Proposal to the Department of Energy, Office of Science

Program Notice DE-FG01-04ER04-03

High-Performance Network Research:
Scientific Discovery through Advanced Computing (SciDAC) and
Mathematical, Informational, and Computational Sciences (MICS)

Contact: Thomas Ndousse

INCITE *Ultra* – New Protocols, Tools, Security, and Testbeds for Ultra High-Speed Networking

---

Richard Baraniuk (PI), Edward Knightly, Rolf Riedi
Rice University
{richb, knightly, riedi}@rice.edu
tel: 713-348-5132, fax: 713-348-6196

Robert Nowak, Paul Barford
University of Wisconsin-Madison
nowak@engr.wisc.edu, pb@cs.wisc.edu
tel: 608-265-3914, fax: 608-262-1267

Wu-chun Feng, Mark Gardner
Los Alamos National Laboratory (LANL)
{feng, mkg}@lanl.gov
tel: 505-665-2730, fax: 505-665-4934

Les Cottrell, Jiri Navratil
Stanford Linear Accelerator Center (SLAC)
{cottrell, jiri}@slac.stanford.edu
tel: 650-926-2523, fax: 650-926-3329

This document fuses ***four proposals*** from an integrated team of Rice University, University of Wisconsin, Los Alamos National Laboratory (LANL), and Stanford Linear Accelerator Center (SLAC).

While aimed at the needs of the *DOE MICS Base Program on Ultra High-Speed Network Engineering*, this research also addresses many of the open issues at the core of the SciDAC Integrated Experimental Ultra High-Speed Networks Program.

Table of Contents

A. Project Summary

INCITE *Ultra* – New Protocols, Tools, Security,and Testbeds for Ultra High-Speed Networking

Richard Baraniuk, Edward Knightly, Rolf Riedi
Rice University

Robert Nowak, Paul Barford
University of Wisconsin-Madison

Wu-chun Feng, Mark Gardner
Los Alamos National Laboratory (LANL)

Les Cottrell, Jiri Navratil
Stanford Linear Accelerator Center

Advances in optical transport technology are yielding high-speed switches with astounding capacity. Simultaneously, distributed scientific computing applications are insatiably demanding ever-increasing bandwidth. The goal of the INCITE *Ultra* project is to develop the protocols and application tools that will provide the vital bridge between application demands and the vast transport capacity of ultra high-speed networks. Using a combination of protocol design, theoretical modeling, and testbed implementation, this project will provide:

1. a TCP protocol suite that scales to ultra high-speed networks while simultaneously achieving fairness and priority objectives,

2. scalable monitoring tools that enable grid applications to robustly and accurately predict and optimize performance,

3. DoS and worm-resilient protocols that ensure that the abundant resources of ultra high-speed networks are not used maliciously, and

4. testbed experimentation that provides validation and proof-of-concept of our designs, as well as valuable experience with real-world applications.

To address these challenges, we have assembled a team spanning four institutions with inter-disciplinary expertise in the fields of high-speed networking, high-performance computing, statistical signal processing, and applied mathematics. Our team has an extensive track record both individually and collaboratively as described in our accomplishments from our current DOE-sponsored INCITE project (incite.rice.edu).

B. Project Description

Motivation and Significance

The notion of a *computational grid* [Steve97,Foste99] has emerged as a way to ubiquitously provide distributed computing resources to high-performance scientific applications around the world. The central premise of the Grid is that by deploying advanced services over a high-performance TCP-based network, the grid can provide access to computing resources to the masses, irrespective of physical location. Access to such a cornucopia of computing resources can then provide the necessary computational cycles to more quickly solve "Grand Challenge Applications" such as climate modeling, human genome sequencing, and hydrodynamics.

Today, designers of the grid have identified "the network" as the fundamental bottleneck, particularly in support of bandwidth-hungry and latency-sensitive applications such as remote visualization. Unfortunately, removing this bottleneck is not merely a matter of upgrading link capacities. Indeed, the critical challenge to enable high-performance grid computing is to design protocols and tools such that applications can harness the vast communication capabilities of emerging optical transport technologies such as those provided by DOE's UltraScience Net [Rao03a,b]. While fully-meshed optical light-paths as provided by the UltraScience Net do remove difficult-to-predict effects of packet-based statistical multiplexing in the network core, three fundamental networking challenges remain: 1) How to design a TCP that can fully and fairly exploit the vast capacity of 10's of Gb/sec wide-area links? 2) How to design inference tools that will aid applications and TCP given a mix of a dynamically circuit-switched core and packet-switched edges? and 3) How to ensure that malicious users cannot utilize ultra-high-speed networks to launch an ultra-devastating denial-of-service attack?

We will design, implement, and develop the theoretical underpinnings of the key building blocks for realizing ultra-high speed end-to-end grid communication via the following four inter-related thrusts (see Figure 1):

1. **Ultra high-speed protocols.** We will design an ultra high-speed TCP suite with four landmark properties: (i) *dynamic right sizing* that exploits passive probing to set a high initial flow control window and adapt to RTT variation; (ii) *fair and prioritized TCP* that enables TCP flows to share bandwidth via application-specific performance objectives; (iii) *parallel download TCP* that enables a receiver to simultaneously download data from multiple replica sites.
2. **Scalable monitoring tools.** We will design scalable monitoring tools that aid applications and TCP in assessing a network's available bandwidth in ultra-high speed networks. Our tools will use end-to-end inference together with accurate and robust traffic models to characterize end-to-end paths containing both packet-switching elements and dynamically configured optical circuits.
3. **Scalable cybersecurity.** We will develop endpoint tools that will ensure that cheaters, misbehavers, and DoS attackers cannot impede the performance of grid applications. Our tools will detect and thwart malicious behavior via the development of (i) robust transport protocol designs that ensure that malicious nodes cannot exploit TCP vulnerabilities, (ii) end-point enforcement to ensure from a host that the other end of the flow is not cheating, and (iii) automated worm detection to ensure that ultra-fast networks do not provide a tool for efficient worm delivery.
4. **Testing and deployment on the DOE UltraScience Net.** We will utilize the LANL 10GigE Testbed and the Wisconsin Advanced Internet Lab as platforms for proof-of-concept implementations and experiments. With these controlled but state-of-the-art environments, we will validate and tune our designs as a stepping stone to wide-scale deployment. We will then deploy our protocols and tools on the UltraScience Net and integrate them with grid applications. Our objective is to establish the world-speed record for high-speed end-to-end data transfer without sacrificing fairness and priority objectives, nor DoS resilience.

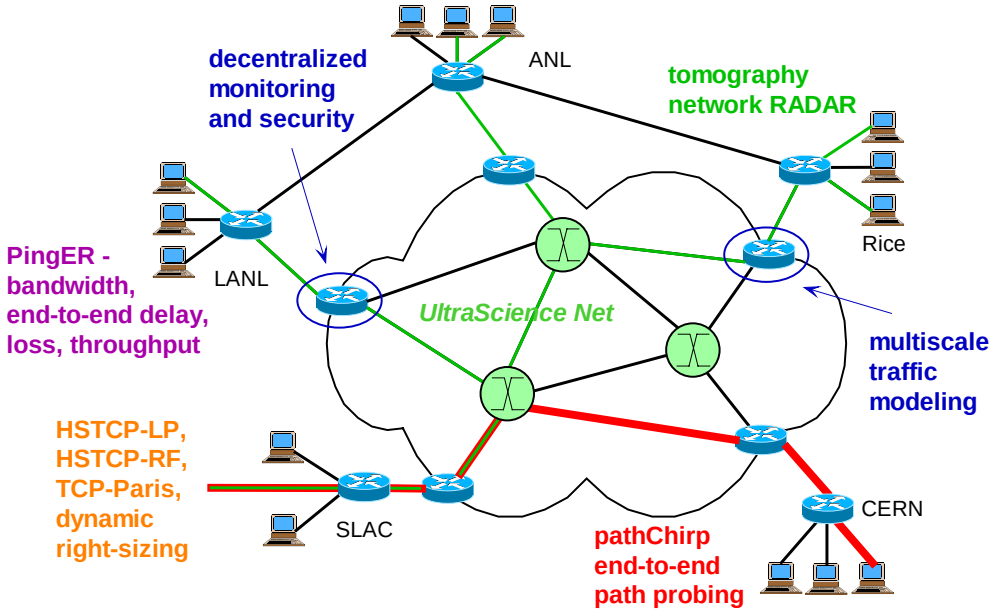

Figure 1: Relationships between the research thrusts.

To achieve these objectives, we have assembled an inter-disciplinary team with expertise ranging from theoretical traffic modeling and protocol design and deployment to supercomputing. Our team leverages past success and collaboration from the DOE INCITE project (2001-2004, incite.rice.edu). Moreover, we leverage the tools, protocols, and testbed infrastructure that we have previously developed, including MAGNET, PingeER HSTCP-LP, pathChirp, and NetTomo. These prior results not only provide a valuable starting point for achieving the above goals, but have already made a significant impact in high-performance networking as detailed in our DOE INCITE progress reports.

Thrust 1: Ultra High-Speed Protocols

Recent breakthroughs in high speed TCP protocol design have demonstrated the feasibility of having a cross-country file download at a blazingly fast pace in the Gb/sec range. Two challenges remain for the future: first, how to enable other TCP flows to share bandwidth with ultra-fast flows according to the desired fairness or application priority objectives? The second challenge is: how to go even faster?

We will develop an ultra high-speed TCP protocol suite with the following break-through characteristics. First, we optimize data transfer using sophisticated passive probing and statistical analysis. Second, we enable a broad class of performance and fairness objectives while simultaneously achieving high utilization of ultra-high speed links. Third, we utilize parallel download in order to dramatically reduce file transfer time for applications with replicated data.

1.1 Dynamic right-sizing and bandwidth prediction

TCP relies on two mechanisms to set its transmission rate: a flow-control window that regulates how much data can be in-flight so as not to overrun the receiver's buffer, and a congestion-control window that regulates how fast data is injected into the network according to congestion. While the congestion-control window varies dynamically as the network state changes, the flow-control window is static in

most TCP implementations. Furthermore, the default size of the static flow-control window (usually 64 KB) is insufficient for contemporary high-speed wide-area networks (WANs) such as ESnet or UltraScience Net. To maximize the transfer rate, the flow-control window should be equal to the bandwidth-delay product (BDP) of the network. While manually tuning is possible as detailed in [RFC2488, PSC, Tiern01a], it is a tedious process. Moreover, while new services can indeed provide a coarse estimate of BDP (e.g., AutoNcFTP [Liu00], Web100 [Web100], Enable [Tiern01b] and Network Weather Service [NWS]) they approximate conditions only at connection set-up time, resulting in significant inefficiencies when the BDP changes. Figure 2 illustrates how the BDP dynamics between Los Alamos and New York can vary by as much as two orders-of-magnitude over 20 s intervals.

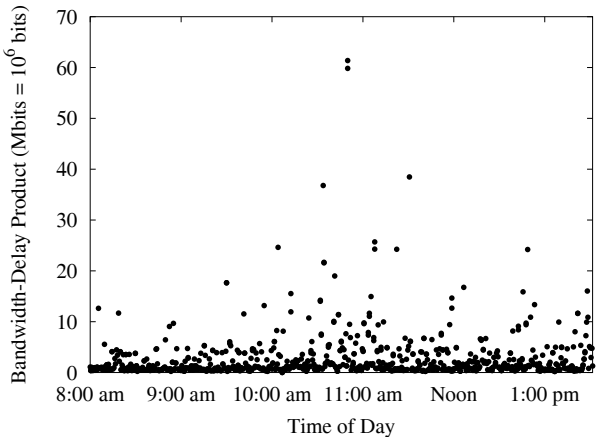

Figure 2: Bandwidth-delay product at 20 s intervals for a Los Alamos-New York path.

We will develop a *dynamic right-sizing* (DRS) protocol that adapts the flow control window to the BDP via robust and passive endpoint measurements. To date, we have demonstrated 7x average and 30x maximum speedup via kernel DRS implementation and 4x to 6x speedup for user-level implementation. Yet, while current implementations of DRS do an excellent job of estimating the currently available bandwidth along a path, they are purely reactive. The bandwidth delivered over the last RTT is used as an estimate of the available bandwidth for the next RTT, a methodology that fails in highly bursty environments. To address the challenge of the vast traffic dynamics expected for grid computing applications, we will develop a DRS that uses packet path *chirps* (see Task 2.1 below) [Rib03a] and multifractal wavelet modeling [Riedi99,Rib03b,Rib04,Abry02] as an accurate and robust predictor of future BDP. (A packet chirp is train of packets with an exponentially increasing flight pattern. Using the information obtained by path chirps, a multi-fractal wavelet model of the cross traffic can be constructed that predicts the amount of cross traffic on the bottleneck link.) Armed with available bandwidth predictions, DRS can anticipate cross traffic and size the flow-control window appropriately. The result is higher throughput for scientific applications.

The ability to accurately predict available bandwidth has profound implications on TCP protocol design. First, observe that today's AIMD congestion-control algorithm in TCP systematically increases its sending rate until the available bandwidth is exceeded, packets are dropped, and the sending rate is cut in half. The result is the characteristic sawtooth transfer-rate pattern of TCP. With an accurate prediction of available bandwidth, DRS can be modified to size the flow-control window just below the congestion-control window rather than at or above it. Instead of exceeding the available bandwidth and cutting the sending rate back, the sending rate stays just below the available bandwidth. Consequently, the transfer rate does not exhibit a sawtooth pattern and the throughput dramatically improves. Second, the time consuming "slow-start" phase of TCP can be altogether eliminated via an accurate initial estimate of available bandwidth, thereby yielding significant throughput and delay improvements especially for short-

lived flows. Finally, while exploiting the information gained by path chirps allows the available bandwidth to be computed accurately, the *active* probing currently used by path chirps introduces additional network load that may itself trigger network congestion. Our methodology is to embed path chirps into the data stream of the connection itself. Consequently, by *passively* inferring available bandwidth, we eliminate the bandwidth consumed by the chirps, increase the bandwidth available to applications, and reduce the risk of triggering congestion.

1.2 High-speed but fair TCP

State-of-the-art TCP protocols designed for high-speed networks cannot be incrementally deployed in the Internet because they are incompatible with TCP-Reno, the default TCP on most Internet hosts. Packet-drop-based protocols such as HSTCP [Flo03] are demonstrably unfair to TCP-Reno on high-speed paths [Sou] despite the fact that fairness is supposedly designed in. A likely cause for this incongruity is the conventional wisdom driving today's TCP designs that the background packet drop probability is time invariant. Quite to the contrary, packet losses, which signal congestion, actually have an erratic bursty nature. One potential solution is to migrate away from congestion control based on packet drops to control based on packet delay. Unfortunately, advanced delay-based protocols such as FAST [Fast04] are starved of bandwidth when competing with a number of TCP-Reno streams: Since packet delays indicate congestion much sooner than packet drops, delay-based schemes react quicker to the onset of congestion than drop-based schemes like TCP-Reno.

We will develop and implement HSTCP-RenoFair, or HSTCP-RF for short, that efficiently exploits the bandwidth available on high-speed paths while being fair to competing TCP-Reno connections. HSTCP-RF is a drop-based protocol that is a hybrid combination of HSTCP and TCP-Reno. HSTCP-RF's congestion window update rules are identical to those of HSTCP in the absence of significant queuing delay and to those of TCP-Reno otherwise. Thus, on uncongested high-speed paths (where there is no queuing delay), HSTCP-RF behaves like HSTCP and utilizes the abundant available bandwidth. On congested high-speed paths (either due to competing TCP-Reno flows or HSTCP-RF flows), HSTCP-RF behaves like TCP-Reno, which ensures fairness.

A preliminary experiment (HSTCP-RF emulated over UDP) running from SLAC to a host on apan.net in Japan indicates the promise of this new protocol. The link had a bottleneck capacity of approximately 1 Gb/s, a minimum round trip time of approximately 114 ms, and an observed packet drop rate during the transfer of approximately 10^{-6}. HSTCP-RF achieved a sustained throughput of 600-650 Mbps while causing the average round trip time to increase only by 3 ms and the drop rate to increase mildly to 10^{-4}. TCP-Reno streams running on this same link were unable to reach higher than 60 Mbps. A more limited second experiment showcases the promising fairness properties of HSTCP-RF. An emulated TCP Reno stream and an emulated HSTCP-RF stream were sent from hosts at Rice to a host at wisc.edu. The two streams achieved bandwidth shares of approximately 30 Mb/s each. When the HSTCP-RF stream was replaced by a Scalable TCP stream, the bandwidth share went to 60 Mb/s for Scalable TCP and only 10 Mb/s for TCP-Reno. The link had a minimum round trip time of 43 milliseconds and a bottleneck capacity of 100 Mb/s. The delay sensitivity of HSTCP-RF allows it to detect that it is competing with TCP-Reno and to reduce its aggressiveness. More experiments are currently underway.

1.3 High-speed prioritized TCP

The classical TCP design problem focuses on *fairness*, that is, how to design a distributed algorithm that can fairly divide network resources via endpoint control. However, fairness is not always the appropriate goal: a file transfer application may seek as high bandwidth as possible but may not wish to impede interactive web traffic; a distributed super-computing application may require near real-time transfer of

large data sets. Thus, it is increasingly apparent that *flexibility* is required in determining an application's priority and fairness objectives.

We have recently developed TCP Low Priority (TCP-LP) [Kuz03a], a distributed algorithm whose goal is to utilize only the excess network bandwidth as compared to the "fair share" of bandwidth as targeted by TCP. Moreover, we recently developed a high-speed version of the protocol termed HSTCP-LP [Kuz03b] with the goal of having both TCP-LP's ability to yield to cross-traffic and HSTCP's [Flo03] efficiency in utilizing high-capacity links. The key mechanisms unique to TCP-LP and HSTCP-LP congestion control are the use of one-way packet delays for early congestion indications and a novel TCP-transparent congestion avoidance policy. Our results have shown that: (i) HSTCP-LP is largely non-intrusive to TCP traffic; (ii) both single and aggregate HSTCP-LP flows are able to successfully utilize excess network bandwidth; moreover, multiple HSTCP-LP flows share excess bandwidth fairly; (iii) substantial amounts of excess bandwidth are available to low-priority class, even in the presence of "greedy" TCP flows; and (iv) response times for interactive flows are dramatically improved when bulk flows use HSTCP-LP.

Yet, TCP-LP represents only one point in a large space of fairness and priority objectives, i.e., it realizes low-priority service for a reference two-level strict-priority system. Thus, we will design and implement a flexible TCP that is able to realize different priorities, performance objectives, and fairness profiles via a purely endpoint protocol. Our methodology is to have different end-points apply different delay thresholds and to use different response functions to achieve service differentiation. However, given that many high bandwidth-delay-product (BDP) paths may be significantly under buffered [Kuz03b], the primary focus of our research will be to develop and explore protocols with novel response functions for high BDP networks. Thus, our design will ensure that achieving high speed and full utilization of ultra-high speed links is not mutually exclusive with diverse fairness and performance-differentiation objectives.

1.4 TCP-Paris: Parallel download for ultra-fast file transfer from replicas

Downloading data in *parallel* from multiple replicas to a single server has the potential to significantly reduce the download time of a large file. For example, as depicted in Figure 3, by simultaneously downloading disjoint parts of a file from multiple replicas, the download time can be reduced by a factor of up to $1/n$ in a homogeneous environment with n replicas.

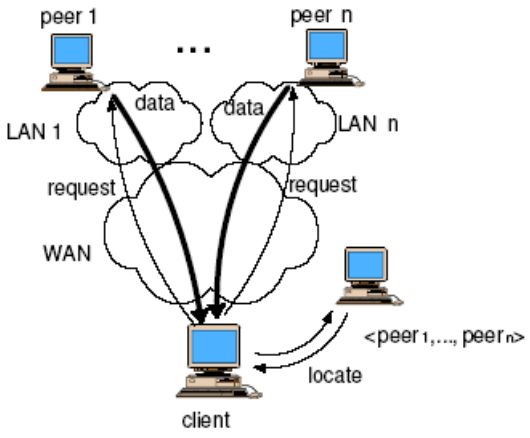

Figure 3: Illustration of a parallel download.

The key challenges for parallel download arise in a dynamic and heterogeneous environment. In this case, the volume of traffic downloaded from each replica must be matched to the available server and network

resources; otherwise a single slow peer can significantly reduce the total download time. Moreover, as server and network resources vary in time, the per-replica volume must be continuously adapted to ensure a match of replica transfer volume to the available resources. Finally, the download must be coordinated to ensure that the partition of the file among replicas is non-overlapping, and this coordination must occur efficiently to ensure that overhead in partitioning resources does not overwhelm the potential performance advantages of parallel download.

To download a file from a single server, today's approaches focus on how to select the "best" server that will result in the fastest download (e.g., [CC97,FBZA98,GS95b,KLL+97,STA01]). While such techniques can reduce delay as compared to systems without replication or to systems that select a random replica, a recent study shows that replica selection is not always successful in finding the best delay-minimizing server: even with a combination of different probing techniques, such as RTT-based probing or bandwidth probing, only 40-50% of the downloads managed to correctly choose the delay-minimizing server [NCR+03]. Moreover, such protocols do not exploit the above benefits of parallel download. Hence, while replica selection protocols may be successful for delay-sensitive applications with small files, they have limited benefits for large-file downloads.

Likewise, today's *parallel* download techniques have a number of significant shortcomings that make them unsuitable for high-performance applications. In particular, *fileslicing* is widely employed in current peer-to-peer systems such as KaZaa. Here, a large file is split into multiple equally-sized slices by the application [RB02,RKB00]. The client can download multiple slices in parallel and can download subsequent slices from the same peer when a slice completes from a fast peer before slower peers. Unfortunately, slicing protocols face an inherent problem in selecting the slice size. On the one hand, large slices limit the ability to adjust the amount of data that is downloaded from each peer and lead to situations in which a single slow peer dominates the download time. On the other hand, small slice sizes, while more responsive to dynamic conditions and client heterogeneity, result in excessive per-slice overhead. This overhead occurs because the application-layer slicing and client-selection protocols require that a new connection be established for every slice, resulting in a connection setup overhead for the initial handshake and reduced throughput due to repeated slow start phases. Moreover, information must be exchanged between the receiver and sender to determine which slice to transmit. While it may appear that a "moderate" size slice may therefore be ideal, no single size can optimally serve replicas with widely varying link capacities and loads. Moreover, the optimal slice size varies in time as network and peer loads change.

We will design, implement, and perform an Internet measurement study of *TCP-PARIS*, a multi-point to point transport-layer protocol (PArallel download protocol for ReplIcaS). Because the ideal partitioning of the transfer volume among servers is a dynamic and difficult-to-predict function of network conditions and server load, TCP-PARIS continuously adapts the transfer rates to available resources to best approximate the optimal partitioning of transfer volume to replicas. Moreover, unlike static, small-slice approaches that continuously incur additional round-trip-time overhead and slow-start phases, TCP-PARIS achieves low overhead by piggybacking partition information in TCP packets sent between receivers and senders. Hence, we will design TCP-PARIS to be a flexible and efficient parallel download protocol that adapts per-server download rates and volumes according to both long-term server heterogeneity and temporal dynamics of network and server congestion. Preliminary simulation results indicate that TCP-PARIS can achieve download time reductions of up to 52% as compared to slicing protocols.

Thrust 2: Scalable Monitoring Tools

Network capacities have increased dramatically in the last several years, rendering many current end-to-end monitoring methods ineffective and inaccurate. For example, to make a stable bandwidth measurement on a 1Gb/s link using the iperf tool requires about 60 s and 6 Gb of data. Indeed, current Abilene traffic statistics indicate that iperf measurements account for more than 5% of all traffic on the network. We will develop new application level tools for measuring traffic, estimating available bandwidth, and for network tomography that are especially designed for ultra high-speed networks. Our ultimate goal is to obtain continuous real-time network conditions (delay, loss, available bandwidth, jitter) estimates along multiple paths from the edge of the network.

2.1 Efficient and scalable end-to-end path inference with pathChirp

Knowledge of the *available* or unused bandwidth on an end-to-end path is critical for improving the performance of high-energy physics applications such as grid computing and those requiring large file transfers. With information on available bandwidth, several important tasks become feasible such as predicting the duration of large file transfers, fine tuning TCP window parameters to achieve higher throughput and improve system resource usage, optimally choosing paths for data downloads, and selecting optimal paths to route data in overlay networks.

Applications can further benefit from knowing not just the available bandwidth on a path, but also the location of the available bandwidth-scarce or *tight* links in the network over space and time. Information as to whether or not different paths share a common tight link can benefit applications downloading data from multiple locations. Tight link localization also enables network troubleshooting as well as helps identify the causes of network congestion. We term the estimation of the available bandwidth on paths and the localization of tight links in space and over time as *spatio-temporal* available bandwidth estimation.

We will dramatically enhance *pathChirp* [Rib03a,Rib04] for application to ultra high-speed networks. (pathChirp was developed within the current DOE INCITE project and won the Best Student Paper Award at PAM 2003.) PathChirp exploits the self-induced congestion principle and a novel "packet chirp train" of exponentially spaced packets to accurately estimate available bandwidth on end-to-end network paths i.e., a chirp probe train probes a network path at successively higher rates over time.

A significant advantage of chirp trains over alternative probing schemes, such as trains with equally spaced packets, is their *efficiency*; they provide accurate estimates of available bandwidth using very few packets, thus keeping the probing traffic load minimal. Indeed, pathChirp balances a "probing uncertainty principle" between the accuracy of an available bandwidth estimate and the number of probe packets involved in making the estimate. High-speed Internet and Gigabit testbed experiments with pathChirp indicate that it can estimate the available bandwidth on an end-to-end path using about *ten times* fewer packets than current techniques (path load, for instance [Jai02]). Because of its low probing traffic load, pathChirp can be deployed widely for network monitoring purposes without congesting the network. Figure 3 illustrates that pathChirp can accurately track available bandwidth fluctuations along one network path (SLAC-to-Rice) resulting from injecting cross-traffic into another (Caltech-to-Rice).

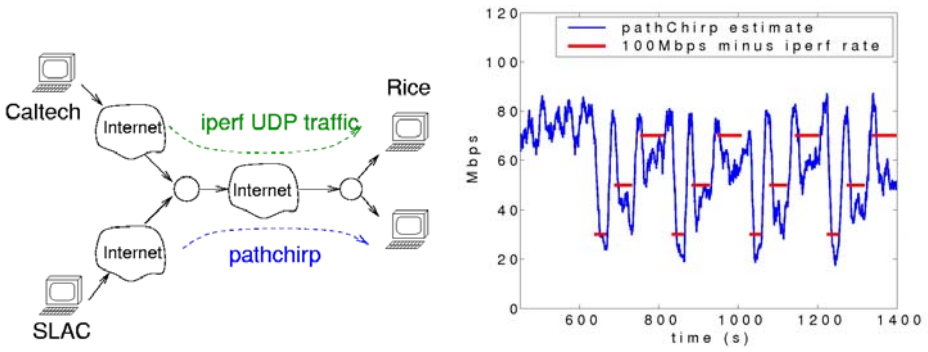

Figure 3: pathChirp available bandwidth tracking experiment.

Ultra high-speed networks present several significant challenges to an estimation technique based on self-induced congestion. First, to deal with the system I/O bandwidth limitations inherent in ultra high-speed networks, we have developed a promising preliminary extension based on packet tailgating and multiple probing sources [Rib03c]. A *packet tailgating chirp* is a chirp with every packet replaced by one large packet followed closely by a small one. The large packet exits the network at the last router on the path due to a limited IP TTL field while the small packets go through to the receiver. This alleviates the problem of system I/O bandwidth constraints on the end-host receiving probe packets. By synchronizing the probing trains from several hosts we propose to increase the net probing rate we inject into the network thereby alleviating the problem of system I/O bandwidth on the end hosts sending probe packets into the network. Additional issues we will address include more accurate software clocks, interrupt coalescence, and router architecture [Tie03].

Second, we will develop a spatio-temporal version of pathChirp that can locate (in space) and track (in time) a network path's tight link (see [Rib03c] for promising preliminary results). Finally, as described in Task 1.1, we will show how pathChirps can be integrated with a transport protocol to quickly converge to the initial available bandwidth (removing the inefficient slow start phase) and to dynamically right-size the flow control window.

2.2 Network RADAR: Network tomography using end-to-end measurements

Tomography is a powerful method for measuring and analyzing link specific characteristics using end-to-end active probes. This capability is important since link specific information such as delays and losses are otherwise only available to network administrators who have direct access to those links. Prior work has established the basic mechanisms for the use of tomographic inference techniques in the networking context [Tsa03,Pad03,Chny02,Ada00,Har00,Cac99]. However, the measurement methods described in all prior network tomography studies require infrastructure that is largely unavailable or that limits the scope of the paths over which the measurements can be made.

The general objective of our research is to develop and evaluate a new network tomographic technique for measuring link delays that can be used much more widely than prior techniques, including our successful *NetTomo* tool developed in the DOE INCITE project. We will investigate a new network tomographic measurement method for this study that is based on the use of round trip time measurements from a single source to two destinations. Our proposed method will use round trip time (RTT) measurements of back-to-back packets sent to different receivers. The key advantage of this approach is that it enables tomographic delay measurement to be conducted widely in the Internet. The challenges include the potential for noisy measurements due to unpredictable reverse path effects, and that instead of localizing all edges of the tree, only the aggregate link delay up to the branching node in the tree can resolved. We

call this approach *network radar* since it is analogous to the idea of standard radar, which sends signals into a medium, collects the "echoes" and compares the signal-to-echo ratio to estimate the distance to the objects. This represents a fundamentally different approach to network tomography. A simple schematic of Network Radar in the simplest case (two endhosts) is shown in Figure 4 below.

Most network tomography tools and studies focus on the single sender and multiple receivers scenario. The topology in this case is tree-structured, with the sender at the root and receivers at the leaves. The spread of the measurable network is mainly determined by the number of accessible measuring points (endhosts running special purpose software). Despite the potential in localizing the performance characteristics [Tsa03,Pad03,Chny02,Ada00,Har00,Cac99], the scalability in the inference techniques is largely restricted by the infrastructure. Our proposed network radar technique eliminates the need for special cooperation from endhosts, thereby enabling network tomography on a large scale.

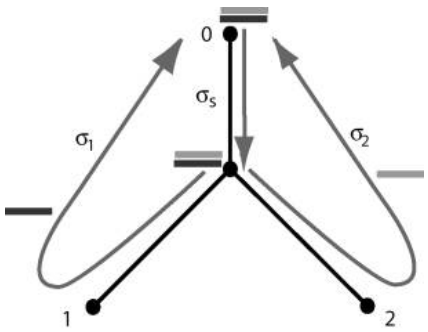

Figure 4: Network radar: A probing host sends back-to-back probes to two endhosts. Using the echoes, the probing host can estimate congestion on the shared link above the branching point to 1 and 2.

Our research strives to answer two major questions: (i) Given limited measurement points, can we come up with an alternative technique with minimum cooperation and still be able to cover the Internet at large? (ii) If we are given more measurement points, what do we gain from these extra measurements, by how much and where they should be placed?

One major challenge we will address is how edge based measurements can be used to study the performance of ultra high-speed networks. As the speed of the networks increases, the ability to measure within the network is reduced. It is important to understand how the edge-based measurements are impacted by the unique aspects of a ultra high speed network, such as the varying packet size, imhomogeneity of the network, and the speed-of-light propagation speed constraint.

Our proposed method will use RTT measurements of two closely time-spaced packets send to different receivers. This method has an advantage that it can be conducted widely method is similar to "ping" in which round trip delay is measured at the sender. However, "ping" depends on Internet Control Message Protocol (ICMP), which can be inaccurate due to the response packet generation time and the potential of ICMP packet filtering for security reasons. Our proposed approach will utilize the TCP's three-way handshake mechanism in the design of more effective probes. The feasibility of the approach poses a challenge. We have to address the potential for noisy measurements due to the unpredictable reverse path effects and the extreme diversity of bandwidth conditions in which the tool will operate.

2.3 Large-scale network monitoring and tomography

A key unsolved problem in network monitoring and tomography is *scaling*, i.e., how to probe and characterize the paths between all grid nodes in a large-scale grid. This problem is directly related to the stability of computing the inverse of the routing matrix that relates the probe measurements to the network path parameters of interest. The stability of the routing matrix, and hence the scalability of network monitoring and tomography, can be gauged by the condition number of the matrix. The condition number is the ratio of the largest to the smallest singular values. The larger the condition number is, the more unstable is the inversion process becomes, and thus the more sensitive network monitoring and tomography is to noisy measurements and other sources of error. A clear understanding of the growth of the condition number as a function of the number of endhosts will allow us to predict how many probes are required to achieve a certain level of accuracy, or conversely how many paths can be reasonably identified using a fixed number of probes. Moreover, the condition number will provide insight into the optimal deployment of additional measurement and probing sites within the network of interest. While of the analysis of condition numbers is well studied in large scale control problems, its use is novel in Internet sensitivity analysis. The diverse nature of envisioned ultra high-speed networks makes sensitivity and scalability studies all the more crucial.

2.4 Alpha-beta connection-level traffic model

As high-speed networks become increasingly complicated and hetrogeneous, network inference and control algorithms are placing escalating demands on traffic models. There is a pressing need for new traffic models and analysis techniques for dealing with the type of bursty traffic that can saturate a link or network, e.g., the traffic seen at the interface between circuit-switched core and packet-switched edges. Our approach leverages advanced techniques from signal processing, statistics, probability theory, and mathematics and our successes in signal and image modeling [CNB98,CB98,B99,CB99,CB99a, RCRB99,N99,TN99,Rib04a]; it is based on *multiscale analysis* to match both the short-range and long-range dependencies of ultra-high-speed traffic.

The sheer number of connections potentially active on a high-speed network forbids monitoring and controlling each of them individually. Fortunately, however, a careful study of many traffic traces acquired in different high-speed networking situations reveals that traffic bursts typically arise from just a few high-volume connections that dominate all others. This observation prompts an *alpha/beta model* for traffic, comprising a dominant, high-volume *alpha* component with relatively few connections but virtually all of the bursts, plus a low-volume aggregate *beta* residual with most connections [Sar01,Sar02]. Interestingly, the beta component is typically close to Gaussian and carries any long-range-dependent correlations seen in the aggregate traffic. The alpha/beta model enables a massive simplification in traffic modelling, since the bursty alpha connections can be analyzed, modelled, and even controlled individually, while the beta component can be analyzed, modelled, and controlled in aggregate. The result is a truly scalable connection-based traffic model.

We will develop an alpha/beta model suitable for use in network provisioning and buffer specification on the key links at the interface between the packet-switched network edge and circuit-switched UltraScience Net. As shown in our previous work [Sar01,Sar02], the heterogeneous RTTs faced by the competing connections arriving at the UltraScience Net will cause the traffic to behave as an alpha/beta process. However, the lack of cross-traffic in the circuit-switched network changes the rules of the alpha/beta game, resulting in several modeling and ultimately design and control challenges. For instance, the high bandwidth of the circuit-switched network and its limited cross-traffic will likely result in an increase in the relative number and strength of alpha connections and many more traffic bursts, which will significantly change the queuing behavior of the buffers [Sar04]. Moreover, a decrease in the relative strength of beta connections will weaken the long-range-dependence of the overall traffic. Finally, the

effect on the alpha and beta components of abrupt changes in available bandwidth as new lambdas are added or removed is completely unknown and must be understood.

Thrust 3: Scalable Cybersecurity

Because grid applications and high-speed testbeds such as UltraScience Net are not isolated from the global Internet, malicious behavior is inevitable. Two types of threats present a severe impediment for achieving robust and high performance networking: Denial of Service (DoS) attacks launched by users who wish to "bring down" communication and data services; and so-called "rational" attacks in which users want to increase their share of bandwidth as high as possible, without concern for fairness or network stability. Our methodology is to view DoS resilience as a key dimension of performance that must be fully incorporated into the protocol design process.

3.1 DoS-resilient transport protocols

While TCP's congestion control algorithm is highly robust to diverse network conditions, its implicit assumption of end-system cooperation results in a well-known vulnerability to attack by *high-rate* non-responsive flows, which is a common property of all TCP stacks. However, the problem of TCP fragility radically magnifies in high-speed environments. In essence, given the fact that a single TCP flow can utilize the bandwidth of the order of tens of Gb/s, the reasonable question becomes how to accurately distinguish such a flow from a malicious high-bandwidth flow.

Moreover, we have recently discovered a class of *low-rate* DoS attacks which, unlike high-rate attacks, are difficult for routers and counter-DoS mechanisms to detect [Kuz03c]. We have shown using Internet experiments that maliciously chosen low-rate DoS traffic patterns that exploit TCP's retransmission time-out mechanism can throttle TCP flows to a small fraction of their ideal rate while eluding detection. We will use a combination of analytical modeling, simulations, and Internet experiments, to design, implement, and evaluate a suite of high-performance and DoS-resilient transport protocols. By viewing performance and DoS-resilience as two tightly coupled aspects of protocol design, we will show that protocols can indeed simultaneously achieve both of these properties. We will explore three mechanisms. First, as such attacks exploit protocol homogeneity, we will study the ability of a class of randomized time-out mechanisms to thwart such low-rate DoS attacks without sacrificing performance due to spurious retransmissions. Second, we will revisit the foundations of TCP's slow-time-scale congestion control mechanisms with a perspective that high loss may be caused by DoS attack as well as high congestion. Third, we will study low-rate active probing to determine available bandwidth even in the presence of attackers. In this way, endpoints will be able to distinguish between congestion vs. DoS scenarios and react accordingly. We will particularly focus our experimental efforts to ultra-high-speed networks, where we expect the performance-security tradeoffs to be pronounced the most.

3.2 Cheat-resilient transport protocols

TCP variants that are widely deployed today are sender-centric protocols in which the sender performs important functions such as congestion control and reliability, whereas the receiver has minimum functionality via transmission of acknowledgements to the sender. Yet, it is becoming evident that increasing the functionality of *receivers* can significantly improve TCP performance [NETBLT87, TFRC00,Mehra03,WTCP99,Spring00,TCP-Real02]. Indeed, a key breakthrough in this design philosophy is represented by fully receiver-centric protocols in which *all* control functions are delegated to receivers [WEBTP00,RCP03]. The benefits that are being established for this design include improved

TCP throughput and an array of performance enhancements: (i) improved loss recovery; (ii) more robust congestion control; (iii) improved bandwidth aggregation; and (iv) improved web response times.

However, both sender- and receiver-centric protocols implicitly rely on the assumption that both endpoints cooperate in determining the proper rate at which to send data, an assumption that is increasingly invalid today. With sender-centric TCP-like congestion control, the sending endpoint may misbehave by disobeying the appropriate congestion control algorithms and send data more quickly. Fortunately, the lack of a strong incentive for selfish Internet users to do so (uploading vs. downloading) appears to be the main guard against such misbehavior. On the other hand, receiver-centric congestion control presents a perfect match for a misbehaving user: the receiving endpoint performs *all* congestion control functions, and has both the incentive (faster web browsing and file downloads) and the opportunity (open source operating systems) to exploit protocol vulnerabilities.

We will develop endpoint mechanisms to detect and thwart DoS attackers and cheaters. Our key objective is to ensure that enhanced TCP variants are resilient to endpoint misbehavior such that they can be safely deployed, while simultaneously ensuring that their full performance advantages can be realized. Our methodology is to first anticipate and analyze a set of possible receiver misbehaviors ranging from classical denial-of-service attacks (e.g., receiver request flooding) to more moderate and consequently harder-to-detect misbehavior. We will then develop a set of sender-side mechanisms designed to detect and thwart receiver misbehavior via dynamic sender-side estimation of the throughput that the sender estimates that the flow should obtain. Our approach is not to attempt to strictly enforce TCP friendliness: while this would catch all cheaters, it would also preclude deployment of advanced TCP stacks that can obtain higher throughput than standard TCP. Instead, our approach will be to let moderate cheating and "petty theft" of bandwidth slide, and instead catch flows engaged in more severe misbehaviors such as Denial of Service or aggressive bandwidth stealing that if not prevented would lead to poor performance for behaving flows or even congestion collapse.

3.3 Distributed security and monitoring of worms

Early detection of the spread of Internet worms is an important problem if we hope to develop a system for counteracting or stopping the spread of worms [Weaver03]. This problem is further amplified in *ultra* high-speed networks since the rate of infection is potentially amplified with a faster medium for propagation. Formally, an Internet worm refers to malicious, self-propagating code. To clarify the difference between a worm and a virus, worms exploit holes in an operating system to gain control of a host machine whereas viruses rely on the action of a human (e.g., opening an email attachment) to execute malicious code. Consequently, worms can spread extremely rapidly, making the prevention thereof an extremely challenging task. Once a new host has been compromised by a worm, the malicious code propagates by scanning new IP addresses at random. The classical model used to describe the spread of an epidemic through a finite population is

$$\frac{dI_t}{dt} = \beta I_t (N - I_t),$$

where N is the number of susceptible hosts, I_t is the number of hosts infected at time t, and β is the rate of infection. The solution to this differential equation (depicted in Figure 5) shows us that there is slowly increasing *initial* phase, when the initial number of infected hosts is small, followed by a *fast-spread* phase where the number of compromised hosts grows extremely rapidly, after which nearly all susceptible hosts are compromised in the final *slow-finish* stage. Thus, it would be most desirable to detect the presence of a worm during the initial phase.

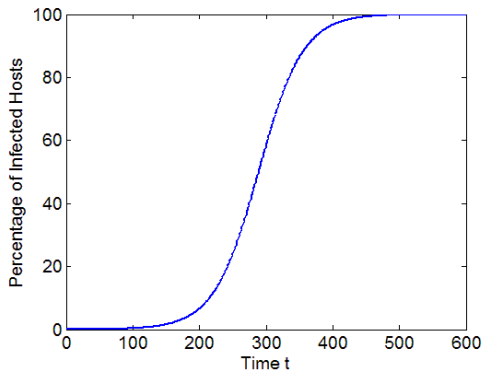

Figure 5: Modeling the spread of Internet worms: initial phase, fast-spread, slow-finish.

One approach to detecting the presence of a worm amounts to estimating β from the data and tracking this parameter over time. A major research challenge in detecting the presence of a worm is to differentiate between worm traffic and other anomalous traffic (e.g., hacker port scans) that may appear as noise. Zou et al. propose an early warning system for Internet worms in [Zou03], where measurement stations are deployed across the Internet. Each measurement station then transmits all of its data to a central location for processing. Their setup appears promising as a mechanism for detecting worms; however it lacks robustness in the sense that if the processing center becomes compromised (e.g., network connectivity is essentially lost due to a DDOS attack), then the entire system fails.

We will develop a completely *decentralized* and *scalable* worm detection system using an incremental update algorithm to estimate and track the worm propagation rate. Similar to [Zou03], our approach uses monitoring stations distributed across the Internet. However, rather than transmitting data to a central location, the monitoring stations circulate small messages amongst themselves. Messages contain a sufficient statistic for the set of parameters being estimated. Upon receiving a message, the monitoring station makes a small adjustment based only on its local data, and then passes the updated statistic to the other monitoring stations. This approach has the benefit of being robust to attacks, since all monitoring stations have an estimate of the state of the system. Techniques from the theory of incremental optimization and stochastic approximation will allow us to analyze the convergence rates of such algorithms and thereby assess the extent to which such a system will be capable of detecting and countering the spread of Internet worms. In transitioning from a centralized processing approach to one which is completely decentralized, challenges arise since multiple monitoring stations may potentially observe scans made by the same infected host. Thus, care must be taken to avoid making biased estimates or over-estimating the amount of anomalous traffic.

Thrust 4: Testing and Deployment

Design and deployment of new Internet protocols, systems and tools requires thorough evaluation and validation under a broad range of operational conditions. Unfortunately, the two most commonly used environments for protocol/tool/system evaluation, simulation and in situ testing, have well known drawbacks which limit their overall effectiveness. Simulations suffer from their inability to capture the robustness and dynamics of live environments, while lack of repeatability, control and access for measurements hinder in situ testing. We will develop two unique testbeds for evaluating and validating our ultra high-speed networking tools. Moreover, we will deploy our tools on the DOE UltraScience Net and facilitate their integration with grid applications.

4.1 LANL 10GigE testbed

LANL has had a *10GigE testbed* operational for the past 18 months and has conducted numerous experiments on high-speed data transfer, culminating in the Sustained Bandwidth Award ("Moore's Law Move Over!") at the SC2003 Bandwidth Challenge [Sus03].

4.2 Wisconsin Advanced Internet Lab (WAIL)

The Wisconsin Advanced Internet Lab (WAIL) [Barford03] is a one-of-its-kind facility for building arbitrary network configurations using actual networking equipment (wail.cs.wisc.edu). Development of the lab has taken place in the UW CS department over the past 20 months and has been made possible primarily through large donations from industrial sponsors. During this time, systems have been installed and configured, and a management environment for conducting experiments both internally and remotely has been developed. The lab currently has over 50 IP routers (ranging from low end access systems to backbone class 10 Gbps systems), 130 general purpose PC's and various other networking equipment (switches, caches, traffic shapers, measurement systems, etc.). Two important capabilities – propagation delay emulation and scalable, representative background traffic generation [Sommers04] – are also available. WAIL is in the process of adopting NetBed [White02] capabilities (through a collaborative grant between Utah and UW from NSF) to enable its PC systems to be used as flexible emulation nodes. Furthermore, over the next year, plans are coming together to develop an "optical core" evaluation capability within WAIL which will give it further relevance to UltraNet. This combination of systems and capabilities enables WAIL to provide a unique environment for investigating a broad range of network and distributed systems research topics.

One of the principle focuses of WAIL is to provide testbed capabilities for INCITE *Ultra* protocol/tool/system development. Part of the vision of this project is to construct a set of canonical network topologies in the lab that match existing environments to as great an extent as possible. An example is the recreation of the entire Internet2/Abilene infrastructure currently available in WAIL. Using this (or other topologies such as the TeraGrid and the UltraNet), we plan to develop a comprehensive test suite (including automated regression tests, evaluation tools and data repositories) for INCITE *Ultra*. The test suite will subject the protocols/tools/systems developed in this project to a broad range of network and use conditions. This unique testing capability should enable tools and protocols to be developed and deployed more rapidly, with higher quality and with more predictable performance than current test environments. Finally, the PI's (Barford - NLR board member) affiliation with the National LambdaRail project [NLR04] should facilitate this project's access to that unique infrastructure for testing and evaluation.

4.3 Deployment into the DOE UltraScience Net

There are few, if any teams in the world that have made more extensive network end-to-end measurements in the last 10 years than the SLAC IEPM-Pinger/BW team. During the 1990s the PingER project [PingER] was created to provide regular, low network, impact end-to-end measurements that now covers over 100 countries that between them contain more than 90% of the world's Internet connected population. More recently, the IEPM team developed a more intensive system [IEPM-BW] for throughput performance monitoring and analysis of high performance HENP, Grid and well provisioned network paths (i.e. mainly OC12 and Gb/s). Currently, the IEPM-BW project as installed at SLAC is monitoring more than 40 paths to their most important scientific partners in Europe, Canada and Japan. There are also nine other IEPM-BW monitoring sites making measurements on paths of particular relevance to them. IEPM-BW, with its focus on high performance (OC12 and Gb/s) paths, is relatively unique among today's measurement infrastructures in being able to test and provide information on the performance of higher-speed paths and of measurement tools in that environment. As part of the IEPM-BW project the team tested and compared many publicly available end-to-end network measurement tools. These included tools developed in the INCITE project, SciDAC and elsewhere. Feedback from the tests was made to the developers, especially to help them tune the tools for high-performance networks. As part of the INCITE project we have developed and deployed a tool *abing* that can be effectively used for monitoring in a continuous mode (each measurement takes less than a second) on many tens of paths. Such a low network impact tool (only 40 1450 Byte packets for a bidirectional available-bandwidth measurement) brings another level of information compared to more network intensive measurements such as iperf [iperf] or a file transfer application such as GridFTP [GridFTP] that can only be run infrequently (e.g. at 90 minute intervals). Abing has now been integrated in the IEPM-BW monitoring infrastructure together with the more successful of the other evaluated measurement tools.

The IEPM team has also deployed abing in the PLANET-LAB network [Planet] at 30 sites, in the US, Canada, Europe, Japan, China, and Brazil. Using this deployment they are studying the behavior of abing tools with a wide range of RTTs and losses. The IEPM team also has set-up and had access to short-term high-speed (2.5Gb/s and 10Gb/s) testbeds between Sunnyvale and StarLight [StarLight] in Chicago and SC03/Phoenix and PAIX in Palo Alto, as well as access to more permanent testbeds such as the DataTAG [DataTAG] and the NetherLight [NetherLight] testbeds and used these to test the measurement tools and techniques in ultra-high speed environments.

SLAC will be an early UltraScience Net site, enabling the testing of new tools on this new backbone network and at the same time providing monitoring results for the network. The emerging optical network technologies proposed for the DoE UltraScienceNet enable new (e.g. circuit oriented) techniques and performances that pose challenges for many of today's measurement tools. We therefore plan to evaluate how today's preferred end-to-end network measurement tools perform in the new ultra-high speed network environment, and to work with the developers to understand and solve problems. From this we will select a preferred set of successful tools and integrate them into the IEPM and other network measurement infrastructures.

Many of the existing and newly developed networking tools are designed to minimize their network intrusiveness and we will further explore this trend. This is critical if we are to make frequent (e.g. once a minute) measurements, with a wide full-mesh deployment, to provide a full picture of the network's status in each minute of its operation. With such a capability, changes in the network performance and structure, (e.g. path changes, congestions, etc.) can be detected and reported on. We believe that in the near future INCITE *Ultra* tools such as pathchirp (Task 2.1 above) [Rib03a,Rib04], abing, and network-radar (Task 2.2 above) can be valuable tools for network administrators. These active (i.e., they inject probes into the network) tools provide network status information complementary to that provided by passive mechanisms such as using the Simple Network Monitoring Protocol to request utilization and

other data from network devices such as routers and switches. SNMP, based tools, including industrial systems as CiscoWorks [CiscoWorks], HPOpenView [HPOpenView], etc. or the public domain MRTG [MRTG] tool are still the main sources of information and troubleshooting tools for most network centers. However, due to security and privacy concerns, access to such information is typically only available to a limited set of network administrators. Also these tools typically only provide information averaged over fairly long time intervals (typically a few minutes). Further though they provide information on the performance of individual components of the network and do not provide information on end-to-end performance. It is therefore critical to also provide complementary end-to-end bandwidth measurement tools that are available to network users at the end sites, can provide more frequent updates of the performance for selected end-to-end paths, and can therefore quickly detect significant changes in performance and network problems.

On the other hand, from experience we are aware that independent monitoring projects developed and managed from outside the network management domain itself, may have difficulty in convincing the network administrators of their relevance. To assist in bridging this gap, we will closely collaborate with teams who are currently responsible for the network management of ESnet and Internet2, and with other network authorities.

Most of the proposed tools and/or data from the measurement tools are designed to be used directly by the individual users and/or their applications. Our goal is to provide access to the tools for end-users to allow them to quickly discover significant changes in end-to-end network paths that usually extend over multiple Internet Service Providers (ISPs). Providing access to network measurements will allow bulk-data transfer applications to estimate and report the expected transfer time, and enable data placement for replication of data. In the ideal case, measurements made by various techniques (active vs. passive, network intensive vs. low-impact, end-to-end vs. backbone, on-demand vs. historical) should give similar or comparable results. This is needed to be able to make and compare instantaneous results with the publicly available historical data, to relate end-to-end performance to router utilization at congestion points, and to diagnose end-to-end problems.

The tools will also be used for the creation of "Virtual Organization" (VO) Monitoring Systems. This is increasingly required for such communities as High Energy and Nuclear Physics (HENP), Astrophysics or the Grid. Each of these communities has unique circles of resources (e.g., in HENP those used by a particular experiment such as BaBar or LHC/CMS) that should be monitored independently with respect to the users and their demands. There are few projects that are trying to cover this problem. One such possible project is the "MAGGIE" SciDAC proposal. We are in contact with MAGGIE developers (one of the PI's for MAGGIE is also a PI on the current proposal) and are prepared to integrate developed tools into MAGGIE as a first step of our future work. Similar practice will be used for the Grid community. Currently we are in close contact with Euro-Grid developers of monitoring tools, and we propose to integrate abing into their monitoring and presentation systems (Mapcenter [Mapcenter] and MonALISA [MonALISA]).

Most major academic backbone networks now support both the IPv4 protocol (as production) and IPv6 protocols (as experimental). IPv6 use is gradually growing and more and more universities and scientific labs are starting to install new hosts in a dual mode and to do the experiments with this type of networking. Currently, there are few monitoring tools available for IPv6. One of the first tools in this field was PingER. It was modified for IPv6 in 2000 [IPv6]. Today it is being used to monitor some tens of IPv6-based hosts worldwide. SLAC is part of the ESnet IPv6 testbead, and has machines connected to it running IPv6.

Data-intensive science and grids have a critical for need for high-performance bulk-data transfer. The current TCP-IP (Reno-based) protocols perform poorly on long-distance high-speed networks. To address

multiple new advanced transport protocols are being developed such as the TCP based HS-TCP and FAST and the UDP rate-based UDT [UDT] and RBUDP [RBUDP] protocols. To meet the needs of circuit switched networks, such as will be experimented with in the UltraScienceNet, new protocols such as IBP [IBP] and remote disk access techniques will be developed and will need evaluating in terms of ease-of-use, scalability, performance, integration with applications etc. The IEPM team is well positioned to take a leadership role in this due to being colocated at the HENP BaBar experiment's tier 0 (accelerator) site which has immediate massive data transfer needs – BaBar has already collected about a petaByte of data, and needs to share over a terabyte/day of data with tier 1 sites in Europe; having close contacts with the BaBar users and developers of BaBar applications such as bbcp [bbcp] and bbftp [bbftp].

We have close connections with the HENP community, including CERN, SLAC, the EU DataTag [DataTag], and the EU DataGrid [EDG] (now called EGEE [EGEE]). We also have strong ties to other high-energy physics projects such as the Particle Physics Data Grid (PPDG), Grid2003 [Grid2003] and the SLAC-led BaBar [BaBar] project. We will work with these groups to determine the requirements needed by the applications community. We have contacted other experimental networking facilities such as Planet-lab and developers of new presentation tools (Caltech and CERN). Finally, we will work closely with the CAIDA networking developers (project titled "Pythia: Automatic Performance Monitoring and Problem Diagnosis in Ultra High-Speed Networks").

Impact

The INCITE *Ultra* Project will provide the key missing link between the capacity offered by ultra high-speed transport technology and the demands of grid computing. Our research will provide fundamental technical contributions in high-speed network protocol design, inference, modeling, DoS resilience, and experimentation. Our research in high-performance TCP (Thrust 1) will ensure that not only can a single TCP flow utilize an ultra-high speed links, but also multiple flows can share bandwidth according to diverse fairness and performance objectives. Our research in path and link inference (Thrust 2) will enable accurate and robust assessment of the network's internal performance via highly scalable techniques. Our research in cybersecurity (Thrust 3) will target robust protocol design and DoS/worm detection to ensure that malicious users cannot attack network protocols and end hosts. Finally, our testbed experiments on UltraScience Net (Thrust 4) will provide a proof-of-concept of our designs and will yield dramatic performance enhancements to grid applications.

Our reference implementations will provide first-of-their-kind platforms for obtaining a deep understanding of ultra high-speed networks and enable principled designs of future network architectures, algorithms, protocols, and models. To ensure that this project impacts DOE, industry, international standards bodies, and advanced development laboratories, Rice University, University of Wisconsin LANL, and SLAC will work closely in a single cooperative effort.

C. Project Management

The research will be conducted in close collaboration at four sites: Rice University, University of Wisconsin-Madison, Los Alamos National Laboratory (LANL), and the Stanford Linear Accelerator Center (SLAC). While we already have a long history of collaboration (most recently through our current DOE INCITE project), we will enhance our communication and collaboration by a number of means:

- Research interactions will be greatly facilitated by LANL's and Rice's close link via the *Los Alamos Computer Science Institute* (LACSI).

- Students and postdocs from Rice will spend summer internships at LANL and SLAC. In addition to visits by the PIs, Rice students Yolanda Tsang, Ryan King, Vinay Ribiero, Shri Sarvotham, and Aleksandar Kuzmanovic have already conducted extended visits/internships at SLAC and LANL in the period 2001-2004.

- The PIs and co-PIs will make frequent visits between the three sites. Progress mini-workshops will be held twice per year in February and July, rotating between the three sites. Finally, we plan a larger project workshop will be held at the completion of the project; it will be open to the networking community at large.

- We will maintain a project web page at incite.rice.edu. This will retain our visibility in the broader high-speed networking community. We maintain a web presence for each of our projects (see the projects under incite.rice.edu, dsp.rice.edu, www.ece.rice.edu/networks, and www.cmc.rice.edu). Through the project web site, outsiders will obtain publications, released software, and general information about the project.

- LANL and SLAC will participate in the Rice University and University of Wisconsin Departments of Electrical and Computer Engineering Affiliates meetings that bring numerous companies and research laboratories to the respective campus to discuss the relationships between our research and their problems. Besides students in academia, Rice students are working at companies such as Texas Instruments, Nokia, Nortel, Lucent, Motorola, Intel, and Cisco. This web of professional and personal contacts opens the door through which we can encourage the adoption of the best of our work.

- PI Baraniuk from Rice University will facilitate integration and dissemination of the results among the co-PIs and project participants, as well as DOE management.

In each of the sections that follow, we delineate the roles of each PI institution.

D. Rice University Details

Deliverables, milestones, and timeline

The primary deliverables will include a suite of scalable and secure transmission protocols (HSTCP-LP, HSTCP-RF protocol, dynamic right-sizing TCP protocol, TCP-Paris), an alpha/beta traffic modeling toolbox, and a tested software implementation of the pathChirp end-to-end path monitoring tool.

- ***Ultra high-speed protocols***
  - Year 1
    - Design enhanced TCP protocol suite. Implement dynamic right-sizing and passive chirp probing in TCP (with LANL). Test on LANL and WAIL. Lay the theoretical foundation for HSTCP-RF performance in terms of end-to-end path characteristics (statistics of delays and losses); study stability and fairness with respect to TCP-Reno.
  - Year 2
    - Begin testing dynamic right-sizing and HSTCP-RF on both high speed packet networks and interfacing with UltraScience Net (with SLAC). Begin coding robust implementations of TCP suite suitable for wide-scale release.
    - Compete for world land speed record.
  - Year 3
    - Release software.
    - Testing and validation.
    - Compete for world land speed record.

- ***Alpha-beta traffic model***

  - Year 1
    - Study theoretical characteristics of alpha/beta model; analyze trace data from high-speed and ultra-high-speed testbeds (LANL and WAIL) and real networks (with SLAC and LANL). Verify that alpha/beta is a good match to the data via real-world situations, in particular at routers on the interface between the packet-switched edge and circuit-switched UltraScience Net. Make improvements to the model as inconsistencies arise.
  - Year 2
    - Code efficient research implementation of alpha/beta analysis and model that could run on edge router CPU. Develop alpha/beta analysis toolbox (research code).
  - Year 3
    - Develop and disseminate release version of alpha/beta toolbox.

- ***PathChirp for ultra high-speed networks***

  - Year 1
    - Develop prototype ultra high-speed pathChirp implementation and test on LANL, WAIL and INDRA testbeds.
  - Year 2
    - Tune implementation and begin testing on UltraScience Net (with SLAC). Begin integrating into several representative grid applications (with SLAC). Investigate

opportunities for fusing tomographic inferences with path inferences. Develop and begin test of spatio-temporal pathChirp to localize tight links.

- Year 3
  - Complete testing and validation on UltraScience Net and test and validate operation with grid applications (with SLAC).

Scalable cybersecurity

- Year 1
  - Anticipate and model protocol attacks and worms.
- Year 2
  - Design DoS resilient protocol enhancements and counter-attack strategies.
- Year 3
  - Integrate cybersecurity protocol enhancements into the project's protocols and tools and perform testing for performance and robustness to attack.

Budget justification - Rice University

Year 1:	$582,016	(1/3 each for Thrusts 1, 2, 3)
Year 2:	$589,541	(1/3 each for Thrusts 1, 2, 3)
Year 3:	$619,442	(1/3 each for Thrusts 1, 2, 3)

Amount Requested:	$**1,799,999**
Project Total:	**$1,799,999**

See detailed budget sheets from Rice University.

The proposed budget provides partial support for the three principal investigators, partial support for three postdoctoral associates (i.e., 2.5 FTE), and full support for a programmer and three graduate students. Baraniuk, Knightly, and Riedi are full-time tenure track faculty at Rice. One month of summer support for each is requested.

Thrusts 1, 2, and 3 will each be staffed by one DOE-supported graduate student. The postdoctoral associates will promote cross-thrust synergy by researching both directly on and "between" the thrusts to ensure smooth and continuous information flow from our modeling and inference research into high-speed protocols and security algorithms (Thrust 1 to Thrust 3). Rice has an excellent track record of attracting top-flight postdocs who go on to academia and leadership positions in industry. One example is Mark Coates (PhD, Univ. Cambridge), a postdoctoral fellow supported by Texas Instruments, who worked on our first DOE INCITE project and is now an assistant professor at McGill University.

The *programmer* plays a key role in our program, with the responsibility to transform research quality code into well-documented ns and IP compatible software that our LANL/SLAC partners and we can use to test and validate our results on real networks. The programmer will also interface with our industrial sponsors to ensure a maximally efficient technology transfer of our algorithms.

The budget includes funds to support the *Fringe Benefit* costs related to the salaries and wages for the Co-PI and the postdoctoral associate. These amounts are in accordance with the rates approved by Rice University's cognizant DHHS agency.

Travel funds are requested to attend regular DOE PI meetings and to collaborate with our LANL/SLAC partners and other investigators in the DOE high-speed networking program.

Funding is requested for *Materials and Supplies* to support the needs of the postdoctoral associates, programmer, and graduate students working on the proposed research program. These funds will be used to purchase workstations (first year) and to provide the necessary operating materials and services (e.g., computer supplies, publication costs, long-distance telephone charges, etc.). Fringe benefits are calculated according to regulations at Rice University.

The budget includes funds for Rice University's Facilities and Administrative expense (i.e., *Indirect Costs*), based on 51% of modified total direct costs. The base excludes tuition remission costs for graduate students and all subaward costs greater than $25,000.

The goals of this project align naturally with a number of other ongoing projects at Rice, which are supported by:

- Six NSF grants related to networking, including two Career awards, two from the Special Projects in Networking program ($1.2M and $1.4M), three from the Information Technology Research (ITR) program ($5M, $7.5M, and $2.4M), and one from the Wireless Information Technology and Networks Initiative.
- A $1m grant from DARPA's Network Modeling and Simulation (NMS) program.
- Grants for signal and image modeling and tomography from DARPA, ONR, ARO, and AFOSR.
- $700k from Nokia and the State of Texas Advanced Technology Program for QoS networking and new high-speed wireless LAN technology.
- $1M from Texas Instruments from its Leadership University Program, which supports research in signal modeling, traffic modeling, tomography, and QoS networking.

Current and Pending Support – Richard G. Baraniuk

Current

DOE "INCITE: Edge-based Traffic Processing and Service Inference for High Performance Networks"; 8/15/01 - 8/14/04; $1,275,000; PI with multiple Co-Investigators

DARPA "Multiscale Traffic Processing Techniques for Network Inference and Control"; 4/28/00 - 4/28/04; $946,000; PI with multiple Co-Investigators

ONR "Coherent Multiscale Statistical Modeling using Complex Wavelets"; 3/01/02 - 11/30/04; $290,000; PI

NSF "A Framework and Methodology of Edge-Based Traffic Processing and Service Inference"; 8/15/01 - 7/31/04; $978,204; Co-Investigator with multiple Investigators

NSF "SAFARI: A Scalable Architecture for Ad Hoc Networking and Services"; 1/1/04 - 12/31/07; $1,450,781; Co-Investigator with multiple Investigators

Pending

DOE "Monitoring, Verification and Risk Assessment for Carbon Sequestration Options"; 11/1/04 - 10/31/07; $200,000; Co-Investigator w/industrial partner (Winrock International) as lead

NSF "S&T Prognosis Center of Excellence for Systems of Systems (ProCESS)"; 6/15/05 - 6/14/10; $850,000; Co-Investigator w/Purdue University and multiple Investigators

NSF "Collaborative Research: SST: Novel Optical Chemical Agent Sensors Based on Quartz-Resonance-Enhanced Mid-Infrared Laser Photoacoustic Spectroscopy" 10/1/04 - 9/30/07; $375,000; Co-Investigator with multiple Investigators

Current and Pending Support – Edward W. Knightly

Current

NSF "ITR: Large: 100 Mb/sec for 100 Million Households" 10/01/03 - 09/30/08; $7,500,000; PI with multiple Investigators

NSF "ITR: Medium: Wireless Transit Access Points – New Foundations for a Scalable, Deployable, high Performance Wireless Internet"; 09/15/03 – 08/31/04; $2,400,000; PI with multiple Co-Investigators

NSF "ITR: Collaborative Research: Scalable Services for the Global Network"; 09/01/00 – 08/31/05; $1,225,074; PI with multiple Investigators

NSF "A Framework and Methodology of Edge-Based Traffic Processing and Service Inference"; 8/15/01 - 7/31/04; $978,204; Co-Investigator with multiple Investigators

DOE "INCITE: Edge-based Traffic Processing and Service Inference for High Performance Networks"; 8/15/01 - 8/14/04; $1,275,000; Co-Investigator with multiple Investigators

TEXAS"TDT: Enabling Technologies for Developing Wireless Transit Access Points"; 01/01/04 – 12/31/05; $200,000; PI with Co-Investigator

TEXAS"ATP: Accelerating the WWW via Integrated Internet and IDC Performance Optimization"; 01/01/04 – 12/31/05; $200,000; PI with Co-Investigator

Current and Pending Support – Rudolf H. Riedi

Current

NSF "SAFARI: A Scalable Architecture for Ad Hoc Networking and Services"; 1/1/04 - 12/31/07; $1,450,781; PI with multiple Co-Investigators

Texas ATP "A Scalable Architecture for Ad Hoc Networking and Services"; 1/1/04 - 12/31/05; $210,000; co-PI with multiple Co-Investigators

DOE "INCITE: Edge-based Traffic Processing and Service Inference for High Performance Networks"; 8/15/01 - 8/14/04; $1,275,000; Co-Investigator with multiple Investigators

NSF "A Framework and Methodology of Edge-Based Traffic Processing and Service Inference"; 8/15/01 - 7/31/04; $978,204; Co-Investigator with multiple Investigators

DARPA "Multiscale Traffic Processing Techniques for Network Inference and Control"; 4/28/00 - 4/28/04; $946,000; Co-Investigator with multiple Investigators

Pending

NSF "EnterPRiSE – Evolving Non-Semi-Parametric Risk Space Exploration"; "; 7/1/04 - 6/30/07; $1,021,498; co-PI with multiple investigators

E. University of Wisconsin Details

Deliverables, milestones, and timeline

The primary deliverables will include the development of the Network Radar software tool based on round-trip-time measurements for monitoring internal network performance characteristics, analysis of the scalability of network tomography and radar algorithms in packet-switched and circuit-switched DOE networks, and distributed monitoring tools for cybersecurity. We will also develop and prototype tools, protocols, and algorithms within the Wisconsin Advanced Internet Laboratory (WAIL).

- ***Network RADAR: Network tomography using end-to-end measurements***
    - Year 1
        - Research, development, and limited deployment of a Network RADAR tool.
    - Year 2
        - Deployment and testing of Network RADAR in WAIL testbed at UW-Madison.
        - Experimentation with alternated back-to-back probes based on different transport protocols, including ICMP and TCP-SYN-ACK.
    - Year 3
        - Testing and further enhancement of RADAR tool at collaborating sites.
        - Assessment of performance and feasibility of RADAR in DOE network infrastructure.
        - Final software-distribution release of RADAR.

- ***Large-scale network tomography***
    - Year 1
        - Analysis of condition numbers (i.e., numerical stability) of network inference tools including network tomography, RADAR, and pathChirp.
    - Year 2
        - Bounds on probing requirements for guaranteed accuracy in inference tools.
        - Development of algorithms for optimizing probing locations and probing sequences.
    - Year 3
        - Testing and validation of proposed algorithms in WAIL testbed.
        - Evaluating potential of inference tools in hybrid packet-switched/circuit-switched networks.

- ***Distributed security and monitoring***
    - Year 1
        - Investigate feasibility of distributed algorithms for worm detection and monitoring.
    - Year 2
        - Development of light-weight distributed tool for network worm monitoring.
        - Initial testing and experimentation in WAIL testbed.
    - Year 3
        - Testing and validation at partner sites, and validation on existing datasets.

- ***Wisconsin Advanced Internet Lab* (WAIL)**

  - Year 1
    - Construction of a set of canonical network topologies in WAIL that match expected DOE networking environments.
  - Year 2
    - Development of optical core evaluation capabilities within WAIL.
  - Year 3
    - Development of a comprehensive test suite (including automated regression tests, evaluation tools, and data repositories).

Budget justification – University of Wisconsin

Year 1:	$309,079	(1/3 each for Thrusts 2, 3, 4)
Year 2:	$320,356	(1/3 each for Thrusts 2, 3, 4)
Year 3:	$332,107	(1/3 each for Thrusts 2, 3, 4)

Amount Requested:	$961,541
Project Total:	**$961,541**

See detailed budget sheets from University of Wisconsin.

A. Senior Personnel: Robert Nowak requests 25% academic year salary to support his efforts on the project. Paul Barford requests 1 month summer salary.
B. Other Personnel: Two graduate students will assist Nowak in the research activities in this project. Funding is requested for two ½-time Research Assistants (20 hours per week for full duration of project at pay rate of approximately $24,000 per year). Barford requests one ½ -time Research Assistant (20 hours per week for full duration of project at pay rate of approximately $24,000 per year) and 1 programmer/technician (20 hours per week for full duration of project at pay rate of approximately $27,000 per year).
C. Fringe Benefits: The fringe benefit rate is 45.5%.
D. Equipment: None.
E. Travel: $18,000 is requested for travel to/from collaboration sites and technical meetings and conferences.
F. Trainee/Participant Costs: None.
G. Materials and Supplies: $30,000 is requested for materials and supplies. This will support the purchase of computing equipment and supplies necessary for completing the proposed research activities.
   Publication Costs: $12,000 is budgeted to cover the costs of publication.
   Computer Services: $18,000 is budgeted to cover the costs of operation and maintenance of computing resources.

Current and pending support – Robert Nowak

Current

R. Nowak, ``Complexity Regularized Signal Processing for Networking Applications," National Science Foundation , grant no. CCR-0310889, $200,000, 9/1/03-8/31/06; 1 month summer support

R. Nowak (co-PI with A. Hero, G. Michaleadis, S. LaFortune, D. Teneketsis, M. Crovella, P. Barford, E. Kolaczyk) , ``Modular Strategies for Internetwork Monitoring," National Science Foundation ITR, (total funding over $1,000,000) grant no. CCR-0325571, 9/1/03--8/31/08; 1 month summer support

R. Nowak (co-PI), Baraniuk (PI), “INCITE: Edge-Based Traffic Processing and Service Inference for High Performance Networks,” Department of Energy, $1,275,000, 8/15/01-8/14/04

R. Nowak, ``Statistical Inference Problems in Communication Networks," Office of Naval Research, $64,908, grant no. N00014-03-1-0966, 8/18/03-7/31/04.

R. Nowak (PI with R. Baraniuk, E. Knightly, and R. Riedi), ``A Framework and Methodology for Edge-Based Traffic Processing and Service Inference," National Science Foundation , grant no. ANI-0099148, $996,704, 8/15/01--7/31/04.

R. Nowak ``Network Topology Investigation," Applied Signal Technology, Inc., 11/10/03-9/30/04.

Pending
R. Nowak (with B. Van Veen) Robust Spatio-Temporal MEG/EEG Functional Imaging, NIH, $800,000.

Current and pending support – Paul Barford

Current
NSF "iSWORD: Instrumented Streaming Research and Testbed"
$1,072,808, 2001-2004 (co-PI)

NSF "Collaborative ITR/CSE: Modular Strategies for Internetwork Monitoring"
$524,270, 2003-2008; 3 months summer support

NSF "Collaborative Research: A Unified Experimental Environment for Diverse Networks"
$250,000, 2003-2006; 3 months summer support

ARO “Coordinated Anomaly Detection and Characterization in Wide Area Network Flows"
$320,000, 2002-2005; 2 months summer support

Cisco Systems
“Empirical Evaluation of Buffer Performance in High Performance Routers”
$92,428, 2003-2004

EMC Corporation
“An Integrated Program for Content Addressable Storage Research”
$92,200, 2003-2004

Intel Corporation
“Data Center Traffic Monitoring and Control”
$82,593, 2003-2004

Pending
NSF "CAREER: A Multiresolution Approach to Network Anomaly and Intrusion Detection”
$627,376, 2004-2008, 3 months summer support

F. LANL Details

Deliverables, milestones, and timeline

The primary deliverable from LANL will include adapting and incorporating *pathChirp* into dynamic right-sizing, a flow-control adaptation technique that is applicable to *any* TCP, for enhanced performance for large-scale bulk data transfer. The adaptation and incorporation will entail transforming path chirps from being an "active probing" mechanism into a "passive probing" mechanism that is embedded in the existing data stream.

To better capture (and control) the high-speed environment that dynamic right-sizing and path chirps will operate in, we will conduct the initial software development, testing, and evaluation in LANL's 10-Gigabit Ethernet testbed environment.

- ***Dynamic Right-Sizing with Path Chirps***
  - Year 1
    - Modify *pathChirp*, as necessary, for incorporation into a kernel implementation of dynamic right-sizing. (Specifically focus on transforming the path chirp from an "active probing" mechanism to a "passive probing" mechanism.)
  - Year 2
    - Deployment and testing of "dynamic right-sizing + pathChirp" in the existing 10-Gigabit Ethernet testbed at Los Alamos National Laboratory.
    - Experimentation with alternated back-to-back *passive* probes in "dynamic right-sizing + pathChirp."
  - Year 3
    - Testing, tuning, and further enhancement of "dynamic right-sizing + pathChirp" at collaborating sites.
    - Deployment and testing of "dynamic right-sizing + pathChirp" to DOE UltraScience Net and ESnet community.
    - Assessment of the performance and feasibility of "dynamic right-sizing + pathChirp" in DOE network infrastructure.
    - Final software-distribution release of RADAR.

Budget justification - LANL

Year 1:	$300,000	(1/3 each for Thrusts 1,4)
Year 2:	$300,000	(1/3 each for Thrusts 1,4)
Year 3:	$300,000	(1/3 each for Thrusts 1,4)

Amount Requested:	$900,000
Project Total:	**$900,000**

See budget detailed sheets from Los Alamos National Laboratory.

Current and pending support – *Wu Feng*

Current

Software-Based Power-Aware Computing. PI. 3 calendar months.
Award: $75,000/year, October 2003 – September 2004.
Agency: DOE Laboratory-Directed Research & Development.

Interface Design for High-Performance Networking. Co-PI. 0.01 calendar months.
Award: $100,000/year, October 2003 – September 2004.
Agency: University of California – Cooperative Agreement on Research and Education.

Collaborators: University of California at Riverside (L. Bhuyan)

Wide-Area Transport and Signaling Protocols for Genome to Life Applications. PI. 0.01 calendar months.
Award: $45,214/year, July 2003 – June 2004.
Agency: University of California – Cooperative Agreement on Research and Education.

Software Technology to Enable Reliable High-Performance Distributed Disk Arrays. Co-PI. 0.01 calendar months.
Award: $540,000/3 years, June 2002 – May 2004.
Agency: NASA Applied Information Systems Research Program.

Reliable Networking in System- and Wide-Area Networks. PI. 1.20 calendar months.
Award: $910,000/3 years, October 2001 – September 2004.
Agency: Los Alamos Computer Science Institute.

Improvements to TCP over the Wide-Area Network. PI. 0.60 calendar months.
Award: $500,000 / 3 years, October 2001 – September 2004.
Agency: DOE ASCI.
Program: Distributed & Distance Computing (DISCOM).

INCITE: Edge-Based Traffic Processing and Service Inference for High-Performance Networks. Co-PI. 2.40 calendar months.
Award: $2,281,000/3 years, July 2002 – July 2004.
Agency: DOE Office of Science.
Program: Scientific Discovery through Advanced Computing (SciDAC).

High-Performance Transport Protocols. PI. 2.40 calendar months.
Award: $750,000/3 years, October 2001 – September 2004.
Agency: DOE Office of Science.
Program: Base Program.

Pending

Improvements to mpiBLAST. PI.
Award: $10,000/year (pending). 0.01 calendar months.
Agency: Advanced Micro Devices, Inc. (AMD).

Current and pending support – *Mark Gardner*

INCITE: Edge-Based Traffic Processing and Service Inference for High-Performance Networks. Co-PI. 1.00 calendar month.
Award: $2,281,000/3 years, July 2002 – July 2004.
Agency: DOE Office of Science.
Program: Scientific Discovery through Advanced Computing (SciDAC).

G. SLAC Details

Deliverables, milestones, and timeline

SLAC will deliver developed tools and make testing measurements and experimental monitoring across multiple administrative domains that include Abilene, ESnet, and UltraScience Net, and others networks (as Canet, Geant, Nordunet, Apan, AARnet) where most of our scientific partners are located. We break the SLAC deliverables into the following areas:

- Integration of developed tools into the infrastructures of existing monitoring systems and coordinate development of new tools according to the new network environments, requests and needs of network administrators and users and the results of analysis characteristics of new tools as effectiveness, correctness and intrusiveness with reality or other tools.
- Make a permanent publication of monitoring data from 50-100 selected locations (in different types of networks all over the world) in several levels of hierarchy (24 hours/weeks/months) with respect to the tools capabilities and potential users (communities) requests.
- Make a case studies and data analysis from different situations. Show, how such situations were visible via different measurement tools, TCP implementations or topology results, classified them into typical categories according to the different environments, traffic circumstances, changes of network topology etc. and make them publicly available.
- We will continue in the development and testing bandwidth estimation tools based on current tools as *pathChirp* and "abing" in new conditions which brings new network infrastructures with the latest optical technology based on traditional SONET/SDH (POS) or new TDM, DWDM technology [Lambda], a high level of aggregation of different type of traffic and the new type of services as MPLS etc.
- As part of INCITE we will develop IPv6 version of "abing" . Following this we will deploy "abing" to a group of selected hosts and start experimental monitoring on this virtual network connected via Internet2, ESnet and potentially via the UltraScienceNet infrastructure. In the future we can use this experience for converting and testing other tools developed originally for Ipv4 in this new type of network.
- SLAC along with our collaborating partners, will select suitable testing infrastructures for tomography tests. We will deploy topographic tools into this infrastructure, start collecting data and run the analysis program in scheduled intervals. We believe the new proposed concept with the link delay strategy can be relatively easily integrated into existing monitoring systems and provide totally new information which would be an interesting subject for further studies.
- SLAC will evaluate and compare the developed transport tools such as HSTCP-LP, with other offerings such as FAST TCP, HS-TCP and derivatives, and rate-based non-TCP transport tools such as RBUDP, UDT etc. in both high-speed production. Networks and testbeds such as UltraScience Net. Realms of applicability (based on throughput, fairness, stability, resource utilization, ease of use/deployment) will be determined, documented, provided to the developers and others, and recommendations on utilization will be provided. The SLAC team will work with developers and users of production applications such as the BaBar data replication services to deploy the chosen transport protocols (e.g. at tier0 and tier1 HENP sites) so that the services and end users can take advantage of them. This in turn will provide further experience for the transport protocol developers, while also bringing improved performance to the applications and their users.

Year 1:

- Integration

  Make deployment of topographic tools into the infrastructure that is used for bandwidth monitoring (first 20 nodes) and start collecting topographical-data

  Make an analysis of characteristics of new communication protocols used on optical networks with the respect to influence on current measurement tools. Based on this study prepare a strategy for developing new generation of measurement tools

- Data publication

  Prepare data from INCITE tools developed in 2002-2004 project for public presentation via presentation tools as MonALISA and start collecting data

- Case studies

  Prepare review about used monitoring tools and compare the results in large scale of time for different paths, environment and protocols.

Year 2:

- Integration

  Continue in deployment of all tools into new infrastructure and extend number of paths in continues monitoring mode

- Data publication

  Continue collecting and publishing data from all monitoring sites in available presentation systems MonALISA, Mapcenter, etc. which will be available in this time or recommended by the community groups

- Case studies

  Report of all exceptional situations recorded during first year of test monitoring

  Make an analysis of correctness of existing tools on large scale of monitored paths.

Year 3:

- Integration

  Deployment of new generation tools into new infrastructure available for monitoring in ultra high speed networks

  Continue to evaluate new and tools. Identify improvements and new needs and communicate to developers

- Data publication

  Continue publish data from test monitoring

- Case studies

  Prepare final study in which we will summing up our results from test monitoring and experimental measurements on high and ultra high speed networks and present an recommendation how tools can be used in most effective mode.

Budget justification - SLAC

Year 1:	$199,880	(Thrust 4)
Year 2:	$205,877	(Thrust 4)
Year 3:	$213,001	(Thrust 4)

Amount Requested:	$618,758
Project Total:	**$618,758**

See budget detailed sheets from SLAC.

SLAC Personnel

R. Les Cottrell — (0.08 FTE) will supervise the work on this project, direct and participate in the research and data analysis, work with developers to evaluate other monitoring tools, interface with the ESnet and HENP communities to gather requirements and promote deployment.

Jiri Navratil - (0.83 FTE) will be responsible for research, detailed design and implementation of testing new measurement tools and TCP stack evaluations. Will work with measurement infrastructures to integrate and deploy applicable measurement tools in the infrastructures. Will write web pages, publish findings, and present results at various conferences and meetings.

SLAC Direct Costs

Cost estimates have been presented in this proposal to be comparable to other research institution's proposals. At the Stanford Linear Accelerator Center, actual costs will be collected and reported in accordance with the Department of Energy (DOE) guidelines. Total cost presented in this proposal and actual cost totals will be equivalent.

Senior Personnel – Item A.1-6
The salary figure listed for Senior Personnel is an estimate based on the current actual salary for an employee in her/his division plus 3% per year for inflation.

Fringe Benefits – Item C
Fringe Benefits for SLAC employees are estimated to be the following percent calculated on labor costs:

- Career Employees – 29%
- Students/Others – 3.5%

Travel – Items E.1 and E.2
The senior staff members plan to attend domestic and/or foreign technical conferences/workshops in the areas of research covered by this proposal. Total cost includes plane fare, housing, meals and other allowable costs under government per diem rules.

Other Direct Costs- Item G.6
The estimated cost of tuition for graduate students.

Indirect Costs – Item I

- Materials and supplies, clerical support, publication costs, computer (including workstations for people) and network support, phone, site support, heating, lighting etc. are examples of activities included under indirect costs. Indirect costs are 36% of the Salaries including the Fringe, 36% of Travel costs and 6.8% on Materials and Supplies.

Current and pending support – Les Cottrell

Current Support:

- Project: DOE/SciDAC Edge-based Traffic Processing and Service Inference for High-Performance Networks (INCITE)
- Percent Support: 5%; Duration: ends October 2004
- DOE HENP base funding: 95%

Other Pending Support:

- Project: DOE/SciDAC TeraPaths: A QoS enabled Collaborative Data Sharing Infrastructure for Peta-scale Computing Research
- Percent support: 8%
- Project: Measurement and Analysis for the Global Grid and Internet End-to-end performance (MAGGIE)
- Percent support: 17%

Current and pending support – Jiri Navratil

Current Support:

- Project: DOE/SciDAC Edge-based Traffic Processing and Service Inference for High-Performance Networks (INCITE)
- Percent Support: 100%; Duration: end October 2004

Other Pending Support:

- None

H. Facilities and Resources

Rice University

Research will be conducted in the new $32M *Computational Engineering Building*, which houses the departments of Electrical and Computer Engineering, Computer Science, Statistics, Applied Mathematics, the NSF/DOE/DOD Center for Research on High-Performance Software (HiPerSoft), and the Rice Computer and Information Technology Institute (CITI), of which all PIs are members.

CITI's access to high-speed national networking through Rice's connection to the Texas GigaPOP allows members to collaborate, share powerful computing and information resources, and explore networking research. The GigaPOP is a collaboration between Rice, the University of Houston (UH), Baylor College of Medicine, and Texas A&M. The GigaPOP currently has a DS-3 connection to the very-high-speed Backbone Network Service (vBNS) and an OC-3 connection to Internet 2's ABILENE project.

The investigators have access to a large number of workstations and personal computers for conducting this research, including software such as ns-2 and MATLAB. These are networked with the extensive computing resources available in through CITI. In parcirular, for large simulations and testing, the PIs have access to the Rice Terascale Cluster (RTC), which comprises:

280 900MHz 1.5MB Itanium2 processors rack mounted cluster:
1 HP zx6000 dual node with 8GB of DDR RAM and 3 73GB Ultra 160 SCSI HD
5 HP zx6000 dual nodes with 8GB of DDR RAM and 73GB Ultra 160 SCSI HD
20 HP zx6000 dual nodes with 4GB of DDR RAM and 73GB Ultra 160 SCSI HD
106 HP zx6000 dual nodes with 4GB of DDR RAM and 36GB Ultra 160 SCSI HD
4 HP rx5670 quad nodes with 16GB of DDR RAM and 2 73GB Ultra 160 SCSI HD
Interconnect:
96 nodes interconnected with Myrinet2000 (compute)
All nodes connected with 1000 Ethernet (compute)
All nodes connected with 10/100 Ethernet (management)
Foundry FastIron 1500 switch for Ethernet
Scalable Cluster File Server
2 HP zx6000 dual nodes with 8GB of DDR RAM and 2 73GB Ultra 160 SCSI HD
2 0.5TB Ultra 160 SCSI disk arrays attached to data servers with dual fiber channels
Shared Front End
1 HP rx5670 dual node with 8GB of DDR RAM and 2 73GB Ultra SCSI HD
2TB Ultra 160 SCSI RAID 5 disk array attached to data server with dual fiber channels
Front End Tape Storage
HP SureStore 4/60 Ultrium Tape Backup with 4 LDVS Ultrium Tape Drives
Attached to front end data server with dual fiber channels

University of Wisconsin-Madison

Research will be conducted in the new Engineering Hall Building and the Computer Sciences Building, which house the departments of Electrical and Computer Engineering and Computer Science, respectively.

The investigators have access to a large number of workstations and personal computers for conducting this research, including software such as ns-2 and MATLAB. These are networked with the extensive computing resources available in the College of Engineering. Moreover, UW-Madison is home to the Wisconsin Advanced Internet Laboratory (WAIL), which is based in the Computer Sciences Building. WAIL is a one-of-a-kind facility for conducting network and distributed systems research. The vision is

to be able to recreate instances of the Internet from end-to-end-through-core in a laboratory environment. What sets WAIL apart from other network test beds is that real IP networking hardware is used to create the network configurations used in tests. WAIL features over 50 IP routers and switches, 100 end hosts and a variety of other networking gear all housed under one roof. The scope and scale of these components will continue to grow over time to reflect technology trends and enable increasingly complex configurations.

Los Alamos National Laboratory (LANL)

LANL has a long history in the area of high-performance computing and networking. From the perspective of high-performance networking, our experience spans both hardware and software, for example, HiPPI-800 (High-Performance Parallel Interface at 800 Mb/s) in the early 1990s, HiPPI-6400 in the mid-1990s, and Scheduled Transfer Protocol (STP) in the mid-to-late 1990s.

At Los Alamos, Dr. Feng's and Gardner's research laboratory includes the following:

- Two 120-node Linux research clusters, with each node composed of a 1-GHz Transmeta processor with high-performance, code-morphing software, 640-MB memory and 20-GB hard disk, all connected by Fast Ethernet. (Each cluster only occupies six square feet.)
- A 240-node Linux research cluster, with each node composed of a 1-GHz Transmeta processor with high-performance, code-morphing software, 640-MB memory, 20-GB hard disk, and connected by Fast Ethernet. (This cluster only occupies six square feet.)
- An 11-node Linux research cluster, with each node composed of a dual-processor AMD processor (eight nodes at 1.4 GHz; three nodes at 2.0 GHz), 1 GB memory, and connected by Gigabit Ethernet and 10-Gigabit Ethernet.
- An 18-node Linux research cluster with each node composed of a 1.13-GHz Intel Pentium III processor, 512-MB memory, 10-GB hard disk, and connected via Gigabit Ethernet.
- A 5-node Linux research cluster with each node composed of a 1-GHz VIA EPIA processor, 128-MB memory, and connected via Fast Ethernet. (This cluster only occupies a 0.5' x 0.5' x 1' footprint.)
- An 8-node Linux research cluster with each node composed of a dual-processor 500-MHz Intel Pentium II, 1 GB memory, and connected via Gigabit Ethernet.

With respect to networking resources, Dr. Feng's research laboratory has a 10-Gigabit Ethernet testbed along with the following sundry items:

- Six Intel 10-Gigabit Ethernet PCI-X adapters. These resources, in particular, will be used extensively in support of this proposed project.
- One Extreme Networks Summit 7i switch : 28-port Gigabit Ethernet (24 copper and 4 fiber).
- One 3Com SuperStack3 4924 Switch: 28-port Gigabit Ethernet (24 copper and 4 fiber).
- Two 3Com SuperStack3 Switch 4400: 48-port Fast Ethernet.

There is also direct access to the following facilities:

- The Los Alamos Access Grid Room with a projection wall that is 20' wide and 10' tall for distributed collaboration.
- An 18-projector (6 x 3) tiled display wall driven by a 32-node Linux cluster, of which only 18 nodes are used to drive the display. We would use this system for distributed collaborations between Rice University, University of Wisconsin, and SLAC.

Stanford Linear Accelerator Center (SLAC)

SLAC has an OC12 Internet connection to ESnet, and a 1 Gigabit Ethernet connection to Stanford University and thus to CalREN/Internet 2. We have also set up experimental OC192 connections to CalRENII and Level(3). The experimental connections are currently not in service, but have been successfully used at SC2000-2003 to demonstrate bulk-throughput rates from SuperComputing to SLAC and other sites at rates increasing over the years from 990 Mbits/s through 13 Gbps to 23.6 Gbps. SLAC is also part of the ESnet QoS pilot with a 3.5 Mbps ATM PVC to LBNL, and SLAC is connected to the IPv6 testbed with three hosts making measurements for the IPv6 community.[1] SLAC has dark fibers to Stanford University and PAIX, and will be connected to the DoE UltraScience Net. SLAC is also a member of the NSF UltraLight proposal.

SLAC hosts network measurement hosts from the following projects: AMP, NIMI, PingER, RIPE, SCNM, and Surveyor. SLAC has two GPS aerials and connections to provide accurate time synchronization. In addition, the SLAC IEPM group has a small cluster of five high performance Linux hosts with dual 2.4 or 3 GHz processors, 2 GB of memory, a 133 MHz PCI-X bus. Two of these hosts have 10GE Intel interfaces and the other have 1 GE interfaces. These are used for high-performance testing, including the successful SC2003 bandwidth challenge and the Internet 2 Land Speed Records.

SLAC is the home of the BaBar experiment and its tier 0 site. The SLAC data center contains two Sun E6800 20 and 24 symmetric multiprocessors. In addition there is a Linux cluster of over 2400 CPUs and an 800 CPU Solaris cluster. For data storage there are 320 TByte of online disk space and automated access tape storage with a capacity of 10 PetaBytes.

[1] See for example *SLAC IPv6 deployment* presented by by Paola Grosso at the Internet2 Member meeting, Indianapolis Oct.13-16

I. Technology Transfer

Technology transfer

This project will produce new theory, algorithms, methods, and software for analyzing, modeling, processing, and controlling network traffic. This work will be of interest to a wide range of companies in the networking arena, from equipment vendors to bandwidth providers. Our results, embodied in papers, talks, and software are candidates for use in commercial systems.

In the past, we have had considerable success in moving the results of our research into commercial practice. For example, Rice University is currently negotiating world-wide exclusive license with Texas Instruments to six patents relating to high-speed wireless modem design (invented by PI Baraniuk and his PhD student).

We have strong indications of commercial interest from several companies, including Nokia and Texas Instruments. Texas Instruments and Nokia have invested over $9M in the Rice networking, signal processing, and communications programs over the past seven years. These companies thus have a vested interest in examining the results of our research and moving the best results into their products. We will work aggressively to explain our results to networking companies and to encourage their adoption in commercial and research systems. The following section explains our general philosophy on technology transfer and outlines our strategy for getting these results adopted.

The UW group has strong industrial ties including close relationships with Cisco, Intel, Sun, EMC and Spirent who have all made major donations - both monetary and equipment - to the Wisconsin Advanced Internet Laboratory. The UW group also has had close collaborations with SPRINT Research Labs. An important objective of the partnerships with these companies is to develop tools and systems to the point where tech transfer is possible, and at present many opportunities are being explored. Areas of particular interest include all of the operational and management systems being developed for the lab itself, and the Internet measurement and analysis tools and techniques that are currently being investigated by the PIs.

Technology Transfer Path

Our prior experience has convinced us that moving new technology into the marketplace requires:

1. good technology that addresses a real problem,
2. credible experimental results,
3. a well-engineered prototype,
4. a close relationship with industrial/laboratory collaborators.

Our research efforts aim to produce the first three; our close relationships with a number of vendors and our previous successes produce the last one.

To increase the visibility of our results, we maintain a web presence for each of our projects (see the projects under incite.rice.edu, dsp.rice.edu, spin.rice.edu, www.ece.rice.edu/networks, and www.cmc.rice.edu). Through the web project site, outsiders can obtain publications, released software, and general information about the project. We also actively participate in the Rice and Wisconsin Departments of Electrical and Computer Engineering Industrial Affiliates meetings that bring numerous companies to campus to discuss the relationships between our research and their problems. Besides

students in academia, our students are working at companies such as Texas Instruments, Intel, Broadcom, Nokia, Microsoft, Nortel, Lucent, Motorola, and Cisco. This web of professional and personal contacts opens the door through which we can encourage adoption of the best of our work.

Finally, throughout this project, Rice and Wisconsin will work closely with LANL and SLAC. These interactions will provide us the best possible opportunities for transferring our new technologies into tools of real use to DOE.

Proprietary Claims

We plan to distribute the results of this research widely. Software produced will be *open source*. In the event that another party uses these ideas in a commercial product, Rice University and University of Wisconsin may require reasonable licensing to protect both Rice/Wisconsin and the investigators from any liability arising from third-party use.

J. Biographical Sketches

A unique aspect of this project is the inter-disciplinary nature of the research team. The investigators have extensive research experience in both the theoretical and applied aspects of networking, supercomputing, grid computing, statistics, and signal processing, including protocols, traffic models, quality-of-service, tomography, and measurement. A diverse team such as this is much more likely to provide the breakthrough insights required to move forward rapidly in this challenging area.

The investigators have had successful and productive collaborations in the past and are currently collaborating on the DOE-supported INCITE project (see incite.rice.edu), in a large NSF-sponsored Special Projects in Networking initiative, in the DARPA-sponsored Network Models and Simulation program, and in the Texas Instruments-sponsored Leadership University program.

Previous research support for the team over the past five years includes: Baraniuk and Nowak on advanced signal and image models and imaging algorithms (supported by NSF, AFOSR, ONR, and DARPA), Riedi and Baraniuk on multiscale network traffic modeling (supported by NSF, DARPA, and Texas Instruments), Knightly, Baraniuk, and Riedi on proxy traffic modeling and QoS in wireless networks (supported by the NSF Wireless Information Technology and Networks Initiative), Riedi and Baraniuk on scalable ad hoc wireless networking and services (supported by the NSF Special Projects in Networking and the Texas Advanced Technology Program) and Baraniuk, Knightly, Nowak, and Riedi on edge-based service inference and control (supported by the NSF Special Projects in Networking).

Nowak recently moved from Rice to UW-Madison, where he has continued research with the Rice group and initiated new activities with Barford. The addition of Barford to the team brings an additional expertise in network monitoring and measurement to the team.

Detailed CVs follow.

RICHARD G. BARANIUK

Professor of Electrical and Computer Engineering, Rice University, dsp.rice.edu/~richb

Education

University of Manitoba (Canada)	BSc in Electrical Engineering (with distinction), 1987
University of Wisconsin-Madison	MSc in Electrical and Computer Engineering, 1988
University of Illinois-Urbana	PhD in Electrical and Computer Engineering, 1992
Ecole Normale Superieure de Lyon (France)	Postdoc, 1992-1993

Appointments

Professor, Rice University, Houston, TX, 2000-present
Rosenbaum Fellow, Isaac Newton Institute, Cambridge University, 1998
Associate Professor, Rice University, Houston, TX, 1996-2000
Assistant Professor, Rice University, Houston, TX, 1993-1996
Research Assistant, National Research Council of Canada, 1987
R&D Engineer, Omron Tateishi Electronics (Kyoto, Japan), 1986
Sabbatical 2001-2002: Ecole Normale Superieure de Telecommunications (Paris)
Ecole Federale Polytechnique de Lausanne (Switzerland)

Honors

Co-Author on Passive and Active Measurement Workshop Best Student Paper Award, 2003
Fellow of the IEEE, 2002
Co-Author on IEEE Signal Processing Society Junior Paper Award (with M. Crouse, R. Nowak), 2001
IEEE NORSIG Best Paper Award (with E. Monsen, J. Odegard, H. Choi, J. Romberg), 2001
George R. Brown Award for Superior Teaching (Rice), 2001, 2003
University of Illinois ECE Young Alumni Achievement Award, 2000
Charles Duncan Junior Faculty Achievement Award (Rice), 2000
C. Holmes MacDonald National Outstanding Teaching Award (Eta Kappa Nu), 1999
Rosenbaum Fellowship, Isaac Newton Institute (Cambridge University), 1998
Office of Naval Research Young Investigator Award, 1995
National Science Foundation National Young Investigator Award, 1994
National Sciences and Engineering Research Council of Canada NATO Postdoctoral Fellowship, 1992

Synergistic Activities

Project Director: *DOE INCITE Project* – Multiscale network traffic measurement and analysis project (collaboration with Stanford Linear Accelerator Center and Los Alamos National Laboratory)
Director: *Connexions Project* – an open, community-based education project (cnx.rice.edu)
Editorial Board: *Applied and Computational Harmonic Analysis*
Chair: *IEEE Signal Processing Society*, Houston Chapter
IEEE Signal Processing Society Technical Committee on Theory and Methods

Selected Publications (see also dsp.rice.edu/publications)

V. J. Ribeiro, R. Riedi, and R. G. Baraniuk, "Spatio-Temporal Available Bandwidth Estimation with pathChirp," *ACM Sigmetrics/Performance* (poster), 2004.

V. Delouille, R. Neelamani, and R. G. Baraniuk, "Robust Distributed Estimation in Sensor Networks using the Embedded Triangles Algorithm," *Int. Symp. Integrated Processing in Sensor Networks*, Berkeley, CA, April 2004.

V. J. Ribeiro, R. H. Riedi, and R. G. Baraniuk, "Internet Path Modeling and Analysis using *pathChirp*," *Passive and Active Measurement Workshop*, San Diego, April 2003.

P. Abry, R. G. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, "Multiscale Nature of Network Traffic," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 28-46, May 2002.

S. Sarvotham, R. Riedi, and R. G. Baraniuk, "Connection-level Analysis and Modeling of Network Traffic," *ACM SIGCOMM Internet Measurement Workshop*, San Francisco, November, 2001.

V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk, "Multiscale Queuing Analysis of Long-Range-Dependent Network Traffic," *IEEE INFOCOM*, Tel Aviv, Israel, March 2000.

V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk, "Simulation of Non-Gaussian Long-Range-Dependent Traffic using Wavelets," *ACM SIGMETRICS*, Atlanta, May 1999.

P. Abry, R. G. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, "Multiscale Nature of Network Traffic," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 28-46, May 2002.

R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, "A Multiplicative Wavelet Model with Application to TCP Network Traffic," *IEEE Transactions on Information Theory*, Vol. 45, pp. 992-1018, April 1999.

M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, ``Wavelet-based Statistical Signal Processing using Hidden Markov Models," *IEEE Transactions on Signal Processing*, Vol. 46, No. 4, April 1998.

Doctoral Students

Current: N. Ahmed, W. Chan, S. Lavu, W. Mantzel, V. Ribiero, S. Sarvotham, M. Wakin, R. Wagner

Completed: J. Romberg, now a postdoc at Caltech

R. Neelamani, now at Exxon-Mobil Research, Houston, 1997-2003

Tim Dorney, now at Texas Instruments, 1996-2001

Rohit Gaikwad, now at Broadcom, San Jose, CA (winner of the Herschel Rich Invention Award, 1999, and Budd Award for best PhD thesis in Rice College of Engineering, 2000), 1996-2000

Roger Claypoole, now Associate Professor at Air Force Institute of Technology (AFIT), Dayton, OH, 1996-1999

Matthew Crouse, now at Duke Energy, Houston, Texas (winner of the Budd Award for best PhD thesis in Rice College of Engineering, 1999, and IEEE Signal Processing Society Junior Best Paper Award, 2001), 1995-1999

Postdocs

Current: Veronique Dellouile (UK-Louvain, Belgium), Dror Baron (Univ. Illinois)

Rutger van Spaendonck (from Delft Univ. Technology, now at Shell Research), 2003

Xin Wang (from Princeton University, now at Lucent Bell Labs), 2002

Maarten Jansen (from University of Leuven, now faculty at Eindhoven Univ. Technology), 2000-2001

Mark Coates (from Cambridge University, now faculty at McGill University), 1999-2000

Hyeoko Choi (from University of Illinois, now Research Faculty Fellow at Rice), 1998-2000

Rolf Riedi (from ETH, Zurich, now Assoc. Prof. of Statistics at Rice), 1997-1999

Jan Odegard (from Rice University, now Exec. Director of CITI, Rice University), 1997-1999

Phillippe Steeghs (from Delft University of Technology, Netherlands), 1997-1999

Ivan Magrin-Chagnolleau (now with CNRS, Lyon, France), 1998-1999

Robert Nowak (now Associate Professor at University of Wisconsin), 1995-1996

Paulo Goncalves (now with INRIA, Grenoble, France), 1994-1996

External Theses

Pier Luigi Dragotti (Ecole Polytechnique Federale de Lausanne, Switzerland), 2002

Minh Do (Ecole Polytechnique Federale de Lausanne, Switzerland), 2001

Pierre Chanais (Ecole Normale Superieure de Lyon, France), 2001

Philippe Steeghs (Delft University of Technology), 1998

Eric Chassande-Motin (Ecole Normal Superieure de Lyon, France), 1998

Edward W. Knightly

Rice University
Department of Electrical and Computer Engineering, MS 380, Houston, TX 77005
Email: knightly@ece.rice.edu

Research Interests

High-performance protocol design, mobile and wireless networks, quality of service, and performance evaluation.

Education

University of California at Berkeley, Ph.D. received 1996 (EECS).
University of California at Berkeley, M.S. received 1992 (EECS).
Auburn University, B.S. received 1991 (EE).

Positions

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, 2003 – 2004. *Visiting Professor*.
Rice University, Houston, TX, August 1996 - present. *Associate Professor*.
University of California at Berkeley, Berkeley, CA, April 1992 - August 1996. *Postgraduate Researcher and member of the Tenet Group (advisor: Domenico Ferrari).*
Sandia National Laboratories, Livermore, CA, summers, May 1993 - August 1996. *Research Consultant for the Networks and Distributed Computing Group.*
Auburn University, Auburn, AL, March 1991 - August 1991. *Research Assistant.*
Columbia University, New York, NY, June 1989 - August 1989. *Research Assistant at the Centerfor Telecommunications Research.*

Honors and Awards

Sloan Fellowship, 2001.
NSF CAREER Award, 1997.
Premium Paper Award, *Distributed Systems Engineering Journal*, 1997.
Sandia National Laboratories Graduate Fellowship, 1993-1996.
University Scholarship, University of California at Berkeley, 1991.
Received Amoco, Citizen's, H.K. Porter, and A.H. Skinner scholarships, 1987-1991.
Selected by peers as 1991 Outstanding Graduating Senior in Electrical Engineering.
Awarded 1989 Outstanding Undergraduate in Physics.

Professional Activities

Technical co-chair, IEEE INFOCOM 2005; Finance Chair, ACM/IEEE MOBICOM 2002, 2003; Tutorial co-chair, ACM MOBIHOC 2003; IEEE ICNP 2001, Program Chair; 1998 IEEE/IFIP International Workshop on Quality of Service; Steering Committee, IEEE/IFIP International Workshop on Quality of Service, 1999 - present.

Editor, *IEEE/ACM Transactions on Multimedia*, 2001 - present, *Computer Networks Journal*, 2000 – present, *IEEE/ACM Transactions on Networking*, 2000 - present, *IEEE Network*, 1999 – present,, Special Issue of *IEEE Network* on Integrated and Differentiated Services for the Internet, September 1999.

Panelist, ACM MOBICOM 2002, (*Evolution of Mobile and Wireless Networks: Enablers and Inhibitors*), IEEE INFOCOM 2002, (*QoS Research in a Complicated World*), NSF Gigabit Kits

Workshop-June 2002, (*Towards a Better Infrastructure for Networking and Distributed Systems Research*), Internet2/DOE QoS Workshop – February 2000, (*New Directions for Internet2 QoS*).

Technical Program Committee Member, IEEE INFOCOM '98, '99 and 2000, IEEE/IFIP IWQoS '99 and 2000, and SIGMETRICS 2000.

Relevant Publications

1) V. Gambiroza, P. Yuan, L. Balzano, Y. Liu, S. Sheafor, and E. Knightly, "Design, Analysis, and Implementation of DVSR: A Fair, High Performance Protocol for Packet Rings," *IEEE/ACM Transactions on Networking,* 12(1), February 2004.
2) V. Gambiroza, P. Yuan, and E. Knightly, "The IEEE 802.17 Media Access Protocol for High-Speed Metropolitan-Area Resilient Packet Rings," to appear in *IEEE Network*.
3) Kuzmanovic and E. Knightly, "Low-Rate TCP-Targeted Denial of Service Attacks (The Shrew vs. the Mice and Elephants)," in *Proceedings of ACM SIGCOMM 2003*, Karlsruhe, Germany, August 2003.
4) Kuzmanovic and E. Knightly, "TCP-LP: A Distributed Algorithm for Low Priority Data Transfer," in *Proceedings of IEEE INFOCOM 2003*, San Francisco, CA, April 2003.
5) H. Fu and E. Knightly, "A Simple Model of Real-Time Flow Aggregation," *IEEE/ACM Transactions on Networking,* 11(3):422-435, June 2003.
6) Kuzmanovic and E. Knightly, "Low-Rate TCP-Targeted Denial of Service Attacks (The Shrew vs. the Mice and Elephants)," in *Proceedings of ACM SIGCOMM 2003*, Karlsruhe, Germany, August 2003.
7) Kuzmanovic and E. Knightly, "TCP-LP: A Distributed Algorithm for Low Priority Data Transfer," in *Proceedings of IEEE INFOCOM 2003*, San Francisco, CA, April 2003.
8) V. Gambiroza, Y. Liu, P. Yuan, and E. Knightly, "High-Performance Fair Bandwidth Allocation for Resilient Packet Rings" in Proceedings of the 15th ITC Specialist Seminar on Internet Traffic Engineering and Traffic Management, Wurzburg, Germany, July 2002.
9) H. Fu and E. Knightly, "A Simple Model of Real-Time Flow Aggregation," *IEEE/ACM Transactions on Networking,* 11(3):422-435, June 2003.
10) B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic Media Access for Multirate Ad Hoc Networks," in Proceedings of ACM MOBICOM 2002, Atlanta, GA, September, 2002.
11) V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, "Ordered Packet Scheduling in Wireless Ad Hoc Networks: Mechanisms and Performance Analysis," in Proceedings of ACM MOBIHOC 2002, Lausanne, Switzerland, June 2002.
12) V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, "Distributed Priority Scheduling and Medium Access in Ad Hoc Networks," ACM Wireless Networks Journal (WINET): Special Issue on Selected Papers from MOBICOM 2001, 8(6):455-466, November 2002.
13) S. Sargento, R. Valadas, and E. Knightly, "Resource Stealing in Endpoint Controlled Multi-class Networks," in *Proceedings of IWDC 2001: Evolutionary Trends of the Internet*, Taormina, Italy, September 2001 (Invited Paper).
14) C. Cetinkaya, V. Kanodia, and E. Knightly, "Scalable Services via Egress Admission Control," in *IEEE Transactions on Multimedia: Special Issue on Multimedia over IP*, 3(1):69-81, March 2001.
15) J. Schlembach, A. Skoe, P. Yuan, and E. Knightly, "Design and Implementation of Scalable Admission Control," in *Proceedings of the International Workshop on QoS in Multiservice IP Networks*, Rome, Italy, January 2001.

Rudolf H. Riedi

Assistant Professor
Departments of Statistics and of
Electrical and Computer Engineering
Rice University
Houston, TX 77251-1892
Tel: (713) 348 3020, Fax: (713) 348 5476, Email: riedi@rice.edu

EDUCATION

1993	Ph.D.in Mathematics	Federal Institute of Technology ETH, Zurich--Switzerland
1986	M.S.in Mathematics	ETH, Zurich—Switzerland, (winner of ETHZ Polya Prize)

POSITIONS

2003-present	Rice University, Houston, TX	Assistant Professor
1997-2003	Rice University, Houston, TX	Faculty Fellow
1995-1997	INRIA (France)	Research Associate
1993-1995	Yale University, New Haven, CT	Postdoctoral Research Fellow
1987-1993	ETH Zurich (Switzerland)	Research Assistant

PROFESSIONAL ACTIVITIES AND AFFILIATIONS

Member of	American Mathematical Society, Mathematical Society of Switzerland, IEEE Information Theory and Signal Processing Societies. Institute of Mathematical Statistics American Statistical Association
Consulting with	AT&T Research Labs, Florham Park, NJ
Long term visitor at	Newton Institute of Mathematical Science, Cambridge, UK Dept. of Computer Science, UFMG, Brazil Dept. of Physics, ENS, Lyon, France Dept. of ECE, Melbourne, Australia
Associate Editor	Special issue on Signal Processing in Networking, IEEE SP
Technical program	InfoComm (2003-2005)

RESEARCH SUPPORT

2003 NSF (PI)	SAFARI: A scalable architecture for ad hoc networking and services
2003 ATP (co-PI)	A scalable architecture for ad hoc networking and services
2001 DoE (co-PI)	INCITE: Edge-based Processing and Inference for High-Performance Networks
2000 NSF (co-PI)	A Framework for Edge-Based Processing and Inference
2000 DARPA, (co-PI)	Multiscale Traffic Processing for Inference and Control
1999 NSF (co-PI)	Seamless Multitier Wireless Networks

AWARDS and HONORS

1995 Centre National des Etudes en Telecommunications, France (research grant)
1993 National Science Foundation of Switzerland (postdoctoral fellowship)
1986 ETHZ Polya prize (best grades in mathematics of the year)

TEN RELEVANT PUBLICATIONS

pathChirp: Efficient Available Bandwidth Estimation for Network Paths
Vinay J. Ribeiro, Rudolf H. Riedi, Jiri Navratil, Les Cottrell, and Richard G. Baraniuk
Proceedings Workshop on Passive and Active Measurement PAM200
Best Student Paper Award

Optimal Sampling Strategies for Multiscale Models with Application to Network Traffic Estimation
Vinay J. Ribeiro, Rudolf H. Riedi and Richard G. Baraniuk
Proceedings Workshop on Statistical Signal Processing SSP03, St. Louis, MO, Sept 2003

Network Traffic Modeling using Connection-Level Information
Xin Wang, Shriram Sarvotham, Rudolf H. Riedi, and Richard G. Baraniuk
Proceedings SPIE ITCom, Boston, MA, August 2002

Network Traffic Analysis and Modeling at the Connection Level
S. Sarvotham, R. Riedi, and R. Baraniuk
Proceedings IEEE/ACM SIGCOMM Internet Measurement Workshop 2001, San Francisco, CA.

The Multiscale Nature of Network Traffic: Discovery, Analysis, and Modelling
Patrice Abry, Richard Baraniuk, Patrick Flandrin, Rudolf Riedi, Darryl Veitch
IEEE Signal Processing Magazine vol 19, no 3, pp 28-46 (May 2002).

Long-Range Dependence and Data Network Traffic
W. Willinger, V. Paxson, R. H. Riedi and M. S. Taqqu,
in *Long range dependence : theory and applications*, ISBN: 0817641688.

Multiscale Queuing Analysis of Long-Range-Dependent Network Traffic
V. J. Ribeiro, R. H. Riedi, M. S. Crouse and R. G. Baraniuk
IEEE Trans. on Networking, submitted

Toward an Improved Understanding of Network Traffic Dynamics
R. H. Riedi and W. Willinger
in: *Self-similar Network Traffic and Performance Evaluation*
eds. Park and Willinger, (Wiley 2000), chapter 20, pp 507-530.

A Multifractal Wavelet Model with Application to Network Traffic
R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk
IEEE Special Issue on Information Theory, **45**. (April 1999), 992-1018.

A Hierarchical and Multiscale Analysis of E-Business Workloads
Daniel Menascé, Virgílio Almeida, Rudolf Riedi, Flávia Ribeiro, Rodrigo Fonseca, Wagner Meira Jr.,
Performance Evaluation **54**(1), Sept 2003, pp 33--57.

RECENT COLLABORATORS
P. Abry, V. Almeida, R. Baraniuk, L. Cottrell, P. Druschel, W. Feng, P. Flandrin, A. Hero, D. Johnson, B. Mandelbrot, S. Marron, I. Norros, V. Paxson, M. Taqqu, D. Veitch, W. Willinger.

Robert D. Nowak

University of Wisconsin-Madison, ECE Department, 1415 Engineering Dr., Madison, WI 53706
Email: nowak@engr.wisc.edu, Web: www.ece.wisc.edu/~nowak

Research Interests

Statistical signal and image processing, communication systems, network inference and tomography

Education

B.S. (highest distinction), 1990, M.S., 1991, Ph.D, 1995, Electrical Engineering, University of Wisconsin-Madison

Recent Honors and Awards

Office of Naval Research Young Investigator Award, 2000
Army Research Office Young Investigator Award, 1999
Invited Visiting Fellowship, Isaac Newton Institute for Mathematical Sciences, Cambridge University, Cambridge, UK, July - August 1998.
National Science Foundation CAREER Award, 1997
Rockwell International Fellow, 1991-1995
General Electric Genius of Invention Award, 1994

Professional Experience

Associate Professor
Electrical and Computer Engineering, University of Wisconsin - 2003-present

Adjunct Professor
Electrical and Computer Engineering, Rice University - 2003-present

Assistant Professor and Associate Professor
Electrical and Computer Engineering, Rice University - July 1999 to 2003

Assistant Professor
Electrical Engineering, Michigan State University - August 1996 - July 1999

Postdoctoral Research Fellow
Electrical and Computer Engineering, Rice University - 1995-1996

Research Scientist
General Electric Medical Systems, Milwaukee, WI - Summers 1987-1991

Patents

R. Nowak and H. Hu, "An Improved Z-wedge Lowpass Filter Algorithm for Reducing Incomplete Data Artifacts in Volume Computed Tomography Images," U.S. Patent #5400377.

Professional Activities

Member, IEEE Signal Processing Society Technical Committee
Member, IEEE International Conference on Acoustics, Speech and Signal Processing Organizing Committee
Organizer, IEEE ICASSP Special Session on Network Inference and Traffic Modeling, Salt Lake City, UT, 2001

Reviewer:

Journal of the American Statistical Association
IEEE Transactions on Image Processing, Signal Processing, Information Theory
IEEE INFOCOM 2001

Ten Relevant Publications

M. Coates and R. Nowak, ``Sequential Monte Carlo Inference of Internal Delays in Nonstationary Communication Networks," IEEE Transactions on Signal Processing, Special Issue on Monte Carlo Methods for Statistical Signal Processing, vol. 50, pp. 366-376, 2002.

M. Coates, A. Hero, R. Nowak, and B. Yu,``Internet Tomography," IEEE Signal Processing Magazine, Special Issue on Signal Processing in Networking, vol. 19, pp. 47-65, 2002.

M. Coates and R. Nowak, "Network Loss Inference using Unicast End-to-end measurement," Proceedings of ITC Conference on IP Traffic, Modelling and Management," Monterey CA, September, 2000.

M. Coates and R. Nowak, "Networks for Networks: Internet Analysis using Bayesian Graphical Models," Proceedings of IEEE Neural Network for Signal Processing Workshop, Sidney Australia, December, 2000.

R. Nowak and E. Kolaczyk, ``A Bayesian Multiscale Framework for Poisson Inverse Problems," IEEE Transactions on Information Theory, Aug. 2000.

R. Nowak, ``Multiscale Hidden Markov Models for Bayesian Image Analysis," Bayesian Inference in Wavelet Based Models, Springer-Verlag, 1999, (Editors: B. Vidakovic and P. Muller).

R. Nowak and R. Baraniuk, ``Nonlinear Wavelet-Based Transformations for Signal Analysis and Processing," IEEE Transactions on Signal Processing, July 1999.

R. Nowak, ``Wavelet-Based Rician Noise Removal for Magnetic Resonance Imaging," IEEE Transactions on Image Processing, October 1999.

K. Timmermann and R. Nowak, ``Multiscale Modeling and Estimation of Poisson Processes with Application to Photon-Limited Imaging," IEEE Transactions on Information Theory, April 1999.

R. Nowak, ``Penalized Least Squares Estimation of Volterra Filters and Higher Order Statistics," IEEE Transactions on Signal Processing, Feb. 1998.

Paul Barford

Computer Science Department
University of Wisconsin - Madison
1210 West Dayton Street
Madison, WI 53706
phone: (608) 262-6609
fax: (617) 265-2635
email: pb@cs.wics.edu

Professional Preparation
Ph.D. Computer Science Boston University, Boston, MA December, 2000
B.S. Electrical Engineering University of Illinois, Urbana, IL May, 1985

Appointments
Computer Science Department, Univ. of Wisconsin, Madison, WI January, 2001 -
Assistant Professor; Director, Wisconsin Advanced Internet Laboratory
Computer Science Department, Boston University, Boston,MA 1995 - 2000
Research Fellow
DeGeorge Financial, Inc., Cheshire, CT 1991 - 1995
Director of Planning and Research
Digital Equipment Corporation, Maynard, MA 1987 - 1991
Senior Engineer

Related Publications
1. Paul Barford and Joel Sommers. A Comparison of Probe-based and Router-based Methods for Measuring Packet Loss", Submitted for Publication, February, 2004.
2. Vinod Yegneswaran, Paul Barford and David Plonka. On the Design and Use of Internet Sinks for Network Intrusion Monitoring", Submitted for Publication, January, 2004.
3. Joel Sommers, Hyungsuk Kim, and Paul Barford. Harpoon: A Flow-Level Tra_c Generator for Router and Network Tests" To appear in Proceedings of ACM SIGMETRICS, poster, New York NY, June,2004.
4. Vinod Yegneswaran, Paul Barford and Somesh Jha. Global Intrusion Detection in the DOMINO Overlay System", In Proceedings of ISOC Network and Distributed Systems Security Symposium (NDSS '04), February, 2004.
5. Paul Barford and Larry Landweber. Bench-style Network Research in an Internet Instance Laboratory", In ACM SIGCOMM Computer Communications Review, 33(3), July, 2003.
6. Vinod Yegneswaran, Paul Barford and Johannus Ullrich. Internet Intrusions: Global Characteristics and Prevalence" In Proceedings of ACM SIGMETRICS, San Diego, CA, June,2003.
7. Paul Barford, Je_ Kline, David Plonka and Amos Ron. A Signal Analysis of Network Trace Anomalies", In Proceedings of ACM SIGCOMM Internet Measurement Workshop '02, Marseille, France, November, 2002.
8. Jim Gast and Paul Barford. Resource Deployment based on Autonomous System Clustering", In Proceedings of IEEE Globcom '02, Taipei, Taiwan, Octpber, 2002.
9. Paul Barford and David Plonka. Characteristics of Network Trace Flow Anomalies", In Proceedings of ACM SIGCOMM Internet Measurement Workshop '01, San Francisco, CA, November, 2001.
10. Paul Barford, Azer Bestavros, John Byers and Mark Crovella. On the Marginal Utility of Network Topology Measurements", In Proceedings of ACM SIGCOMM Internet Measurement Workshop '01, San Francisco, CA, November, 2001.

Professional Activities
_ Board Member, National LambdaRail.
_ Program Committee, ACM WORM 2004.
_ Program Committee, ACM SIGCOMM 2004.
_ Organizing Committee, OpenSig 2003.
_ Organizing Committee, IEEE ICNP 2003.
_ Program Committee, ACM SIGMETRICS 2003, 2004.
_ Program Committee, ACM SIGCOMM Internet Measurement Conference, 2001, 2004.
_ Program Committee, IEEE Workshop on Internet Applications 2003.
_ Organizing Committee, Computer Science and Telecommunications Board of the National Research Council study on Internet Under Crisis Conditions," 2002.
_ Organizing Committee, Institute for Pure and Applied Mathematics workshop on \Large Scale Communications Networks: Topology, Routing, Trace and Control," 2002.
_ Program Committee, IEEE 6th International Computer Performance and Dependability Symposium, Washington DC, 2002.
_ Program Committee, Multiresolution Analysis of the Global Internet Workshop, Palo Alto, CA, 2000.
_ Program Committee, Boston University Workshop on Internet Measurement, Instrumentation and Characterization, Boston, MA, 1999.

Collaborators
John Byers (Boston University), Mark Crovella (Boston University), David Donoho (Stanford University), Nick Duffield (AT&T), Jay Lepreau (Utah), Nick McKeown (Stanford), Vern Paxson(ACIRI), Walter Willinger(AT&T)

Thesis Advisor
Mark Crovella (Boston University)

Ph.D. Students
Shilpi Agarwal, Ryan Kern, Joel Sommers, Janani Thanigachalam, Badhri Varanasi, Vinod Yegneswaran

Wu-chun Feng

Professional Preparation

Penn State University	Computer Engineering	B.S. (*summa cum laude*), 1988
Penn State University	Computer Engineering	M.S., 1990
University of Illinois at Urbana-Champaign	Computer Science	Ph.D., 1996

Appointments

Team Leader, Los Alamos National Laboratory, 2000-
Adjunct Assistant Professor, The Ohio State University, 2000-2003
Technical Staff Member, Los Alamos National Laboratory, 1998-
Institute Fellow, Los Alamos Computer Science Institute, 1998-
Adjunct Assistant Professor, Purdue University, 1998-2000
Research Scientist, Vosaic Corporation, 1997
Visiting Assistant Professor, University of Illinois at Urbana-Champaign, 1996-1998
Research Consultant, NASA Ames Research Center, 1993
Applications Researcher, IBM T.J. Watson Research Center, 1990
Teaching Fellow, Penn State University, 1988-1989

Related Publications (Available at www.lanl.gov/radiant)

[1] J. Hurwitz and W. Feng, "End-to-End Performance of 10-Gigabit Ethernet on Commodity Systems," *IEEE Micro*, January-February 2004.

[2] M. Gardner, W. Deng, T. S. Markham, C. Mendes, W. Feng, and D. Reed, "A High-Fidelity Software Oscilloscope for Globus," *GlobusWORLD 2004*, January 2004.

[3] M. Veeraraghavan, X. Zheng, H. Lee, M. Gardner, and W. Feng, "CHEETAH: Circuit-Switched High-Speed End-to-End Transport ArcHitecture," Best Paper Award, *SPIE/IEEE Optical Networking and Computer Communications Conference*, Dallas, TX, October 2003.

[4] W. Feng, "Green Destiny + mpiBLAST = Bioinfomagic," *10th International Conference on Parallel Computing 2003: Bioinformatics Minisymposium*, September 2003.

[5] M. Gardner, W. Feng, M. Broxton, A. Engelhart, and J. Hurwitz, "MAGNET: A Tool for Debugging, Analysis and Adaptation in Computing Systems," *3rd IEEE International Symposium on Cluster Computing and the Grid*, Tokyo, Japan, May 2003.

[6] M. Gardner, M. Broxton, A. Engelhart, and W. Feng, "MUSE: A Software Oscilloscope for Clusters and Grids," *17th IEEE International Parallel & Distributed Processing Symposium*, Nice, France, April 2003.

Other Relevant Publications

[1] A. Engelhart, M. Gardner, and W. Feng, "Re-Architecting Flow-Control Adaptation for Grid Environments," *18th IEEE International Parallel & Distributed Processing Symposium*, Santa Fe, NM, April 2004.

[2] S. Ayyorgun and W. Feng, "A Deterministic Characterization of Network Traffic for Average Performance Guarantees," *38th Annual Conference on Information Sciences and Systems (CISS'04)*, March 2004.

[3] W. Feng, "Making a Case for Efficient Supercomputing," *ACM Queue*, October 2003.

[4] W. Feng, M. Gardner, M. Fisk, and E. Weigle, "Automatic Flow-Control Adaptation for Enhancing Network Performance in Computational Grids," *Journal of Grid Computing*, Vol. 1, No. 1, 2003.

[5] A. Darling, L. Carey, and W. Feng, "The Design, Implementation, and Evaluation of mpiBLAST," Best Paper: Applications Track, *ClusterWorld Conference & Expo 2003* in conjunction with the *4th International Conference on Linux Clusters: The HPC Revolution 2003*, June 2003.

[6] S. Thulasidasan, W. Feng, and M. Gardner, "Optimizing GridFTP Through Dynamic Right-Sizing," *IEEE Symposium on High-Performance Distributed Computing*, June 2003.

[7] W. Feng and S. Vanichpun, "Ensuring Compatibility Between TCP Reno and TCP Vegas," *IEEE Symposium on Applications and the Internet*, January 2003.

[8] F. Petrini, W. Feng, A. Hoisie, S. Coll, and E. Frachtenberg, "The Quadrics Network (QsNet): High-Performance Clustering Technology" (Extended Version), *IEEE Micro*, January-February 2002.

[9] F. Petrini and W. Feng, "Improved Resource Utilization with Buffered Coscheduling," *Journal of Parallel Algorithms & Applications* (Special Issue), Vol. 16, 2001.

Synergistic Activities

- *Recent Program Chairs and Vice Chairs:* 34th International Conference on Parallel Processing (2005); DOE Workshop on Ultra High-Speed Transport Protocols and Dynamic Network Provisioning for Large-Scale Scientific Applications (2003); 28th International Conference on Parallel Processing (1999).
- *Recent Program Committees:* 1st-4th IEEE Workshop on Communication Architectures for Clusters in conjunction with the IEEE Int'l Parallel & Distributed Processing Symposium (2001-2004); 1st Workshop on Grids and Advanced Networks in conjunction with the 4th IEEE/ACM Symposium on Cluster Computing and the Grid (2004), 2nd International Workshop on Protocols for Fast Long-Distance Networks (2004); 26th-28th IEEE International Conference on Local Computer Networks (2001-2003); 10th and 12th IEEE International Symposium on High-Performance Distributed Computing (2001, 2003); 17th IEEE International Parallel & Distributed Processing Symposium (2003); 30th International Conference on Parallel Processing (2001).
- *Recent Awards and Honors:* On the Road to a Gigabit Award for Partnership, Sponsored by the Corporation of Education Network Initiatives in California (CENIC) and California Institute for Telecommunications and Information Technology, Cal-(IT)2 (2004); Asian-American Engineer of the Year Award (2004); Sustained Bandwidth Award a.k.a. "Moore's Law Move Over!" Award, SC2003 Bandwidth Challenge (2003); R&D 100 Award for *Green Destiny: Super-Efficient Supercomputing* (2003); Best Paper Award, SPIE/IEEE Optical Networking and Computer Communications Conference (2003); On the Road to a Gigabit Award: Biggest, Fastest in the West, Sponsored by CENIC and Cal-(IT)2 (2003).

Dr. Feng is the author of over 80 research papers in the general area of high-performance computing and networking. Four projects of recent note include [1] Green Destiny, a tiny 240-node supercomputer in six square feet that sips as little as 3.2 killowatts of power and requires no special infrastructure to operate (http://sss.lanl.gov), [2] mpiBLAST, an open-source bioinformatics code that delivers super-linear speed-up in parallel computing systems (http://mpiblast.lanl.gov), [3] 10-Gigabit Ethernet for networks of workstations, clusters, and grids – a project which then led to key collaborations with the high-energy physics community and the subsequent smashing of the Internet2 Land Speed Record, and [4] MAGNET, a systems-level toolkit for enabling (high-fidelity) "software oscilloscope" functionality into distributed computing systems.

Collaborators

W. Allcock (ANL), R. Baraniuk (Rice), M. Beck (Tennessee), L. Bhuyan (UC-Riverside), A. Chien (UCSD), R. L. Cottrell (SLAC), I. Foster (ANL), S. Kim (Purdue), E. Knightly (Rice), L. Liu (Illinois), S. Low (Caltech), O. Martin (CERN), B. Mukherjee (UC-Davis), H. Newman (Caltech), R. Nowak, (U. Wisconsin), D. Panda (Ohio State University), R. Riedi (Rice), C. Stewart (Indiana), B. Tierney (LBNL), M. Veeraraghavan (Virginia).

Graduate Advisor: Jane W.-S. Liu, University of Illinois at Urbana-Champaign (now at Microsoft Corporation)

Recent Graduate Student and Postdoctoral Advisees
J. Archuleta (UC-Berkeley/LANL), S. Ayyorgun (UCSD/LANL), M. Broxton (MIT/UC-Berkeley), L. Carey (SUNY), A. Darling (U. Wisconsin), C. Hsu (Rutgers/LANL), J. Hurwitz (St. John's College), A. Kapadia (Illinois), H. Kettani (Jackson State), F. Moraes (Columbia), F. Petrini (U. Pisa/LANL), U. Syyid (Hughes Network Systems), S. Vanichpun (U. Maryland), E. Weigle (UCSD), X. Zheng (Virginia).

Contact Information

Computer & Computational Sciences Division
Los Alamos National Laboratory
P.O. Box 1663, M.S. D451
Los Alamos, NM 87545

Phone: +1-505-665-2730
FAX: +1-505-665-4934
E-mail: feng@lanl.gov
WWW: http://public.lanl.gov/feng

Mark K. Gardner

Professional Preparation

Brigham Young University	Mechanical Engineering	B.S. (*summa cum laude*), 1986
Brigham Young University	Computer Science	M.S., 1994
University of Illinois at Urbana-Champaign	Computer Science	Ph.D., 1999

Appointments

Technical Staff Member, Los Alamos National Laboratory, 1999-Present
Applications Researcher, U.S. Army Construction Engineering Research Laboratory, 1995
Aerodynamicist, Allied-Signal Aerospace, Garrett Auxilliary Power Division, 1986-1991

Related Publications (Available at www.lanl.gov/radiant)

[1] M. Gardner, W. Deng, T. S. Markham, C. Mendes, W. Feng, and D. Reed, "A High-Fidelity Software Oscilloscope for Globus," *GlobusWORLD 2004*, January 2004.

[2] M. Veeraraghavan, X. Zheng, H. Lee, M. Gardner, and W. Feng, "CHEETAH: Circuit-Switched High-Speed End-to-End Transport ArcHitecture," Best Paper Award, *SPIE/IEEE Optical Networking and Computer Communications Conference*, Dallas, TX, October 2003.

[3] M. Gardner, W. Feng, M. Broxton, A. Engelhart, and J. Hurwitz, "MAGNET: A Tool for Debugging, Analysis and Adaptation in Computing Systems," *3rd IEEE International Symposium on Cluster Computing and the Grid*, Tokyo, Japan, May 2003.

[4] M. Gardner, M. Broxton, A. Engelhart, and W. Feng, "MUSE: A Software Oscilloscope for Clusters and Grids," *17th IEEE International Parallel & Distributed Processing Symposium*, Nice, France, April 2003.

Other Relevant Publications

[1] A. Engelhart, M. Gardner, and W. Feng, "Re-Architecting Flow-Control Adaptation for Grid Environments," *18th IEEE International Parallel & Distributed Processing Symposium*, Santa Fe, NM, April 2004.

[2] W. Feng, M. Gardner, M. Fisk, and E. Weigle, "Automatic Flow-Control Adaptation for Enhancing Network Performance in Computational Grids," *Journal of Grid Computing*, Vol. 1, No. 1, 2003.

[3] S. Thulasidasan, W. Feng, and M. Gardner, "Optimizing GridFTP Through Dynamic Right-Sizing," *IEEE Symposium on High-Performance Distributed Computing*, June 2003.

[5] W. Feng, M. Gardner, and J. Hay, "The MAGNeT Toolkit: Design, Evaluation, and Implementation," *Journal of Supercomputing*, (23)1, August 2003, pp. 67-79.

Synergistic Activities

- *Recent Awards and Honors:* Sustained Bandwidth Award a.k.a. "Moore's Law Move Over!" Award, SC2003 Bandwidth Challenge (2003); R&D 100 Award for *Green Destiny: Super-Efficient Supercomputing* (2003); Best Paper Award, SPIE/IEEE Optical Networking and Computer Communications Conference (2003).

Dr. Gardner is the author of over 20 research papers in the general area of high-performance computing and networking. Four projects of recent note include [1] MAGNET, a systems-level toolkit for enabling (high-fidelity) "software oscilloscope" functionality into distributed computing systems, and [2] Dynamic Right-Sizing (DRS), an automatic flow-control adaptation mechanism for TCP that eliminates the need for manual tuning to achieve high throughput on high-performance WANs, [3] drsFTP, an implementation of DRS in user-space for FTP that exhibits up to an 8-fold increase in throughput without the need to hand-tune buffers, [4] DRS-enabled GridFTP, an implementation of DRS in GridFTP that brings the automatic tuning of TCP buffers to the Grid community.

Collaborators

W. Allcock (ANL), R. Baraniuk (Rice), M. Beck (Tennessee), A. Chien (UCSD), R. L. Cottrell (SLAC), I. Foster (ANL), S. Low (Caltech), D. Panda (Ohio State University), R. Riedi (Rice), B. Tierney (LBNL), M. Veeraraghavan (Virginia).

Graduate Advisor: Jane W.-S. Liu, University of Illinois at Urbana-Champaign (now at Microsoft Corporation)

Contact Information

Computer & Computational Sciences Division
Los Alamos National Laboratory
P.O. Box 1663, M.S. D451
Los Alamos, NM 87545

Phone: +1-505-665-4953
FAX: +1-505-665-4934
E-mail: mkg@lanl.gov
WWW: http://public.lanl.gov/mkg

Roger Leslie Anderton Cottrell

Stanford Linear Accelerator Center
Mail Stop 97, P.O. Box 4349
Stanford, California 94309
Telephone: (650) 926 2523 Fax: (650) 926 3329
E-Mail: cottrell@stanford.edu

EMPLOYMENT SUMMARY

Period	Employer	Job Title	Activities
1982 on	Stanford Linear Accelerator Center	Assistant Director, Compu Services	Management of networking and computing
1980-82	Stanford Linear Accelerator Center	Manager SLAC Computer Network	Management of all SLAC's computing acti
1979-80	IBM U.K. Laboratories, Hursley, England	Visiting Scientist	Graphics and intelligent distributed worksta
1967-79	Stanford Linear Accelerator Center	Staff Physicist	Inelastic e-p scattering experiments, physics computing
1972-73	CERN	Visiting Scientist	Split Field Magnet experiment

EDUCATION SUMMARY

Period	Institution	Examinations	
1962-67	Manchester University	Ph.D.	Interactions of Deuterons with Carbon Isotopes
1959-62	Manchester University	B.Sc.	Physics

NARRATIVE

I joined SLAC as a research physicist in High Energy Physics, focusing on real-time data acquisition and analysis in the Nobel prize winning group that discovered the quark. In 1973/3, I spent a year's leave of absence as a visiting scientist at CERN in Geneva, Switzerland, and in 1979/80 at the IBM U.K. Laboratories at Hursley, England, where I obtained United States Patent 4,688,181 for a dynamic graphical cursor. I am currently the Assistant Director of the SLAC Computing Services group and lead the computer networking and telecommunications areas. I am also a member of the Energy Sciences Network Site Coordinating Committee (ESCC) and the chairman of the ESnet Network Monitoring Task Force. I was a leader of the effort that, in 1994, resulted in the first Internet connection to mainland China. I am also the leader/PI of the DoE sponsored Internet End-to-end Performance Monitoring (IEPM) effort, and the ICFA network monitoring working group.

PUBLICATIONS

The full list of 70 publications is readily available from online databases. I include here only a limited number of recent publications relevant to networking.

DEVELOPING COUNTRIES AND THE GLOBAL SCIENCE WEB, H. Cerdeira, E. Canessa, C. Fonda, R. L. Cottrell, CERN Courier December 2003.

OPTIMIZING 10-Gigabit ETHERNET FOR NETWORKS OF WORKSTATIONS, CLUSTER & GRIDS: A CASE STUDY, Wu-chun Feng, Justin Hurwitz, Harvey Newman, Sylvain Ravot, R. Les Cottrell, Olivier Martin, Fabrizio

Coccetti, Cheng Jin, Xiaoliang Wei, Steven Low, SC'03, Phoenix Arizona, November 15-21, 2003, also SLAC-PUB-10198.

MEASURING THE DIGITAL DIVIDE WITH PINGER, R. Les Cottrell and Warren Matthews, Developing Countries Access to Scientific Knowledge: Quantifying the Digital Divide, ICTP Trieste, October 2003; also SLAC-PUB-10186.

INTERNET PERFORMANCE TO AFRICA, R. Les Cottrell and Enrique Canessa, Developing Countries Access to Scientific Knowledge: Quantifying the Digital Divide, ICTP Trieste, October 2003; also SLAC-PUB-10188.

ABWE: A PRACTICAL APPROACH TO AVAILABLE BANDWIDTH ESTIMATION, Jiri Navratil, Les Cottrell, SLAC-PUB-9622, published at PAM 2003.

MEASURING END-TO-END BANDWIDTH WITH IPERF & WEB100, Ajay Tirumala, Les Cottrell, Tom Dunigan, SLAC-PUB-9733, published at PAM2003, April 2003.

EXPERIENCES AND RESULTS FROM A NEW HIGH PERFORMANCE NETWORK AND APPLICATION MONITORING TOOLKIT, Les Cottrell, Connie Logg, I-Heng Mei, SLAC-PUB-9641, published at PAM 2003, April 2003.

IGRID2002 DEMONSTRATION BANDWIDTH FROM THE LOW LANDS, R. Les Cottrell, Antony Antony, Connie Logg and Jiri Navratil, in Future Generation Computer Systems 19 (2003) 825-837, published by Elsevier Science B. V.; also SLAC-PUB-9560, October 31, 2002

NETWORK SCAVENGERS, By Warren Matthews, Les Cottrell and Paola Grosso, InterAct, Vol 2, Spring 2002.

PEER TO PEER COMPUTING FOR SECURE HIGH PERFORMANCE DATA COPYING.
By Andrew Hanushevsky, Artem Trunov, Les Cottrell (SLAC). SLAC-PUB-9173, Sep 2001. 4pp.
Presented at CHEP'01: Computing in High-Energy Physics and Nuclear, Beijing, China, 3-7 Sep 2001.

PASSIVE PERFORMANCE MONITORING AND TRAFFIC CHARACTERISTICS ON THE SLAC INTERNET BORDER.
By Connie Logg, Les Cottrell (SLAC). SLAC-PUB-9174, Sep 2001. 4pp.
To appear in the proceedings of CHEP'01: Computing in, High-Energy Physics and Nuclear, Beijing, China, 3-7 Sep 2001.

THE PINGER PROJECT: ACTIVE INTERNET PERFORMANCE MONITORING FOR THE HENP COMMUNITY.
By W. Matthews, L. Cottrell (SLAC). SLAC-REPRINT-2000-008, May 2000. 7pp.
Published in IEEE Commun.Mag.38:130-136, 2000.

1-800-CALL-H.E.P.: EXPERIENCES ON A VOICE OVER IP TEST BED. By W. Matthews, L. Cottrell (SLAC), R. Nitzan (Energy Sciences Network). SLAC-PUB-8384, Feb 2000. 5pp. Presented at International Conference on Computing in High Energy Physics and Nuclear Physics (CHEP 2000), Padova, Italy, 7-11 Feb 2000.

LECTURE COURSES

HOW THE INTERNET WORKS: International Nathiagali Summer College Lecture course, given by Les Cottrell in Pakistan, Summer 2001.

Jiri Navratil

Stanford Linear Accelerator Center
Mail Stop 97, P.O. Box 4349
Stanford, California 94309
Telephone: (650) 926 3332 Fax: (650) 926 3329
E-Mail: jiri@stanford.edu

Narrative

Jiri Navratil is a Staff Scientist in SLAC's networking group. He joined SLAC in 2002 as a networking specialist for collaboration on the DOE Project INCITE. He has many years of practice in the development, testing and tuning of several different packages analyzing scientific data. (In 70s in FORTRAN). Later, he specialized for interactive applications. He has a very good level of competency in C. His latest programming experiences are linked with Perl and networking software. He worked with many type of Unix since 1985 (HPUX on HP9000 workstations, AIX on IBM RS6000 and SP2, Solaris on Sun and currently Linux. He is using PC (MS Windows/XP) for his personal and administrative work.

He has had many experiences with leading software teams. In 1976 he began to lead a small system software group. Later his group has been responsible for technical and software support on university level. In 80ties he was a leader of several resort-wide projects, The management system for Ministry of Education (1985-1989) and Academic Initiative of IBM (1990-1994). In the period of 1986-1994 he was a member of team that created CESNET (Czech Scientific network) that connect all universities in the country and later I was participating on the similar project TEN-34 CZ. In 1994 -1999 (before his move to CERN) he was responsible for scientific computing and networking in Computing and Information Center at Czech Technical University.

During his work abroad (CERN, KEK) he participated on development software tools for large project as LEP (Large Electron-Positron Accelerator at CERN), rejuvenation of TRISTAN accelerator and design of B-factory Control system at KEK. During his stay at CERN since June 1998 he worked at IT Division as network specialist with responsibility for daily monitoring of CERN international connections, making analysis and web presentation as well as a creation of new tools for network monitoring. His current SLAC activity is linked with the project INCITE. In frame of this he collaborates with Rice University team on development of new tools for end2end monitoring.

Education

Dipl. Engineer (Czech Technical University at Prague) 1972
PhD (Czech Technical University at Prague) 1982

Positions

Application programmer (University Regional Computing Center CTU Prague)	1972 - 1976
System Administrator (University Regional Computing Center CTU Prague)	1976 - 1980
Head of Computing Department (Computing Center CTU Prague)	1980 - 1985
Head of Information Department (Computing Center CTU Prague)	1986 - 1989
Scientific Associate at CERN (PS,SPS/LEP)	1985, 1989 - 1990
Guest Professor at KEK (ACC.Div)	1992, 1994
Head of Computing and Networking Services (CTU Prague)	1995 – 1998
Scientific Associate at CERN (IT-CS)	1998 – 2000
Head of Computing and Networking Services (CTU Prague)	2000 – 2001
Visiting scientist SLAC	2002 – current

Selected Publications

- Connie Logg, Jiri Navratil, Les Cottrell: To be submitted in February 2004 to PAM2004: CORRELATING INTERNET PERFORMANCE CHANGES AND ROUTE CHANGES TO ASSIST IN TROUBLE-SHOOTING FROM END-USER-PROSPECTIVE

- Jiri Navratil, Les Cottrell: What we have learned from developing and running AbwE. BEst-Bandwidth Estimation workshop, CAIDA December 2003, http://www.caida.org/outreach/isma/0312/slides/jnavratil.pdf
- Jiri Navratil, Les Cottrell, SLAC-PUB-9622, published ABWE: A PRACTICAL APPROACH TO AVAILABLE BANDWIDTH ESTIMATION, Jiri at PAM 2003.
- PATHCHIRP: EFFICIENT AVAILABLE BANDWIDTH ESTIMATION FOR NETWORK PATHS, Vinay Ribeiro, Rudolf Reidi, Richard Baraniuk, Jiri Navratil, Les Cottrell, SLAC-PUB-9732, published at PAM 2003, April 2003.
- IGRID2002 DEMONSTRATION BANDWIDTH FROM THE LOW LANDS, R. Les Cottrell, Antony Antony, Connie Logg and Jiri Navratil, in Future Generation Computer Systems 19 (2003) 825-837, published by Elsevier Science B. V.; also SLAC-PUB-9560, October 31, 2002
- J.Navrátil, Z.Bittnar, M.Císlerová, V.Vacek, K.Kozel, J.Macháč: High Performance Computing at Czech Technical University, WORKSHOP 2001,CTU Prague, February 2001
- J.Navratil and col. Review of results and the set of selected publications achieved on IBM RS/6000 SP, ISBN 80-01-02334-6, CTU Prague, January 2001, 216 pages
- J.Navratil, P.Bures,J.Krupova: Upgrade and Extension of IBM Supercomputer RS/6000 SP, WORKSHOP 2001, CTU Prague, February 2001
- J.Navratil: The role of computers IBM RS/6000 SP in research and science at the universities (published in Czech), IBM CR, 1998,The BlueRose, zari 98,pp 16-17
- J.Navrátil, O.Plachý, P.Kolman, P.Bureš: The first experiences with IBM SP2, SUPEUR'96, International Conference, Krakow, September 8-11, 1996
- J.Navratil: The experience with parallel programming, Proceedings of the 2nd Japan-Central Europe Joint Workshop on Modeling Materials and Combustion, November 7-9, 1996, Budapest,
- J.Navratil: The Tcl/Tk A Programming System, X11 User Interface for EPICS Users, textbook of lectures and examples, KEK Tsukuba, February 1995, 150 pages
- J.Navratil: How I see the Epics and the Vsystem, The internal report of Accelerator Division, KEK Tsukuba, March 1995
- J.Navratil, P.Bures, T.Mimashi, V.Novak, J.Safar: CIV Center for High Intensive Computing, Proceedings of the Japan-Central Europe Joint Workshop on Advanced Computing in Engineering., Warsaw, 1994
- J.Navratil, A.Akiyama, H.Fukuma, S.Kamada, S.Kurukawa, T.Mimashi, K.Oide, T.Shintake, J.Urakawa, N.Yamamoto: System for Monitoring Betatron Tune, Proceedings of the XVth International Conference on High Energy Accelerators, HEACC'92, Hamburg, Germany, July 20-24,1992, Int. J. Mod. Physics A (Proc. Suppl) 2A (1993),pp 290-292
- J.Navratil: How to program XWindows, Textbook of lectures and examples, KEK Tsukuba, September 1992, 101 pages
- J.Navratil, A.Akiyama, T. Mimashi: Visual observation of digitalised signals by workstations, Nuclear Instruments and Methods in Physics Research, Section A, pp. 361-365, Proceedings of the International Conference ICALEPCS'93, Berlin, Germany, October 18-23,1993

K. References

[ABING] J. Navratil and L. Cottrell, "ABWE: A Practical Approach to Available Bandwidth Estimation," SLAC-*PUB-9622,* Passive and Active Measurement Workshop, 2003.

[Abry02] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, "Multiscale Nature of Network Traffic," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 28-46, May 2002.

[Ada00] A. Adams, T. Bu, R. Caceres, N. Duffield, T. Friedman, J. Horowitz, F. Lo Presti, S.B. Moon, V. Paxson, D. Towsley, "The use of end-to-end multicast measurements for characterizing Internet network behaviour," *IEEE Communications Magazine*, May 2000.

[B99] R. G. Baraniuk, "Optimal tree approximation using wavelets," Proceedings of SPIE Technical Conference on Wavelet Applications in Signal Processing VII, Denver, July 1999.

[BaBar] Available at http://www-public.slac.stanford.edu/babar/

[bbftp] Available at http://doc.in2p3.fr/bbftp/

[bbcp] Available at http://www.slac.stanford.edu/~abh/bbcp/

[Cac99] R. Caceres, N. Duffield, J. Horowitz, D. Towsley, "Multicast-based Inference of Network-Internal loss characteristics," *IEEE Trans. Info. Theory*, November 1999.

[Car97] R. Carter and M. Crovella, "Server selection using dynamic path characterization in wide-area networks," *Joint Conference of the IEEE Computer and Communications Societies (IEEE Infocom'97)*, April 1997.

[CB98] M. S. Crouse and R. G. Baraniuk, "Contextual hidden Markov models for wavelet-domain signal processing," Proceedings *of 31st Asilomar Conference,* Nov. 1997.

[CB99] H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," submitted to *IEEE Trans. on Image Processing*, October 1999.

[CB99a] H. Choi and R. G. Baraniuk, "Wavelet statistical models and Besov spaces," *Proceedings of SPIE Technical Conference on Wavelet Applications in Signal Processing VII,* Denver, July 1999.

[Chny02] M. Coates, A. Hero, R. Nowak, B. Yu, "Internet Tomography," *IEEE Signal Processing Magazine*, May 2002.

[CiscoWorks] http://www.cisco.com/en/US/products/sw/cscowork/ps2426/index.html

[CNB98] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Trans. Signal Processing* (Special Issue on Wavelets and Filterbanks), vol. 46, pp. 886-902, April 1998.

[Clark87] D. Clark, M. Lambert, and L. Zhang, "NETBLT: A High Throughput Transport Protocol," *Proceedings of ACM Sigcomm '87,* August 1987.

[DataTAG] DataTAG on-line pages and documents http://datatag.web.cern.ch/datatag/

[EDG] http://edg-wp2.web.cern.ch/edg-wp2/

[EGEE] http://egee-intranet.web.cern.ch/egee-intranet/gateway.html

[Fast04] C. Jin, D. X. Wei and S. H. Lo, "FAST TCP: motivation, architecture, algorithms, performance," IEEE Infocom, March 2004.

[Fei98] Z. Fei and S. Bhattacharjee and E. Zegura and M. Ammar, "A novel server selection technique for improving the response time of a replicated service," *Joint Conference of the IEEE Computer and Communications Societies (IEEE Infocom'98),* April 1998.

[Feng00] W. Feng and P. Tinnakornsrisuphap, "The Failure of TCP in High-Performance Computational Grids," *IEEE/ACM SC2000*, November 2000.

[Feng02] W. Feng, A. Kapadia, and S. Thulasidasan, "GREEN: Proactive Queue Management over a Best-Effort Network," *IEEE GLOBECOM 2002*, November 2002.

[Feng03] W. Feng and S. Vanichpun, "Ensuring Compatibility Between TCP Reno and TCP Vegas," *IEEE Symposium on Applications and the Internet (SAINT'03)*, January 2003.

[Feng03b] W. Feng, G. Hurwitz, H. Newman, S. Ravot, R. L. Cottrell, O. Martin, F. Coccetti, C. Jin, D. Wei, and S. Low, "Optimizing 10-Gigabit Ethernet for Networks of Workstations, Clusters, and Grids: A Case Study," *SC2003*, November 2003.

[Feng03c] W. Feng, M. Gardner, M. Fisk, and E. Weigle, "Automatic Flow-Control Adaptation for Enhancing Network Performance in Computational Grids," *Journal of Grid Gomputing*, Vol. 1, No. 1, 2003, pp.63-74.

[Fisk00] M. Fisk and W. Feng. "Dynamic Adjustment of TCP Window Size," *Los Alamos Unclassified Report (LAUR) 00-3221*, July 2000.

[Fisk01] M. Fisk and W. Feng. "Dynamic Right-Sizing: TCP Flow-Control Adaptation," *Proceedings of SC 2001: High-Performance Networking and Computing* Conference, November 2001.

[Fisk01] M. Fisk and W. Feng, "Dynamic Right-Sizing in TCP," *Los Alamos Computer Science Institute Symposium*, October 2001.

[Flo03] S. Floyd, "Highspeed TCP for Large Congestion Windows," Internet Draft *draft-ietf-tsvwg-highspeed-01.txt*, August 2003.

[Floyd00] S. Floyd and M. Handley, J. Padhye, and J. Widmer, "Equation-Based Congestion Control for Unicast Applications," *Proceedings of ACM Sigcomm '00*, August 2000.

[Floyd03] S. Floyd. "High Speed TCP for Large Congestion Windows," *RFC 3649*, December 2003.

[Foste99] I. Foster and C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan-Kaufmann Publishing, 1999.

[Gardn02] M. Gardner, W. Feng, and M. Fisk, "Dynamic Right-Sizing in FTP (drsFTP): An Automatic Technique for Enhancing Grid Performance," *IEEE Symposium on High-Performance Distributed Computing (HPDC'02)*, July 2002.

[Gardn02] M. Gardner and W. Feng and M. Fisk. "Dynamic Right-Sizing in FTP: Enhancing Grid Performance in User Space," *Proceedings of the IEEE Symposium on High-Performance Distributed Computing*, July 2002.

[Gardn03a] M. Gardner, M. Broxton, A. Engelhart, and W. Feng, "MUSE: A Software Oscilloscope for Clusters and Grids," *IEEE International Parallel & Distributed Processing Symposium*, April 2003.

[Gardn03b] M. Gardner, W. Feng, M. Broxton, G. Hurwitz, and A. Engelhart, "Online Monitoring of Computing Systems with MAGNET," *IEEE/ACM Symposium on Cluster Computing and the Grid (CCGrid'03)*, May 2003.

[Gardn03c] M. Gardner, S. Thulasidasan, and W. Feng, "User-Space Auto-Tuning for TCP Flow Control in Computational Grids," accepted for publication in *Computer Communications*, 2003.

[Grid2003] http://www.ivdgl.org/grid2003/

[Gupta00] R. Gupta, M. Chen, S. McCanne, and J. Walrand, "A Receiver-Driven Transport Protocol for the Web," *Proceedings of INFORMS '00,* November, 2000.

[Guy95] J. Guyton and M. Schwartz, ""Locating Nearby Copies of Replicated Internet Servers," *Conference of the Special Interest Group on Data Communications (ACM Sigcomm '95)*, August 1995.

[gridFTP] http://datatag.web.cern.ch/datatag/WP3/grid_app_mon/gridftp.htm

[HP OpenView] http://www.openview.hp.com/

[Hur03] G. Hurwitz and W. Feng, "Initial Performance Evaluation of 10-Gigabit Ethernet," *IEEE Hot Interconnects: A Symposium on High-Performance Interconnects*, August 2003.

[Hur04] G. Hurwitz and W. Feng, "Performance Evaluation and Implications of 10-Gigabit Ethernet," *IEEE Micro*, January-February 2004.

[Hsieh03] H.-Y. Hsieh, K.-H. Kim, Y. Zhu, and R. Sivakumar, "A Receiver-Centric Transport Protocol for Mobile Hosts with Heterogeneous Wireless Interfaces," *Proceedings of ACM Mobicom '03*, September 2003.

[IEPM-BW] L. Cottrell and C. Logg, "Overview of IEPM-BW Bandwidth Testing of Bulk Transfer Data," *SLAC-PUB-9202*, July 2003. http://www.planet-lab.org/consortium/Governance_1203.pdf

[Iperf] http://dast.nlanr.net/Projects/Iperf/

[IPv6] W. Matthews, *IPv6 Performance and Reliability*, Presented at *IPv6-2000*, Washington DC, October 2000.

[Jai02] M. Jain, C. Dovrolis, "End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relationship with TCP Throughput," *ACM SIGCOMM* 2002, Pittsburgh, August 2002.

[Kar97] D. Karger and E. Lehman and T. Leighton and M. Levine and D. Lewin and R. Panigrahy, "Consistent Hashing and Random Trees: Distributed Caching Protocols for Relieving Hot Spots on the World Wide Web," *Symposium on Theory of Computing*, May 1997.

[Kuz03a] A. Kuzmanovic and E. Knightly, "TCP-LP: A Distributed Algorithm for Low Priority Data Transfer," *Joint Conference of the IEEE Computer and Communications Societies (IEEE Infocom'03)*, April 2003.

[Kuz03b] A. Kuzmanovic, E. Knightly, and R. L. Cottrell, "HSTCP-LP: A Protocol for Low-Priority Bulk Data Transfer in High-Speed High-RTT Networks," *Second International Workshop on Protocols for Fast Long-Distance Networks (PFLDnet'04)*, February 2004.

[Kuz03c] A. Kuzmanovic and E. Knightly, "Low-Rate TCP-Targeted Denial of Service Attacks (The Shrew vs. the Mice and Elephants)," *Conference of the Special Interest Group on Data Communications (ACM Sigcomm '03)*, August 2003.

[Kuz04] A. Kuzmanovic and E. W. Knightly and R. Les Cottrell. "Bulk Data Transfer in High-Speed High-{RTT} Networks," *Proceedings of the 2nd International Workshop on Protocols for Fast Long-Distance Networks*, February 2004.

[Lambda] http://www.nwfusion.com/news/tech/2001/0108tech.html

[Liu00] J. Liu and J. Ferguson. "Automatic {TCP} Socket Buffer Tuning," *Proceedings of SC 2000: High-Performance Networking and Computing Conference (Research Gem)*, November 2000. http://dast.nlanr.net/Projects/Autobuf.

[Mapcenter] http://mapcenter.in2p3.fr/user-guide.html

[Mathi99] M. Mathis. "Pushing Up Performance for Everyone," http://www.ncne.nlanr.net/news/workshop/19999/991205/Talks/ mathis_991205_Pushing_Up_Performance/

[Mehra03] P. Mehra, A. Zakhor, and C. De Vleeschouwer, "Receiver-Driven Bandwidth Sharing for TCP," *Proceedings of Infocom '03*, April 2003.

[monALISA] http://monalisa.cacr.caltech.edu/

[MRTG] http://people.ee.ethz.ch/~oetiker/webtools/mrtg/

[N99] R. Nowak, "Multiscale hidden markov models for bayesian image analysis," *Bayesian Inference in Wavelet Based Models*, pp. 243-266, *Springer-Verlag,* 1999.

[Netherlight] http://www.surfnet.nl/innovatie/netherlight/

[Ng03] E. Ng and Y. Chu and S. Rao and K. Sripanidkulchai and H. Zhang, "Enhancing Measurement-Based Optimization Techniques for Bandwidth-Demanding Peer-to-Peer Systems," *Joint Conference of the IEEE Computer and Communications Societies (IEEE Infocom'03)*, April 2003.

[NWS] Network Weather Service. http://nws.cs.ucsb.edu/.

[Pad03] V. N. Padmanabhan, L. Qiu, H. Wang, "Server-based Inference of Internet Link Lossiness", *Proceedings of IEEE Infocom*, San Francisco, CA, April 2003.

[PINGER] PINGER HISTORY & METHODOLOGY, R. Les Cottrell and Connie Logg, Developing Countries Access to Scientific Knowledge: Quantifying the Digital Divide, *ICTP Trieste*, October 2003; also SLAC-PUB-10187

[PLANET] Planet-lab on-line and Planet-lab documents: http://www.planet-lab.org/

[PSC] Pittsburgh Supercomputing Center. "Enabling High-Performance Data Transfers on Hosts." http://www.psc.edu/networking-/perf_tune.html.

[RBUDP] http://dsd.lbl.gov/DIDC/PFLDnet2004/papers/Wu.pdf

[Rib03a] V. Riberio, R. Riedi, R. G. Baraniuk, J. Navratil, L. Cottrell, "pathChirp: Efficient Available Bandwidth Estimation for Network Paths," *Passive Active Measurement Workshop -- PAM2003*, Jolla, CA, April 2003 (winner of best student paper award).

[Rib03b] V. Ribeiro, R. Riedi, and R. G. Baraniuk, "Optimal Sampling Strategies for Multiscale Models with an Application to Network Traffic Measurement," *IEEE Statistical Signal Processing Workshop*, St. Louis, September 2003.

[Rib03c] V. Ribeiro, R. Riedi, and R. G. Baraniuk, " Spatio-Temporal Available Bandwidth Estimation for High-Speed Networks," *ISMA 2003 Bandwidth Estimation Workshop (Best)*, CAIDA, La Jolla, CA, December 2003.

[Rib04] V. Riberio, R. Riedi, R. G. Baraniuk, "Spatio-Temporal Available Bandwidth Estimation with pathchirp," *ACM SIGMETRICS* 2004.

[Rib04a] V. Ribeiro, R. H. Riedi, and R. G. Baraniuk, "Multiscale Queuing Analysis of Long-Range-Dependent Network Traffic," submitted to *IEEE Transactions on Networks*, 2004.

[Riedi99] R. Riedi, M. Crouse, V. Ribeiro, and R. Baraniuk, "A Multiplicative Wavelet Model with Application to TCP Network Traffic," *IEEE Transactions on Information Theory*, April 1999.

[Rao03a] N. Rao, "Network-Intensive Science: Data Transfers, Visualization, and Steering," *DOE Science UltraScience Net Kickoff Meeting*, November 2003. www.csm.ornl.gov/UltraScience Net

[Rao03a] N. Rao, B. Wing, "UltraScience Net: Usage Modes and User Interfaces," *DOE Science UltraScience Net Kickoff Meeting*, November 2003. www.csm.ornl.gov/UltraScience Net

[RCRB99] R. Riedi, M. S. Crouse, V. Ribeiro and R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Info. Theory*, vol. 45, 992-1018, 1999.

[RFC2488] M. Allman and D. Glover and L. Sanchez. "Enhancing TCP Over Satellite Channels Using Standard Mechanisms," *IETF RFC 2488*, 1999.

[Rod00] P. Rodriguez and A. Kirpal and E. Biersack, "Parallel-Access for Mirror Sites in the Internet," Joint Conference of the IEEE Computer and Communications Societies (*IEEE Infocom'00)*, March 2000.

[Rod02] P. Rodriguez and E. Biersack, "Dynamic parallel-access to replicated content in the Internet," ACM/IEEE Transactions on Networking, August 2002.

[Sar01] S. Sarvotham, R. Riedi, and R. G. Baraniuk "Connection-level Analysis and Modeling of Network Traffic," *ACM SIGCOMM Internet Measurement Workshop*, San Francisco, November 2001.

[Sar02] S. Sarvotham, R. H. Riedi, and R. G. Baraniuk, "Connection-Level Modeling of Network Traffic," *36th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November, 2002.

[Sar04] S. Sarvotham, R. Riedi, and R. G. Baraniuk "Connection-level network traffic modeling: From network topology to traffic dynamics," *Computer Networks Journal*, 2004 (invited).

[Semke98] J. Semke and J. Mahdavi and M. Mathis. "Automatic TCP Buffer Tuning", *Computer Communications Review, ACM SIGCOMM*, vol. 28, no. 4, October

1998.

[Sha01] A. Shaikh and R. Tewari and M. Agrawal, "On the effectiveness of DNS-based server selection," *Joint Conference of the IEEE Computer and Communications Societies (IEEE Infocom'01)*, April 2001.

[Sinha99] P. Sinha, N. Venkitaraman, R. Sivakumar, and V. Bharghavan, "WTCP: A Reliable Transport Protocol for Wireless Wide-Area Networks," *Proceedings of ACM Mobicom '99*, August 1999.

[SNMP] http://www.net-snmp.org/tutorial/

[Sou] E. Souza and D. Agarwal, "A HighSpeed TCP Study: Characteristics and Deployment Issues," LBNL Technical Report LBNL-53215. http://www-itg.lbl.gov/~evandro/hstcp/hstcp-lbnl-53215.pdf

[Spring00] N. Spring and M. Chesire and M. Berryman and V. Sahasranaman and T. Anderson and B. Bershad," Receiver Based Management of Low Bandwidth Access Links," *Proceedings of Infocom '00*, March 2000.

[Steve97] R. Stevens, P. Woodward, T. DeFanti, and C. Catlett. "From the I-WAY to the National Technology Grid," *Communications of the ACM,* 40(11): pp. 55-60, November 1997.

[Sus03] Sustained Bandwidth Award ("Moore's Law Move Over!"), *SC2003 Bandwidth Challenge*, November 2003.

[Tie03] B. Tierney and G. Jin, "System capability effects on algorithms for network bandwidth estimation," *Internet Measurement Conference,* October 2003.

[Tiern01a] B. Tierney. "TCP Tuning Guide for Distributed Applications onWide-Area Networks," *USENIX & SAGE Login*, February, 2001. http://wwwdidc.lbl.gov/tcp-wan.html.

[Tiern01b] B. Tierney and D. Gunter and J. Lee and M. Stoufer. "Enabling Network-Aware Applications", *Proceedings of the IEEE International Symposium on High-Performance Distributed Computing*, August 2001.

[Thula03] S. Thulasidasan, W. Feng, and M. Gardner, "Optimizing GridFTP Through Dynamic Right-Sizing," *IEEE Symposium on High-Performance Distributed Computing (HPDC'03)*, June 2003.

[TN99] K. Timmermann and R. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging," *IEEE Trans. Info. Theory,* vol. 45, pp. 846-862, 1999.

[Tsao02] V. Tsaoussidis and C. Zhang, "TCP-Real: Receiver-Oriented Congestion Control," *The Journal of Computer Networks*, 40(4):477-497, April 2002.

[Tsa03] Y. Tsang, M. Coates, R. Nowak, "Network Delay Tomography", *IEEE Transactions on Signal Processing*, vol. 51, No. 8, August, 2003, 2125-2136.

[Vanic02] S. Vanichpun and W. Feng, "On the Transient Behavior of TCP Vegas," *IEEE International Conference on Computer Communications and Networks (IC3N'02)*, October 2002.

[Web100] National Center for Atmospheric Research and Pittsburgh Supercomputing Center and National Center for Supercomputing Applications. "The Web100 Project." http://www.web100.org/.

[Weaver03] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, "Large Scale Malicious Code: A Research Agenda," DARPA-sponsored report, 2003.

[Weigle01] E. Weigle and W. Feng. "Dynamic Right-Sizing: A Simulation Study," *Proceedings of IEEE International Conference on Computer Communications and Networks*, 2001.

[Zou03] C. Zou, L. Gao, W. Gong, D. Towsley, "Monitoring and Early Warning for Internet Worms," *Technical Report TR-CSE-03-01*, Department of Computer Science, University of Massachusetts, March 2003.