# Applied network research: 1994 annual status report*

Hans-Werner Braun
Bilal Chinoy
Kimberly C. Claffy

Applied Network Research
San Diego Supercomputer Center
and
University of California, San Diego

April 19, 1995

**Abstract**

We present the 1994 annual report on research efforts of the San Diego Supercomputer Center's Applied Network Research group (ANR). Our projects included collaboration with other researchers across the country. This year, we continued our research on network analysis, modeling, engineering, as well as advanced network technologies including the CASA gigabit testbed. Our research topics include issues of network access, the transition of the Internet environment as the NSF dismantles the NSFNET this year; instrumentation for the accurate gathering of performance statistics, accounting, and quality of service in the Internet. In the coming year we hope to complement our studies of lower layer network issues with investigations into the critical architectural limits we are reaching at the higher layers of the national network infrastructure. We hope interested parties within the Internet research community will find this report helpful in understanding our research agenda. We also wish to use this report to stimulate further discussion and interaction among researchers pursuing similar projects and goals.

*Any opinions, conclusions, or recommendations in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation, other supporting organizations, General Atomics, SDSC, UCSD or the SDSC Consortium members.

# Contents

# 1 introduction

The Applied Network Research group agenda focuses on the evolution and future requirements of the data networking infrastructure. Research efforts include protocol performance evaluation for leading edge gigabit and ATM networks, traffic characterization for existing and future high speed networks, evolution and integration of higher level services and resources into the Internet fabric, and network engineering and planning support to the National Science Foundation.

In advanced network technologies, the ANR group participates in the CASA gigabit testbed project, specifically evaluating the performance of large-scale, distributed applications (including global climate change modeling) running across a heterogeneous supercomputer environment interconnected on a high-speed HIPPI based substrate.

In addition to investigation of experimental technologies infrastructure, ANR also participates in several collaborative research efforts with a perspective toward short to medium-term improvement of the existing operational infrastructure. In collaboration with other groups at SDSC, other NSF centers, and elsewhere, ANR pursues research in traffic and performance characterization to support the seamless integration of high end applications into the operational Internet. Because the nature of Internet workloads is changing so rapidly, ANR has undertaken studies on how operational Internet statistics collection should evolve with the changing workload in order to gain better insight into the increasingly dynamic nature of traffic. Also in support of the advanced functionality of the environment, ANR projects include issues in supporting servers for general data and information depositories, sophisticated user interfaces, collaborative and educational virtual environments, and high performance visualization of traffic characteristics.

ANR also participates in political developments in the Internet arena, working with NSF and other agencies in their role in advancing Internet evolution. The combination of the two perspectives, on advanced as well as existing network infrastructures, is in line with the ANR research agenda to develop synergy between political and technical objectives in Internet evolution.

This 1994 annual status report of the Applied Network Research (ANR) group at the San Diego Supercomputer Center (SDSC) and the University of California, San Diego (UCSD) represents an update on our research activities and outlines our future directions.

During 1994, the third year since the establishment of the Applied Network Research (ANR) Group, we have continued our work in the area of traffic characterization of strategic networking locations, in particular those which aggregate a large volume and range of traffic, typically NSFNET backbone nodes. An important component of our current research is the modeling of Internet traffic *flows*. In section 2, we discuss motivation for and potential applications of such an effort and describe recent findings. Topics of particular recent interest include traffic visualization, privacy concerns, metrics of burstiness, web traffic characterization, and caching and mirroring architectures for information resources.

We have and continue to apply the results of our investigations to operational and planned components of the Internet. Sections 4 discusses SDSC's involvement in the CASA gigabit testbed research project and in a multi-agency national ATM network project. Section 5 presents ANR efforts in NSF's NREN Engineering Group. In section 6 we discuss ANR's participation in the development, operation, and traffic characterization of network access points. In section 7 we discuss a method for dealing with privacy concerns on collected data.

In the coming year our agenda includes investigations of three Internet resource classes: information resources, communication resources, and computation resources. We discuss these areas in sections 8 through 10. In section 8 we discuss a proposed architecture for the establishment of a national hierarchical system for caching information resources. In section 9 we discuss our work on virtual collaborative educational projects, reflecting the second category, communication resources. In section 10 we discuss our planned projects in the third area, computation resources. Specifically we have proposed an infrastructure for supporting research on NSF's very high speed backbone network service (vBNS).

Section 11 describes the contributions made to the ANR research and analysis effort by three students from the Computer Science Department at the University of California, San Diego. Section 13 lists ANR publications released during the past year.

# 2 network analysis

The research efforts we describe in this section are all applicable not only to current network infrastructures, but also to planned and emerging advanced Internet technologies, around which we focus Section 4.

## 2.1 Internet traffic flow profiling

ANR has devoted much attention in the last two years to an investigation of the usefulness, relevance, and practicality of a wide variety of operationally collected statistics for wide area backbone networks. In particular, we have undertaken several studies on to what extent much of the statistics that the NSFNET project has collected over the life of the NSFNET backbone are useful for a variety of workload characterization efforts. We have also undertaken several studies which collect more comprehensive Internet traffic flow statistics and developed a methodology for describing those flows in terms of their impact on an aggregate Internet workload. We have developed a methodology for profiling Internet traffic flows which draws on previous flow models [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. The model of flows we are using depends on a flow specification (e.g., host pairs) which we then apply to packets traversing a specific network location. We create flows as packets between two entities appear, and time them out after periods of inactivity.

Our methodology for modeling flows differs from many previous studies that have concentrated on end-point definitions of flows, by mainly focusing on TCP flows delimited by SYN and FIN packets. Instead, we focus on the IP layer and define flows based on traffic satisfying various temporal and spatial locality conditions, as observed at internal points of the network. This approach to the definition and characterization of network flows has significant applications to various central problems for networking based on the Internet model. Among them, optimization of caches for routing, feasibility studies and optimization of routing based on quality-of-service considerations, usage based accounting, and optimization of the transport of IP traffic over an ATM fabric. In this section we concentrate on metrics characterizing individual flows including, volume in packets and bytes per flow, and flow duration, at various granularities of the definition of a flow, such as by destination network, host-pair, or host and port quadruple. Our measurements demonstrate

- the brevity of a significant segment of IP flows
- that the number of host-pair IP flows is not significantly higher than destination network flows, and
- that schemes that rely on traffic caching would significantly benefit if they make caching decisions taking into account higher layer information.

Our measurements have implications for: performance requirements of routers at Internet "hotspots"; general and specialized flow-based routing algorithms; future usage-based accounting requirements; and traffic prioritization. We have presented the results of our work in several forums, including to industry, NSF, and journal publications.

### 2.2 network workload visualization

The most interesting flow metrics are also the most difficult to visualize; we have explored methods for visualizing network phenomena and characteristics. We have previously focused on the geographic representation of traffic flow between countries, animating country-by-country traffic matrices to produce a visual time series.

This year we have made considerable progress with software for visualizing IP flows, using a variety of tools, including xgraph, xmgr, splus, data explorer, gnuplot, and customized programs that we developed using the SGI GL library. Progress with commercially available visualization systems was slow, as they are written for visualization experts to optimize flexibility. We eventually decided to design our own software for traffic flow visualization using the GL library that is available on our RS6000 as well as SGI workstations. Jarom Smith, an SDSC REU, took an NSFNET data visualization project he completed using GL and installed it on a Cerfnet web server. We would like to expand the portability of these tools, and compare their utility and functionality to other tools in the community, for example Netramet from the University of Auckland, New Zealand [12].

### 2.3 metrics for burstiness

As the NSFNET service provision agreement ceases and Internet connectivity becomes an industrial commodity, metrics for describing the quality of connectivity will be important to market efficiency. One of our goals is to determine how to describe Internet workload using metrics that will enable customers and service providers to agree on a definition of a given grade of service. One common notion of network workload is 'burstiness', but there is not yet agreement in the Internet community on the best metric to define burstiness. Network behavioral patterns of burstiness are important for defining service specifications. As such, finding metrics to specify expectations and define observations are critical for the evaluation of services, and the need for them intensifies as services in today's internetworking environments become more commercialized, rather than procured via collaborative undertakings between the federal government and academia and industry.

Several researchers [13, 14, 15] have explored the failure of Poisson models to adequately characterize both local and wide area Internet traffic. Based on multiple packet traces, we started some work on traffic self similarity considerations, including variance-time plots to compare them to those in the above studies. We have also explored alternative metrics, e.g, peak/average ratio plots, and are still investigating possible metrics that may serve NSF, or Internet service providers, most effectively.

This task relies critically on accurate packet arrival timestamps, and thus has required finding tools adequate for packet tracing at accurate (microsecond) time granularities. We were previously using SGI R4000s for packet collection; however when the traffic level grew beyond loads with which the SGI's could keep up, i.e., at NCSA or even at SDSC during current busy periods, we investigated possible alternative architectures. Jeff Mogul of DEC assisted us in porting our SGI code to the DEC Alpha workstations, which have performed considerably better. (In our measurements, our SGI platforms delivered FDDI packets to the application up to a rate of approximately 3500 packets per second, after which the CPU became the bottleneck for the application to get the data from the kernel.

From our initial measurements and discussions with Jeff Mogul, a higher end DEC Alpha platform will deliver at least 3 to 4 times as many packets, but we have not yet done experiments to conclusively verify the performance limits.)

# 3 web traffic characterization

Similar to our NSFNET analysis work, we have begun to explore the utility of operationally collected web statistics, generally in the form of *http* logs. We analyzed two days of queries to the popular mosaic server at NCSA to assess the geographic distribution of transaction requests. The wide geographic diversity of query sources and popularity of a relatively small portion of the web server file set present a strong case for deployment of geographically distributed caching mechanisms to improve server and network efficiency.

At the time of our measurements, the NCSA web server consisted of four servers in a cluster. We show time series of bandwidth and transaction demands for the server cluster and break these demands down into components according to geographical source of the query. We analyze the impact of caching the results of queries within the geographic zone from which the request was sourced, in terms of reduction of transactions with and bandwidth volume from the main server. We find that a cache document timeout even as low as 1024 seconds (about 17 minutes) during the two days that we analyzed would have saved between 40% and 70% of the bytes transferred from the central server. We investigate a range of timeouts for flushing documents from the cache, outlining the tradeoff between bandwidth savings and memory/cache management costs. We discuss the implications of this tradeoff in the face of possible future usage-based pricing of backbone services that may connect several cache sites.

We also discuss other issues that caching inevitably poses, such as how to redirect queries initially destined for a central server to a preferred cache site. The preference of a cache site may be a function of not only geographic proximity, but also current load on nearby servers or network links. Such refinements in the web architecture will be essential to the stability of the network as the web continues to grow, and operational geographic analysis of queries to archive and library servers will be fundamental to its effective evolution. An obvious area for future study is to apply the flow methodology described in section 2.1 above to Internet resource discovery services (irds) traffic such as the web.

## 3.1 evaluation of existing web tools for log analysis

We experimented with several *www/httpd* tools, including servers (NCSA, CERN, plexus), client browsing software (Mosaic, netscape [16], html/http utilities and text conversion tools (e.g., libwww [17] latex2html [18]). We also investigated the features and limitations of several web statistics log analysis packages: [19, 20, 21, 22] web navigating tools, e.g., harvest [23].

# 4 emerging network technologies

## 4.1 CASA testbed facilities

We continued our effort on the CASA testbed project in 1994. The long-awaited OC-48 link between SDSC and Los Alamos National Labs (LANL) was turned over to us by MCI and US-West in June 1994. The HIPPI/SONET (H/S) gateways, however, were limited at that time to supporting only 5 OC-3c channels. Our effective link bandwidth was thus 5 * 132 Mbps, or 660 Mbps (each OC-3c channel provides about 132 Mbps after SONET overhead is subtracted).

### 4.1.1 link layer performance

We continued several basic performance tests to calibrate the behavior of the CASA network and understand its performance characteristics. Using IOSC (Input Output Systems Corporation) HIPPI testers, we measured the available HIPPI-to-HIPPI bandwidth at approximately 132 Mbps over a single OC-3 fiber pair. In striped mode, we used a maximum of 6 OC-3 pairs and measured approximately 660 Mbps. Note that the H/S gateways optionally implement Forward Error Correction (FEC) and use one channel to transmit redundant data, reducing the number of available data channels to five.

### 4.1.2 HIPPI logical routing

A key issue in the scalability of HIPPI-SONET networks is the manageability of HIPPI routing. HIPPI routing can be done one of two ways: *source routing* or *logical* *routing*. We first used source routing because the size of the network, both in terms of end hosts as well as HIPPI switches, was small (under 10 hosts and 4 switches). As the network grew in size, source routing became inadequate and we turned to logical routing. This technique involves building switch routing tables and assigning HIPPI addresses (called I-fields) that are location independent.

Since there are no HIPPI routing protocols and no automatic means of configuring and updating host and switch tables, the configuration effort was substantial. We conclude that to build a large network from HIPPI bridges, dynamic routing update protocols as well as ARP servers to map network addresses to link-layer addresses are required.

#### 4.1.3 TCP congestion control and HIPPI flow control

We investigated the packet delay and loss characteristics of the wide-area HIPPI-based CASA gigabit testbed. *Blocking* may occur because HIPPI is a connection-oriented point-to-point protocol that does not allow simultaneous connections. Blocking implies that data may not be delivered to a destination because of a competing (unrelated) concurrent connection. We show that HIPPI blocking can degrade performance by increasing delay and/or packet loss. In the CASA network under conditions of blocking, a tradeoff exists between packet loss and delay variance. The tradeoff point is determined by a combination of factors: source packet rate, mean blocking rate, and a *configurable* connection establishment timeout threshold.

We demonstrate that the delay/loss tradeoff manifests itself in TCP either by triggering the *slow-start* algorithm or inducing TCP to adjust retransmission timeout values due to increased delay variance. In the first case, where blocking triggers a slow-start, we have measured TCP throughput to drop by as much as 97% of its unblocked maximum. The latter case, retransmission/timeout adjustment, results in a 75% drop. The delay/loss tradeoff allows one to match network characteristics with application/transport requirements. With respect to TCP, high packet loss induces the slow-start policy even when the network may not be congested, leading to low throughput. In addition, the variance-sensitive TCP round trip time (RTT) estimator is sufficiently robust to avoid spurious retransmissions in HIPPI-based networks such as CASA.

Thus, the delay/loss tradeoff for TCP implies that avoiding packet loss is more important than minimizing delay. Other transport protocols or loss-tolerant applications may prefer a less reliable channel in exchange for bounded delay variance.

A paper detailing our observations and conclusions appeared in "Usenix Symposium on High-Speed Networking, Aug 1-3, Oakland, CA. pp.45-59. [24]"

#### 4.1.4 TCP performance

We have a number of supercomputers connected to the CASA network. These include the Cray C90 and Intel Paragon at SDSC, the Intel Delta and Paragon at the California Institute of Technology (Caltech), and a Cray Y-MP at Jet Propulsion Laboratories (JPL). The Cray Y-MP and C90 are capable of driving a HIPPI channel at or close to full bandwidth, using full sized IP packets (64 KB) because these machines are limited by per-packet interrupt latencies.

We have reported previously on TCP performance tests performed on the SDSC-Caltech CASA link, between 2 Cray supercomputers. The delay-bandwidth product (DBP) of the SDSC-Caltech link is approximately 0.3 MB, or about four packets of 64 KB each in flight. The DBP product on the SDSC - LANL link is much higher, given the longer geographical distance and hence latency. We observe a 30 msec. RTT for 64 KB IP packets, giving us a DBP of approximately 2 MB. We use the TCP window scaling option for TCP, allowing the Crays to use the entire available TCP bandwidth if possible. We additionally had to increase the Cray advertised TCP window beyond 378 KB (which is the maximum available in default configuration) by increasing the amount of memory allocatable to socket mbufs.

With the following specifications,

| | |
|---|---|
| IP packet size | 64 KB |
| IP RTT | 30 ms |
| TCP window scale factor | 4 |
| TCP socket buffer, advertised TCP window | 2 MB |
| end hosts | Cray C90 (SDSC)<br>Cray Y-MP (T3D front end, LANL) |
| available CASA link bandwidth | 660 Mb/s |

We achieved 500 Mbps TCP sustained throughput (as reported by *TTCP* and *netperf*). Note that the test occurred in *production* mode on the supercomputers, i.e., CPU resources were timeshared amongst competing applications. Large supercomputers such as the SDSC Cray are timeshared machines, with a CPU scheduler controlling the sharing of CPU resources. The effect of scheduling on a network-intensive job such as a large bulk transfer can be critical; interrupting such jobs adds significant delay to packet transmissions and degrades performance.

### 4.2 interagency ATM connection

In 1992, SDSC received an NSF award "Interagency ATM Connection at the San Diego Supercomputer Center" to collaborate with DOE and NASA to explore the suitability of ATM technologies for building and operating interagency networks. Network implementation of the ATM DOE/ESnet testbed started in February 1995, and we are now revising the workscope to reflect the evolution of the NSFNET to its new architecture.

## 5 NSF NREN Engineering

During NSF's fiscal year 1994, SDSC continued to collaborate with NSF/DNCRI on NREN Engineering Group (NEG) activities. NSF's Division for Advanced Scientific Computing funded this as a supplement to the base funding of the San Diego Supercomputer Center. The overall NEG effort included collaborative work among NSF (particularly Steve Wolff and his staff), LANL (Peter Ford) and SDSC (Hans-Werner Braun and Kimberly C. Claffy).

A major objective of NSF's NREN Engineering Group is to facilitate a graceful evolution of the national networking agenda via NSF's involvement as a federal agency. Domestic networking requirements and architectural issues are prominently on the agenda, as are design criteria for international connections to the U.S. portion of the Internet. As such, the NEG has been participating in the conceptualization of network implementation plans, and network engineering studies for NSF/DNCRI. We have written a paper on provisioning the Internet with multiple service levels to address some difficulties under resource contention, co-authored with Steve Wolff of NSF, and UCSD economist Roger Bohn [25]. Other discussions with NSF and the Internet community surrounded issues of policy, such as network acceptable use policies.

In the future, as the NSFNET backbone service provisioning agreement ceases and Internet connectivity becomes an industrial commodity, metrics for describing the quality of connectivity will be important to market efficiency. One of our goals has been and will be to determine how to describe Internet workload using metrics that will enable customers and service providers to agree on a definition of a given grade of service. One common notion of network workload is 'burstiness', as we discuss in section 2.3, but there is not yet agreement in the Internet community on the best metric to define burstiness. Such notions will become critical to the commercial Internet marketplace as NSF gradually decreases funding to Internet service providers.

### 5.1 NSF very high speed backbone network service (vBNS)

In October 1994 MCI visited SDSC to discuss vBNS installation and research opportunities surrounding the vBNS project. Braun participated in discussions in June 1994 at Cornell University, as part of a Metacenter meeting, leading to a draft AUP submitted by the supercomputing centers to NSF. Wendy Huntoon of PSC coordinated, and the centers collectively submitted the following suggested draft of an AUP to NSF in July of 1994, and again in December, in order that DNCRI could ensure it was in place before Stephen Wolff left NSF:

> All vBNS traffic is subject to the NSFnet AUP dated June 1992. Furthermore, the vBNS is an experimental network resource dedicated to the development and execution of meritorious high bandwidth applications and services. As such, periods of reduced services or unavailability must be expected.
>
> Since the vBNS is experimental, all vBNS connected sites must maintain non-vBNS connectivity sufficient to support their average production traffic load.
>
> The NSF Metacenter is willing to collaborate with the NSF to develop methodologies and architectures in order to segregate traffic in support of this policy.

### 5.2 infrastructural data collection requirements

As the Internet develops support for a wider variety of applications and offered service qualities, tracking the composition of network traffic will be important to both meet current demand as well as evolve toward future applications. In particular, as many new applications consume considerably more bandwidth, and for longer continuous periods of time, than those with which the Internet has fundamentally grown to where it is today, it will be important to measure traffic in order to determine the extent and range of different user requirements.

We thus consider it critical to ensure that the follow-on NSFNET activities, including the NAPs, recognize the need for and in fact expand on the operational statistics collection paradigm during the lifetime of the NSFNET. The architecture will likely alter requirements for earlier NSFNET statistics collection techniques and requirements. In particular we think it will be important to foster a multi-agency effort to support the aggregation of network statistics data from multiple service providers.

In pursuit of this agenda we have proposed to deploy operational software for characterizing Internet flows at the four NSF-sponsored network access points (NAPs) which were planned to be operational as of 1 November 1994. The Sprint NAP was operational at that time; it was not clear as of February 95 when the others would be operational. We also suggested other strategic locations such as the FIXes and their successors, but required appropriate equipment that could keep up with the traffic rate in order to perform accurate workload characterization. NSF has specified that they want statistics collected at the NAPs, but without very clear outline of deliverables, and we are working with both NSF and the NAPs to clarify the objectives.

Claffy visited MCI in Reston, Virginia at the end of December 1994 to meet with the MCI engineers responsible for statistics collection and general operation of the vBNS (Rick Wilder, John Jamison, Joe Lawrence, and Dennis Ferguson). MCI will provide each vBNS site with a DEC Alpha 900 server to support vBNS statistics, and delivered one to SDSC in December 1994 for preliminary testing and development of statistics tools.

We discussed MCI's plans for collecting statistics for InternetMCI, which uses a model similar to that of the NSFNET. MCI intends to extend this same architecture to the vBNS to the extent possible, but it is not yet clear how to deal with statistics on the newer switches that are not so conducive to statistics collection. We have made available to MCI tools that we have developed and used for the last three years to process and analyze NSFNET data, including SNMP and ARTS, as well as tools to analyze dedicated packet traces to perform flow characterization.

We hope to continue to collaborate, and offer MCI and vBNS users counsel and support for traffic characterization on the vBNS service when it becomes operational, both for individual vBNS flows as well as the aggregate vBNS workload. Other vBNS sites also plan to contribute to the determination of what vBNS service providers should track regularly, and to coordinate placement of these statistics into a database to facilities studies similar to those of supercomputer usage, e.g., workload burstiness characterization (see sections 2.3 and 10). An important objective is to ensure that metrics and models derived from them are well documented so they can become applicable for the Internet community.

### 5.3 accounting investigations

Although in previous years we investigated telecommunications regulation issues in both the federal and state arenas, we have focused instead this year on research that might diminish the need for regulatory attention to the Internet industry. For example, we have proposed an IP traffic precedence policy that can enable a graceful transition into a more competitively supplied Internet market that might obviate the need for federal interest in regulation. Claffy and Braun outline this proposal with UCSD Professor Roger Bohn and Stephen Wolff of NSF, in "Mitigating the coming Internet crunch: multiple service levels via Precedence". The paper appeared in a 1994 special issue of the Journal of High Speed Networks and is also available in postscript and html from *www.sdsc.edu*.

The paper discusses how taking advantage of existing IP functionality to use multiple levels of service precedence can begin to address disparities between the requirements of conventional and more new and highly demanding applications.

We have had discussions with NSF and NASA representatives regarding accounting/pricing experiments on NASA Sciences Internet (NSI) and concerns of accounting for federal IP networks. We hosted a meeting in San Diego in August with Joe Bailey, of MIT and David Brown from NASA, UCSD professor Roger Bohn, Hans-Werner Braun, and Kimberly Claffy. We have also actively read and contributed to discussion on the *com-priv* mailing list with respect to accounting and pricing methodologies.

Our approach has assumed the belief that at least so far, the less regulation of the Internet, and in fact the more progress in removing regulatory barriers for existing telecommunications companies so they can participate more effectively in the Internet market, the better. However we recognize that regulatory issues may become quite prominent as the Internet and telecommunications industry increasingly overlap in functionality and market targets, and we will keep a close eye on the likely two-way interaction between regulatory trends and Internet marketization.

### 5.4 security

Claffy co-authored a paper with Reagan Moore, Jay Dombrowski, and Tom Hutton of SDSC, entitled "Security models in the Metacenter", sponsored in part by NSF Cooperative Agreement ASC-8902825. ANR has gotten more involved in security issues this year. Driven by the need to access some ANS machine(s) that converted to requiring kerberized access, we have started to use the Cygnus version of kerberos 4 on all ANR machines. We have also offered architectural support to SDSC's security agenda.

# 6 network access points

## 6.1 project description

An integral part of the new NSFnet architecture is the presence of *Network Access Points*, or *NAPs*. NAPs are networks envisioned to provide a neutral, AUP-free meeting point for Network Service Providers to exchange routing and traffic. SDSC staff has teamed up with Sprint and the Minnesota Supercomputer Center (MSC) to design, establish and maintain one such NAP. The Sprint NAP is located physically in Pennsauken, NJ., and its initial design is based on a FDDI ring and switch substrate.

## 6.2 work to date

Since the inception of this project in July 1994, SDSC and its partners have successfully designed and built the NAP in Pennsauken, N.J. Our NAP is completely operational and most of the major NSPs have co-located equipment at the NAP and are currently exchanging traffic.

Having a reliable and robust working NAP is critical to the US Internet transition from being built around the NSFnet backbone to a more commercial network of networks.

## 6.3 workload characterization

We have performed initial throughput estimates on NAP cross-traffic, based on division of current Internet traffic between prospective NSPs. We are beginning to undertake other workload characterization studies in preparation for transitioning to new architectures.

## 6.4 future work

We are in the process of moving to a FDDI switch based architecture, in anticipation of high traffic volumes. We are also starting the process of designing an ATM based NAP, to take advantage of the scaling inherent in the ATM technology.

# 7 privacy issues

Many of our investigations involve IP address information in headers, which are sensitive to privacy issues. Because part of our mission is also to encourage graduate student research that can further contribute to the research field, we developed a mechanism for masking IP addresses in operational and trace statistics collection so that we can make them available to students without concern for compromising privacy.

To secure IP address privacy we wrote a *perl* script that encodes the IP address in a trace file as increasing numbers starting from the beginning of the file, while retaining the net/host relationships (e.g., "net 3 host 4" corresponds to the third unique network seen, and the fourth unique host on that network). There is no mapping back to the real addresses, or even the classes. But the mapping does allow one to see multiple hosts that belong to the same network number, which will be useful for many network workload studies. A resulting trace segment looks as follows (the columns are: timestamp, packet interarrival time, packet size, src:dst masked address pair, protocol, source port, destination port, and TCP flags):

| time | $\Delta$t | sz | src:dst | pr | sprt | dprt | fl |
|---|---|---|---|---|---|---|---|
| 0.000044 | 44 | 112 | 20002:30001 | 17 | 4368 | 518 | 0 |
| 0.000470 | 426 | 552 | 10001:20001 | 6 | 1165 | 6000 | 10 |
| 0.000891 | 421 | 552 | 40001:50001 | 6 | 20 | 4827 | 18 |
| 0.000935 | 44 | 552 | 10001:20001 | 6 | 1165 | 6000 | 10 |
| 0.000970 | 35 | 40 | 60001:70001 | 6 | 1043 | 4242 | 10 |
| 0.001456 | 486 | 552 | 20003:80001 | 6 | 2254 | 119 | 10 |
| 0.001528 | 72 | 1236 | 90001:20004 | 17 | 7648 | 7648 | 0 |
| 0.001599 | 71 | 552 | 10001:20001 | 6 | 1165 | 6000 | 10 |

Having no table to map the IP addresses back (since they are just monotonically increasing from the start of the trace file as new net:hst pairs occur) should address the privacy concerns so that students can use traces for research. The script is available from `hwb@sdsc.edu`.

**The next three sections outline the projects we are planning for the next year. The ANR group is investigating how to make facilities, information, and opportunities for collaboration available on the network in an integrated fashion. In this context, facilities include supercomputers and workstations, information includes world wide web and gopher servers, and opportunities for collaboration include e-mail and multi-user domains (MUDs). One could think of these three categories as corresponding to three types of communication: machine-to-machine, people-to-machine, and people-to-people.**

**Within each category, multiple dimensions emerge:**

- **context: geographical, such as for local news, or topical, such as cognitive neuroscience research**
- **temporal: automated management of dynamically created communication resources used for brief periods, e.g, distributed classroom lecture or seminar**
- **geographic distribution: which may require transparency at times and boundary visibility at another**

**ANR is collaborating with others on the definition and development of service interfaces that will accommodate, with as much consistency as possible, all three dimensions.**

## 8 caching and mirroring architecture

Building on the study we described in section 3, we have emphasized the need for efficient access to information to influence the design of network infrastructure. In that study we conducted simulations to understand how intermediate caching of information at the supercomputer centers or NAPs might speed information delivery. For very heavily accessed servers, we evaluated the relative benefit of establishing mirror sites, which could provide easier access but at the cost of extra (and distributed) maintenance of equipment and software.

However, arbitrarily scattered mirror sites will not be sufficient. The Internet's sustained explosive growth calls for an architected solution to the problem of scalable wide area information dissemination. While increasing network bandwidths help, the rapidly growing populace will continue to outstrip network and server capacity as they attempt to access widely popular pools of data throughout the network. The need for more efficient bandwidth and server utilization transcends any single protocol such as *ftp*, *http*, or whatever protocol next becomes popular.

We have proposed to develop and prototype wide area information provisioning mechanisms that support both caching and replication, using the supercomputer centers as 'root' caches. A nationally sanctioned and sponsored hierarchical caching and replicating architecture would be ideally aligned with NSF's mission, serving the community by offering a basic support structure and setting an example that would encourage other service providers to maintain such operations.

We have proposed, in collaboration with Digital Equipment Corporation as a partner on this project, to develop and deploy a scalable caching and replication system across strategic network locations within the United States. Digital Equipment Corporation has committed to significant cost sharing and collaboration, as an integral part of this proposal. Our goal is to facilitate the evolution of U.S. information provisioning with an efficient national architecture for handling highly popular information. Our proposed effort is unique in tying together research and infrastructure. The research and deployment experience of the principal investigators and their proximity to NSF backbone sites will leverage existing NSF-sponsored activities, at the supercomputer centers as well as activities related to resource discovery at the University of Colorado in Boulder.

## 9 collaboration in virtual environments

In addition to providing previously unimagined access to information, the Internet is also making it possible to communicate and collaborate in unprecedented ways. Online role-playing games such as MUDs are evolving into real-time collaborative environments with their own themes, objectives, and social cultures. SDSC is supporting and participating in the development of several such environments. One, called Oceana, targets the K-12 community[1].

The mission of Oceana is to provide a dynamic learning environment that emphasizes ongoing construction of a virtual world. Oceana encourages participants to explore, work together, use their imaginations, and express themselves. Through this experience, the participants can become more aware of their changing physical and cultural environment, learn about group dynamics by working through group problems, and develop self respect. Digital Equipment Corporation, which shares SDSC's interest in collaborative environments, has committed to donating an Alpha workstation to the project as an enhanced replacement for the machine that is the current server.

As public interest in these virtual worlds increases, scaling of the servers will become important to accommodate greater participation. Scalability, in turn, will depend on developing a functional structure to organize the servers and make them easier to find. Currently, people find out about them mainly by word of mouth so tend to access them somewhat randomly. The ANR group is trying to make them more logically accessible by developing mechanisms to organize them based on semantic relationships.

Several servers currently allow intercommunication in the form of sending messages among multiple servers, and some clients support moving from one server to another in a fairly ad hoc manner. We are interested in developing more sophisticated communication and relational channels that allow an object, such as a 'room', to appear simultaneously on multiple servers with people from those various servers freely communicating within this virtual room. This model fosters social interaction across the individual 'worlds' (servers) while obscuring the geographic distribution of the participants. Discussions are under way to implement this capability across a multicasting environment, such as the mbone, to facilitate this kind of communication without a need for a point-to-point communications mesh among all the servers.

[1] telnet://oceana.sdsc.edu:4201/

## 10 NSF very high speed backbone network service (vBNS) and NLANR

The vBNS represents a testbed for the emerging broadband Internet service infrastructure in which all parts of the network will be experimented with: switches, protocols, software, etc., as well as applications. It will be a unique resource for network and application researchers nationwide to explore performance issues with the new technologies (e.g., how host systems and interfaces interact with ATM components of the wide area network). We have submitted a proposal for a National Laboratory for Applied Network Research (NLANR) to NSF, as a vehicle for providing additional facility and research support, necessary to ensure the success of the vBNS.

The NLANR cooperative agreement will focus on research and engineering support for the vBNS, including issues related to facility maintenance and optimization, resource management, and research coordination.

The NLANR cooperative agreement will provide NSF with several important strategic functions:

1. local support at the vBNS sites to facilitate maintenance and operation of the vBNS facility. A 0.25 FTE technician will be responsible for technical support such as maintaining the link between the vBNS and the local site LAN, coordinating installation and maintenance of vBNS equipment, integrating equipment needed to support research efforts, maintaining routing tables, and monitoring network availability.

2. local support at the vBNS sites for research and engineering activities, in particular for vBNS workload characterization and performance evaluation based on detailed usage data. Collaborating with other parties, this is targeted to result in metrics and definitions of system service specifications. A 0.75 FTE researcher/engineer will support usage of the vBNS and facilitate NLANR research. NLANR participants will focus on specific components of vBNS usage:

   - performance and scalability of vBNS usage for high end data movement
   - usage and performance of vBNS for interactive collaboration technologies.
   - usage and performance of TCP/IP over high speed ATM infrastructure.
   - flow characterization of both individual vBNS flows as well as the aggregate workload.
   - work with the vBNS service providers and the NSF to determine what metrics can and should be tracked on a regular basis, and coordinate placement of the statistics into a database for making it broadly available in both standard and customized reports.

3. participation in the vBNS Resource Allocation Committee (RAC). The objective of these meetings will be to track extensions to the vBNS, coordinate tests of new equipment, discuss usage statistics, and coordinate access by new research projects. The NLANR, as a collaborator of the vBNS project, will assist in developing strategies for the implementation of new projects on the vBNS.

4. resource management for the vBNS. It is expected that NSF will grant use of the vBNS to research projects that occasionally will require significant or even dedicated use of the network. NLANR participants will work with the RAC on availability schedules for their local site, and will help establish schedules for when dedicated access to the entire network is feasible.

5. research coordination. A 0.5-FTE Research Coordinator (RC), Dr. Kimberly Claffy at SDSC, will coordinate research usage of the vBNS, and communicate and collaborate as need be with vBNS researchers. The RC will work with the RAC for the day-to-day scheduling of the vBNS as a resource. A schedule will be communicated to vBNS researchers to ensure that research project staff using the vBNS are aware of scheduled usage times for other projects. The NLANR research coordinator will serve as a central coordinator for resource management, performance studies, and vBNS activities, in collaboration with staff of the vBNS service provider, and the NSF.

6. information services to disseminate results to the community. This will facilitate information distribution both among the vBNS and NLANR collaborators as well as to the networking community.

7. overall project coordination and management by the Principal Investigators, and the research coordinator. They will serve as an interface to the NSF and the vBNS researchers.

8. an NLANR Executive Committee (EC) to work collaboratively with NSF/DNCRI on furthering goals and objectives for the national networking agenda, with specific focus on the requirements of the NSF and the vBNS. The EC will consist of the principal investigators of all the sites, the research coordinator, and NSF officials. The EC will provide

a vehicle for addressing implementation issues and coordination with the NLANR sites. The Principal Investigators for the NLANR sites are Doug Carlson (CTC), Marla Meehl (NCAR), Randy Butler (NCSA), Matt Mathis (PSC), and Bilal Chinoy (SDSC).

These support functions will enable network researchers to focus on their particular research while a vBNS research support vehicle is in place.

We have proposed the NLANR project in support of the networking agenda of NSF DNCRI and related divisions for the next several years, particularly in concert with the DNCRI vBNS activity. Both NSF and the NLANR participants understand that this proposal in its current form only constructs a framework for supporting coordinated research on the national vBNS network research facility, but does not include the resources to carry out significant research itself.

## 11 student involvement with ANR

UCSD students Kimberly Claffy and Kevin Fall have made significant contributions to the ANR groups efforts in analysis and modeling of high speed networks. Both have received their PhDs from the Computer Science and Engineering Departments at UCSD in 1994.

1. *Kimberly Claffy* completed her Ph.D degree on *Internet traffic characterization* in June 1994 under Professor George Polyzos in the Computer Systems Laboratory within the Department of Computer Science and Engineering at UCSD. She is currently interested in several areas of development of Internet communication and information middleware.

2. *Kevin Fall* completed his Ph.D. degree in computer science in December 1994 under the direction of Professor Joseph Pasquale in the Computer Systems Laboratory at UCSD. His dissertation is entitled *A Peer to Peer I/O system in Support of I/O Intensive Workloads.* As part of the network research group at SDSC, Kevin has investigated protocol and performance issues in the wide-area HIPPI environment of the CASA gigabit testbed.

3. Henry Sariowan is currently a Ph.D. student in the Computer Systems Laboratory investigating statistical models of Internet flows under the direction of Professor George Polyzos.

## 12 summary

We reported on results as well as work in progress for multiple studies in the area of analysis and modeling of both research and operational wide area networking. Characterization of the traffic composition and performance of wide area data networking environments continues to be an important focus of the activities of the Applied Network Research group. As the emphasis of the ANR group includes aspects of the operational infrastructure as well as academic research activities, the complexities of both areas have influenced the work scope that we describe for 1995.

For 1995 we expect to focus our agenda in three areas: information resources, collaboration resources, and computation resources. We have described our efforts in these areas. In particular, in the first area, we have proposed an architecture for the establishment of information based infrastructure as an application of network provisioning. This project may include hierarchical models, and mirroring, prefetching and caching of information at strategic locations (e.g., NAPs, FIXes) We are also working with virtual collaborative educational projects, reflecting the second category of communication resources. Finally, we presented our project in the third area, computational resources. Specifically we have proposed an national laboratory for applied network research (NLANR) infrastructure for supporting research on NSF's very high speed backbone network service (vBNS). We expect the initial NLANR framework to grow into an important national asset for the NSF and the U.S. research and education community. Underlying all three dimensions of this research agenda is the objective of evolving national networking not only from a networking perspective, but also including the needs of applications and users.

## 13 publications

ANR has published, submitted, or otherwise made available several papers and reports of research activities during 1994. Several papers included collaborative efforts with outside people, who co-author those papers. Among the co-authors are Steve Wolff (NSF/DNCRI), and Roger Bohn (UCSD economist).

ANR's 1993 annual progress report:

*"1993 Annual Report of the Research Progress of the Applied Network Research (ANR) Group", Feb 1994, H.-W. Braun, B. Chinoy, K. Claffy, and G. C. Polyzos, SDSC Report GA-A21648*

Our activities in the area of flow analysis:

*"A parameterizable methodology for Internet traffic flow*

*profiling", K. Claffy, G. C. Polyzos and H.-W. Braun, Jan 1994, IEEE JSAC Special Issue on the Global Internet, to appear. SDSC Report GA-A21526, UCSD Report CS93-328,*

*"Internet workload characterization", K. Claffy, UCSD thesis, June 1994.*

Policy:

*"Mitigating the coming Internet crunch: multiple service levels via Precedence", R. Bohn, H.-W. Braun, K. Claffy and S. Wolff, to appear in Journal of High Speed Networks, 1994.*

Network analysis and modeling:

*"Web traffic characterization: an assessment of the impact of caching documents from NCSA's web server, H.-W. Braun and K. Claffy, Second international world wide web conference, Chicago, IL, October 17-20, 1994.*

*"TCP/IP performance in the CASA gigabit testbed, B. Chinoy and K. Fall, Usenix Symposium on High Performance Networking, 1-3 August 1994, Oakland, CA, pp. 45-59.*

## 14 Acknowledgements

The ANR research was supported by a grant on "Analysis and Modeling Tools for High-Speed Networks" from the National Science Foundation (NCR-9119473), an NSF augmentation of the SDSC base award for work on NREN Engineering issues, and a NSF grant for an "Interagency ATM Connection at the San Diego Supercomputer Center" (NCR-9214104). The CASA component was supported by National Science Foundation and the Advanced Research Projects Agency under Cooperative Agreement NCR-8919038 with the Corporation for National Research Initiatives.

## References

[1] R. Jain and S. A. Routhier, "Packet trains—measurement and a new model for computer network traffic", *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 986–995, September 1986.

[2] J. Mogul, "Network locality at the scale of processes", in *Proceedings of ACM SIGCOMM '91*, September 1991, pp. 273–285.

[3] M. Acharya and B. Bhalla, "A flow model for computer network traffic using real-time measurements", in *Second International Conference on Telecommunications Systems, Modeling and Analysis*, March 24-27 1994.

[4] J.-S. Ahn, Peter B. Danzig, D. Estrin, and B. Timmerman, "Hybrid technique for simulating high bandwidth delay computer networks", in *Proceedings of ACM SIGMETRICS '93*, May 1993.

[5] R. Caceres, P. Danzig, S. Jamin, and D. Mitzel, "Characteristics of wide-area TCP/IP conversations", in *Proceedings of ACM SIGCOMM '91*, September 1991, pp. 101–112.

[6] V. Paxson, "Empirically-Derived Analytic Models of Wide Area TCP Connections: Extended Report", Technical Report LBL-34086, May 1993.

[7] A. Schmidt and R. Campbell, "Internet protocol traffic analysis with applications for ATM switch design", *Computer Communiations Review*, vol. 23, no. 2, pp. 39–52, April 1993.

[8] K. Claffy, G. C. Polyzos, and H.-W. Braun, "Traffic characteristics of the T1 NSFNET backbone", in *Proceedings of IEEE Infocom 93*, March 1993, pp. 885–892.

[9] David C. Feldmeier, "Improving gateway performance with a routing table cache", in *Proceedings of IEEE Infocom 88*, March 1988, pp. 298–307.

[10] R. Jain, "Characteristics of destination address locality in computer networks: a comparison of caching schemes", *Computer networks and ISDN systems*, vol. 18, no. 4, pp. 243–54, May 1990.

[11] D. Estrin and D. Mitzel, "An assessment of state and lookup overhead in routers", in *Proceedings of IEEE Infocom 92*, May 1992, pp. 2332–42.

[12] Nevil Brownlee, "Netramet flow based accounting tool", 1994.

[13] W. Willinger, "Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements", *Statistical Science*, 1994.

[14] Mark W. Garrett and Walter Willinger, "Analysis, modeling and generation of self-similar VBR video traffic", *Proceedings of ACM SIGCOMM '94*, September 1994.

[15] Vern Paxson and Sally Floyd, "Wide-area traffic: the failure of Poisson modeling", in *Proceedings of ACM SIGCOMM '94*, February 1994.

[16] Mosaic Communications Corporartion, "netscape".

[17] Roy Fielding, "*libwww-perl*", http://www.ics.uci.-edu/WebSoft/libwww-perl/archive/, May 1994.

[18] Nikos Drakos, "*latex2html*", http://cbl.leeds.ac.uk/-nikos/tex2html/doc/latex2html/, May 1994.

[19] Thomas Boutell, "*wusage*", http://siva.cshl.org/-wusage.html, National Institutes of Health, Grant P41-RR02188, 1994.

[20] Roy Fielding, "*wwwstat*", http://www.ics.uci.edu/-WebSoft/wwwstat/, July 1994.

[21] Kevin Hughes, "*getstats*", http://www.eit.com/-software/getstats/getstats.html, May 1994.

[22] Mike Vore, "*gnlog*", http://hopf.math.nwu.edu:70/-1/logging, May 1994.

[23] Michael Schwartz, "*harvest*", http://harvest.cs.-colorado.edu/, May 1994.

[24] Bilal Chinoy and Kevin Fall, "TCP/IP performance in the CASA gigabit testbed", in *Usenix Symposium on High Performance Networking*, 1-3 August 1994, pp. 45–59, Oakland, CA.

[25] R. Bohn, H.-W. Braun, K. Claffy, and S. Wolff, "Mitigating the coming Internet crunch: multiple service levels via precedence", *Journal of High Speed Networks*, forthcoming 1994, available for anon-ftp at `ftp.sdsc.edu:pub/sdsc/-anr/papers/`, and `http://www.sdsc.edu/0/SDSC/-Research/ANR/kc/precedence/precedence.html`.