSI Technical Report No. 2004-8.

- Stoev, S., Taqqu, M., Park, C., and Marron, J. S. *"On the Wavelet Spectrum Diagnostic for Hurst Parameter Estimation in the Analysis of Internet Traffic"* Computer Networks, 48, 423-445. (2005)

- Xi, B., Michailidis, G. and Nair, V.N. *"Estimating Network Internal Losses using a New Class of Probing Experiments Schemes"* To appear in Journal of the American Statistical Association (2005)

- Xu, P., Devetsikiotis, M. and Michailidis, G. *"Adaptive Scheduling using Online Measurements for Efficient Delivery of Quality of Service"* SAMSI Technical Report No. 2004-12.

- Xu, P., Devetsikiotis, M. and Michailidis, G. *"Online Scheduling for Resource Allocation of Differentiated Services: Optimal Settings and Sensitivity Analysis"* SAMSI Technical Report No. 2004-13.

- Xu, P., Michailidis, G. and Devetsikiotis, M. *"Large Profit-Oriented Resource Allocation Using Online Scheduling In Flexible Heterogeneous Networks"* To appear in Telecommunications Systems (2005)

- Xu, P., Michailidis, G. and Devetsikiotis, M. *"Online Scheduling For Optimal Resource Allocation In Flexible Heterogeneous Networks"* Proceedings of the 39th Conference on Information Sciences and Systems, Baltimore, MD (2005)

- Zhu, Z. and Taqqu, M. S. *"Impact Of The Sampling Rate On The Estimation Of The Parameters Of Fractional Brownian Motion"* SAMSI Technical Report No. 2004-14.

Reports in Preparation

- Abry, P. and Pipiras, V. *"Wavelet-Based Synthesis Of The Rosenblatt Process"*

- Park, C., Hernandez-Campos, F., Marron, J. S. and Jeffay, K. "Thresholded Log-Log Correlation Analyses of TCP Characteristics"

- Park, C., Veitch, D., Shen, H., Hernandez-Campos, F., and Marron J. S. "Semi Experiment Analysis Of The Shifting Knee Wavelet Spectrum"

- Park, C., Hernandez-Campos, F., Marron, J.S., Shen, H., and Veitch, D., *"Capturing the Elusive Poissonity in Web Traffic"* In preparation.

- Park, C. and Kang, K. *"Sizer Analysis For Nonparametric Comparison Of Several Regression Curves"* In preparation.

- Park, J. and Park, C., *"Robust Estimation of Hurst Parameter and Selection of an Onset Scaling"* In preparation.

- Park, J. and Park C. *"Robust H Estimation, Automatic Choice Of Parameters"*

- Pipiras, V. *"On The Use And Usefulness Of Wavelet-Based Simulation Of Fractional Brownian Motion"*

- Piparas, V. and Taqqu, M. S. *"Identification Of Periodic And Cyclic Fractional Stable Motions"*

- Piparas, V. and Taqqu, M. S. *"Integral Representations Of Periodic And Cyclic Fractional Stable Motions"*

- Piparas, V. and Taqqu, M. S. *"Semi-Additive Functionals And Cocycles In The Context Of Self-Similarity"*

- Smith, R. L., Taqqu, M., Shen, H., Park, J., Zhu, Z. and Park, C. "Change Points and Long Range Dependence"

- Zhu, Z., Shen, H., Park, C., Hernandez-Campos, F and Marron, J. S."Shot Noise Model, Start Times, Micro-Bursts"

8. Continuing Collaboration

In addition to many new individual collaborations that started during the program (as indicated by co-authorship of the papers listed above), a larger scale effort is also currently under way, through a planned NSF IGERT proposal. Main partners on this proposal include many of the research partners during the SAMSI Program, and also a number of the most active participants during the Workshops.

The main goal of the IGERT proposal, named the Internet Statistics Education and Research Consortium (ISERC) is to continue to foster the special collaboration between statisticians, probabilists and network researchers that happened in the SAMSI program.

To achieve this goal, ISERC will depend upon leaders from all of these disciplines. Because the key players are geographically distributed, ISERC is envisioned as a Consortium (in contrast to the more common Center approach to collaborative research). The major activity of ISERC will be an annual research summer month, where travel expenses of all participants will be paid so they can come to NISS to work in an atmosphere aimed at reproducing and multiplying the strong inter-disciplinary collaborations created by the SAMSI program. The SAMSI pilot effort has clearly proven the success of this approach in terms of both research and education.

As part of the educational component, ISERC cross-disciplinary course development will be encouraged (with progress reports featured at ISERC workshops). The challenge of doing this in a distributed fashion will be addressed using the Connexions collaborative approach to online learning (see http://cnx.rice.edu/), headed by ISERC Partner Richard Baraniuk.

Another goal of ISERC will be the collection, and web posting of high quality publicly available data sets. The value of such data sets, for a wide variety of purposes, became clear from several of the SAMSI workshops. Ongoing SAMSI activities include pilot data sets of this type, but there will be large value added from ISERC doing this on a much large scale (straightforward because of the geographical spread of the partners).

8. Program Assessment and Summary of Lessons Learned

In summary, the SAMSI Program on Network Modeling for the Internet was quite successful in terms of its main goal of establishing new interdisciplinary contacts. Less success was achieved on the goals of combining with industry, and internet posting of high quality data sets.

The collaborative efforts were a joy to behold. The excellent attitudes, and willingness to explore new research directions by the on site participants, gave a clear success in this direction. This was clearly the biggest success of the program.

The attempted contacts with industry took two major directions. Contacts with local industry fell through, perhaps because of the recent crash in the telecommuncations industry. There were massive layoffs (even complete company location closures) at a very important time, so people that were expected to be interested were too distracted by other matters. Contacts with nationwide industry was attempted through workshop attendance, and this generally much more successful. Industry attendees reported a very positive experience, but we were not successful in engaging them in on-going activites. It is not clear that alternative strategies could have helped, given the unfortunate economic timing.

The development of high quality data sets, for internet posting has not been successful to date. Serious efforts were made in two directions: an internet tomography data set, and a set of simultaneous tie series of router data. While both may still actually be posted, progress has been disappointingly slow. In retrospect, a central problem is that the "owners" of both projects were volunteer faculty (unsupported by SAMSI), which seemed to make sense at the time, as they were the most interested people. However, things moved too slowly, other interests got in the way, and momentum was clearly lost. A clear lesson is that such projects require an "owner" who has serious obligation to SAMSI.