# Activity Report

# 1999

Contact Information

Postal Address International Computer Science Institute
1947 Center Street
Suite 600
Berkeley, CA 94704-1198
USA

Phone (510)-666-2900

Fax (510)-666-2956

E-mail info@icsi.berkeley.edu
Internet http://www.icsi.berkeley.edu

# Introduction

**Sponsors**

AT&T
Nortel
Förderverein (German industrial organization)
German Ministry of Research and Technology
European Media Lab
Spanish Office of Science and Technology
European Union
US Department of Defense
National Science Foundation
DARPA

**Corporate Officers**

Prof. Nelson Morgan (President and Institute Director)
Dr. Scott Shenker (Vice President)
Prof. Christos Papadimitriou (Secretary and Treasurer)

**Board of Trustees**

Dr. H. Ahmadi, AT&T Labs-Research
Prof. G. Ausiello, U. of Rome
Vice Chancellor J. Cerny, UC Berkeley
Prof. F. Denoth, IAT-CNR
Dr. T. Grandke, Siemens CRC
Prof. D. Hodges (Chair), UC Berkeley
Prof. R. Karp, ICSI and UC Berkeley
Mr. R. Kay, former Acting Director of ICSI (until 1988)
Mr. P. Lizcano, Spanish Office of Science and Technology
Prof. K. Mehlhorn, Max-Planck-Institute for Informatics
Prof. C. Papadimitriou, CS Division Chair, UC Berkeley
Prof. D. Schütt, Siemens AG
Prof. H. Schwärtzel, FAST
Prof. D. Tsichritzis, Head of the GMD IT Research Center

**Senior Trustees**

Prof. Domenico Ferrari, U. Cattolica
Prof. Jerome Feldman, ICSI and UC Berkeley
Prof. Gerhard Goos, U. Karlsruhe
Prof. Norbert Szyperski
Prof. Gerhard Barth, Dresdner Bank AG

## Institute Overview

The International Computer Science Institute (ICSI) is an independent, nonprofit basic research institute affiliated with the University of California campus in Berkeley, California. Its establishment was motivated by a recognition of the need for an international fundamental research facility in the field of computer science. ICSI was started in 1986 and inaugurated in 1988 as a joint project of the Computer Science Division of UC Berkeley and the GMD, the Research Center for Information Technology GmbH in Germany. Since then, Institute collaborations within the university have broadened (for instance, with the Electrical Engineering Division as well as to other departments, such as Linguistics). In addition, Institute support has broadened to include a range of international collaborations, US Federal grants, and most recently, direct industrial sponsorship. Throughout these changes, the Institute has maintained its commitment to a pre-competitive, open research program. The goal of the Institute continues to be the creation of synergy between world leading academic and industrial researchers in an international environment through excellence in fundamental research in computer science and engineering.

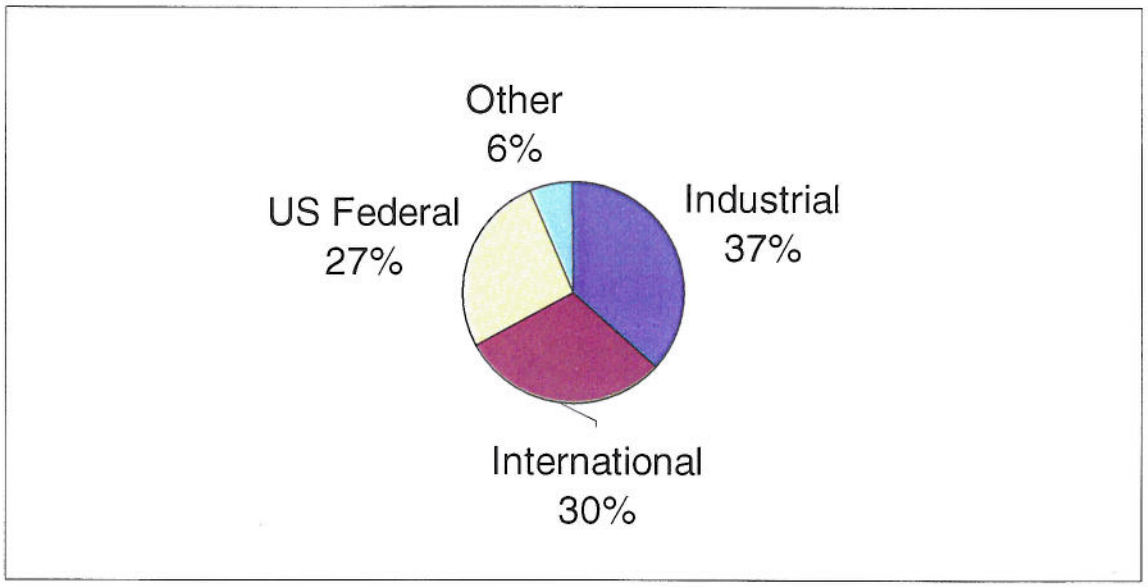
The particular areas of concentration have varied over time, but are always chosen for their fundamental importance and their compatibility with the strengths of the Institute and affiliated UC Berkeley staff. ICSI currently has a major focus on two areas: Internet research, including Internet architecture, related theoretical questions, and network services and applications; and Human Centered Computing (with the main research efforts focusing on natural speech and language processing). Other interests of the Institute scientists spawn smaller collaborative efforts (such as continuing work in theoretical computer science, and bioinformatics), which could potentially lead to more extended activities.

The Institute occupies a 28,000 square foot research facility at 1947 Center Street, just off the central UC campus in downtown Berkeley. Administrative staff provide support for researchers: housing, visas, computational requirements, grants administration, etc. There are approximately eighty scientists in residence at ICSI including permanent staff, postdoctoral Fellows, visitors, affiliated faculty, and students. The senior investigators are listed at the end of this overview, along with their current interests.

## Institute Sponsorship – 1999 fiscal year

As noted earlier, ICSI is sponsored by a range of US Federal, international, an industrial sources. The figure below gives the relative distribution of funding between these different sponsoring mechanisms.

**Figure 1: Distribution of sources of ICSI revenue for 1999.**




US Federal funding comes from a range of grants that support research Institute wide. Most of this funding comes from the National Science Foundation, but significant support is also received from the Department of Defense. International support in 1999 was primarily from government and industrial programs in Germany, with significant new support from the Office of Science and Technology in Spain. Industrial support was primarily from AT&T, but other companies provided additional funding as well.

## Institutional structure of ICSI

ICSI is a nonprofit California corporation with an organizational structure and bylaws consistent with that classification and with the institutional goals described in this document. In the following sections we describe the two major components of the Institute's structure: the Administrative and Research organizations.

## Management and Administration

The corporate responsibility for ICSI is ultimately vested in the person of the Board of Trustees, listed in the first part of this document. Day-to-day operation of the Institute is handled by Corporation Officers, in particular, by the President, and Vice President. The President also serves as the Director of the Institute, and as such takes responsibility for ongoing Institute operations.

The administrative arm of the institute consists of three divisions: Office Administration, System Administration, and Financial Administration. The Office Administration is led by the Office Manager, and is responsible for human resources, office assignments, housing, and in general all office tasks outside of science, computing resources, and finance. The system administration provides support for the ICSI computational infrastructure, and is led by the System Manager. The financial administration is responsible for payroll, grants administration, benefits, and generally all Institute financial matters; it is led by the Controller.

## Research

Research at ICSI is overwhelmingly investigator-driven, and themes change over time as they would in an academic department. Consequently, the interests of the senior research staff are a more reliable guide to future research directions than any particular structural formalism. Nonetheless, through much of its history, ICSI research has been organized into Groups: the Networks Group (internet research), the Theory Group, the Applications/AI Group, and the Realization Group (parallel computer architecture and speech processing). In 1999 we added a fifth group, the AT&T Center for Internet Research at ICSI (ACIRI). Consistent with this history, the bulk of this report is organized along these lines, with one sub-report for each of the five groups.

However, given the change in ICSI research topics since our original group structure was devised, a better way to visualize the work here is as a scatter plot of centers, topics, and projects that can be clustered into a small number of major themes (currently two). Figure 2 shows a caricature of this structure. The elements in the scatter plot correspond to groups of investigators (sometimes overlapping) that work together in some topic area. The large ovals that encompass the majority of these elements correspond to significant clusters of research activity. Currently, these clusters are in Internet Research (IR) and Human Centered Computational Intelligence (HCCI), a term intended to encompass a range of topics related to the use of computing for improved human-machine interfaces and human-human cooperation. Other topics of study are sometimes investigated despite not being a good match to these topic clusters; these are represented by elements outside of the ovals in the figure. Some of the research topics outside of the major clusters will be short-lived, and others may remain small in the future. However, it is likely that at least one of these efforts will ultimately grow into a new cluster of investigations, given sufficient Investigator and Sponsor interest.

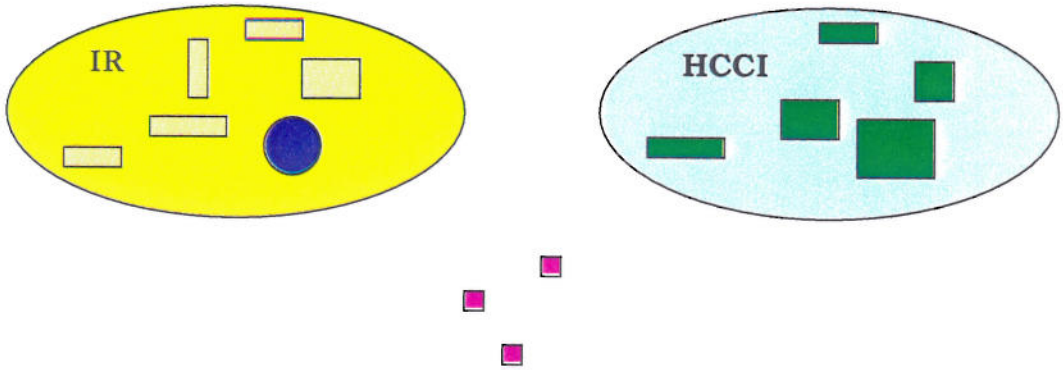



**Figure 2: Caricature of current research efforts at ICSI. A range of investigations can be clustered into the two principle themes of Internet Research (IR) and Human-Centered Computational Intelligence (HCCI); principal efforts in IR include internet architecture and internet services, and for HCCI include speech recognition and natural language understanding. Both sides encompass both system-oriented and theoretical components. Other investigations fall outside of these clusters; current such elements include small efforts in computational molecular biology and in computational complexity.**

## Senior Research Staff

The previous section briefly described the clustering of ICSI research into two major research themes. These themes are descriptions of a current research snapshot, and future descriptions of our work could be extended to new major areas of work based on strategic Institutional decisions and on the availability of funding to support the development of the necessary infrastructure. At any given time, though, ICSI research is best seen as a set of topics that are consistent with the interests of the Research Staff. In this section, we give the names of current (June 2000) senior research staff members at ICSI, along with a brief description of their current interests, the current group name, and the Research Theme (IR or HCCI) that the researcher is most closely associated with. This is probably the best current snapshot of research directions for potential visitors or collaborators. Not shown here are the range of postdocs, visitors, and graduate students who are also key contributors to the intellectual environment at ICSI.

**Jerome Feldman (AI Group, HCCI):** neural plausible (connectionist) models of language, perception and learning and their applications.

**Charles Fillmore (AI Group, HCCI):** building a lexical database for English (and the basis for multilingual expansion) which records facts about semantic and syntactic combinatorial possibilities for lexical items, capable of functioning in various applications: word sense disambiguation, computer-assisted translation, information extraction, etc.

**Sally Floyd (ACIRI Group, IR):** congestion control, transport protocols, queue management, and network simulation.

**Paul Francis (ACIRI Group, IR):** client-based content, distribution and multicast, next generation internet protocols, internet host location service.

**Steven Greenberg (Realization Group, HCCI):** Spoken language processing by humans and machines, analysis of spontaneous speech at the phonetic and prosodic levels, automatic labeling and segmentation of phonetic and prosodic material in spontaneous speech corpora, analysis of automatic speech recognition systems, computational models of auditory processing, hearing-aid design for improving speech intelligibility, auditory-visual speech processing.

**Mark Handley (ACIRI Group, IR):** scalable multimedia conferencing systems, reliable multicast protocols, multicast routing and address allocation, and network simulation and visualization.

**Hynek Hermansky (Realization Group, HCCI):** acoustic processing for automatic speech and speaker recognition, improvement of quality of corrupted speech, human speech communication. (Also with the Oregon Graduate Institute).

**Richard Karp (Theory and ACIRI Groups, IR):** Combinatorial algorithms, computational complexity, randomized algorithms, distributed computation, Internet algorithms, congestion control, genomics, computational molecular biology.

**Nelson Morgan (Realization Group, HCCI):** signal processing and pattern recognition, particularly for speech and biomedical classification tasks

**Vern Paxson (ACIRI Group, IR):** intrusion detection; Internet measurement; measurement infrastructure; packet dynamics; self-similarity.

**Lokendra Shastri (AI Group, HCCI):** Artificial Intelligence, Cognitive Science, and Neural Computation: neurally motivated computational models of learning, knowledge representation and inference; rapid memory formation in the hippocampal system; inference with very large knowledge-bases; neural network models for speech recognition; inferential search and retrieval.

**Scott Shenker (ACIRI Group, IR):** congestion control, internet topology, game theory and mechanism design, scalable content distribution architectures, and quality of service.

**Elizabeth Shriberg (Realization Group, HCCI):** Modeling spontaneous conversation, disfluencies and repair, prosody modeling, dialog modeling, automatic speech recognition, utterance and topic segmentation, psycholinguistics, computational psycholinguistics. (Also with SRI International).

**Andreas Stolcke (Realization Group, HCCI):** probabilistic methods for modeling and learning natural languages, in particular in connection with automatic speech recognition and understanding. (Also with SRI International).

# Research Group Reports

Contents

1 AT&T Center for Internet Research - ACIRI ..... 1
1.1 Overview ..... 1
1.2 Research on Internet Architecture ..... 1
1.3 Research on Internet Operations ..... 6
1.4 Research on Internet Theory ..... 10
1.5 Community Activities ..... 11
2 Realization Group - Speech Processing: Signals and Systems ..... 16
2.1 Signal Modeling ..... 16
2.2 Recognition ..... 19
2.3 Perception and Understanding ..... 22
2.4 Applications ..... 23
3 Artificial Intelligence and its Applications ..... 29
3.1 Learning and Language Use ..... 29
3.2 FrameNet Project ..... 31
3.3 Connectionist Modeling ..... 32
3.4 Biological grounding of recruitment learning and modeling episodic memory formation 35
3.5 Temporal flow models for speech processing ..... 35
4 Network Services and Applications ..... 37
4.1 Overview ..... 37
4.2 Simulation of a Mobile Network Architecture for Vehicles ..... 37
4.3 Equation Based Congestion Control (joint project with ACIRI) ..... 38
4.4 Projects of the Middleware Technology Research Group at ICSI ..... 38
4.5 Anonymity and Unobservability in the Internet ..... 39
4.6 High Speed Cable Access for Internet and Telelearning Applications ..... 40
4.7 Multiuser Detection Receivers in CDMA systems ..... 41
5 Theory of Computation ..... 43
5.1 Introduction ..... 43
5.2 Mathematics of Computer Networking ..... 43
5.3 Computational Molecular Biology ..... 44
5.4 Computational Complexity ..... 45
5.5 Combinatorial Optimization ..... 46

## AT&T Center for Internet Research - ACIRI

### 1.1 Overview

ACIRI's biggest achievement in 1999 was its creation. In the eighteen months of its existence, ACIRI has already become quite visible in the Internet community. For example, at the 1999 SIGCOMM conference, 6 out of the 24 papers had ACIRI authors. Moreover, many of the relevant committees and research groups in the Internet community have significant, and sometimes overwhelming, ACIRI representation. Fulfilling the hopes we all held for it, ACIRI is now widely recognized as a center of excellence for Internet research.

The research efforts at ACIRI can be roughly divided into three main areas:

**Internet Architecture:** The basic Internet architecture, by which we mean the IP protocol family and closely related issues, is the fundamental technical underpinning of the Internet.

**Internet Operations:** This area encompasses issues that are important in keeping the Internet operating efficiently and securely.

**Internet Theory:** One of the goals in setting up ACIRI was to become a leader in developing the *science* of datagram networks. A necessary component of this is to bridge the gap between traditional theoretical computer science and the Internet community.

Sections 1.2 – 1.4 of this report describe ACIRI's major research projects in these areas.

Active participation in the Internet community is another important aspect of ACIRI's mission. ACIRI members are active in several different organizations of the Internet community, most notably:

**IETF:** The Internet Engineering Task Force is the standards body for the Internet community.

**SIGCOMM:** SIGCOMM is the ACM special interest group on communications, and the conference it sponsors is the premier academic conference for the Internet community.

**VINT:** The Virtual Internet Testbed project (joint collaboration between USC/ISI, UCB/LBL, and Xerox PARC) aims to create and support a common network simulator environment for the Internet community. The goal is to encourage more synergy between different research groups by allowing them to share simulator modules. ACIRI members have been quite involved in VINT from the beginning.

Section 1.5 contains a more detailed list of ACIRI contributions to the research and standards activities in the Internet community.

### 1.2 Research on Internet Architecture

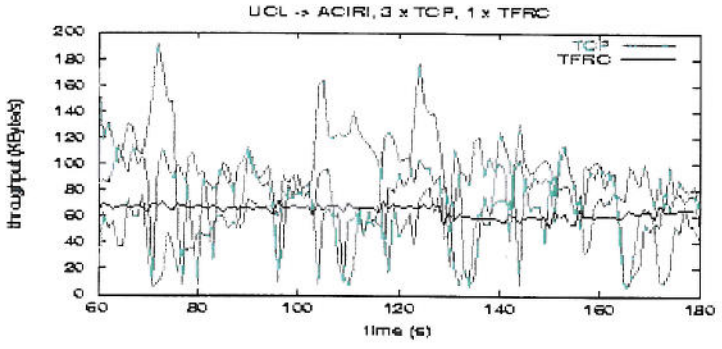
There are five main efforts in ACIRI's research program on Internet architecture: congestion control, multicast, measurement infrastructure, web caching infrastructure, and admission control.

**Congestion Control:** The Internet relies on end-to-end congestion control to keep it from becoming overloaded. As such, congestion control is crucial to the stable functioning of the Internet. The predominant form of congestion control is that embodied in TCP. The basics

of this congestion control algorithm are well understood, but it remains an open question how to extend congestion control to applications such as streaming media that don't use TCP. Since streaming media applications are likely to become an increasingly important source of Internet traffic (and revenue), finding an adequate congestion control solution for this class of applications is of paramount importance to the future stability of the Internet. Other open problems in congestion control are how to provide better congestion signals to TCP and how streaming media applications should manage their quality in the presence of congestion control.

- **Equation-Based Congestion Control:** Because TCP is the most prevalent form of congestion control on the Internet, there is widespread consensus that any non-TCP congestion control algorithm must be *TCP-friendly* or *TCP-compatible*; that is, it must coexist with TCP without consuming an undue share of the bandwidth. Most attempts to create TCP-friendly congestion control algorithms involve imitating TCP's window adjustment algorithm. However, the TCP window adjustment algorithm is not suitable for multicast applications (because of ACK implosion) and streaming media applications (because they cannot tolerate the highly variable bandwidths provided by TCP). ACIRI researchers, along with others, are working on a different approach to TCP-friendly congestion control called *equation-based* congestion control. Equation-based congestion control, rather than imitating TCP's window adjustment algorithm, seeks to match TCP's bandwidth usage equation (the equation that describes the bandwidth usage that results in the presence of a given packet-drop rate). That is, the equation-based approach imitates the end result of TCP but not TCP's means of achieving that end. During this past year, a unicast version of this approach has been developed and extensively analyzed. Work has also been done on how to use the resulting rate adaptation algorithm in a sender-based multicast congestion control algorithm [45]. This work will likely result in a forthcoming SIGCOMM submission.

TFRC (TCP-Friendly Rate Control) is a proposed mechanism for equation-based congestion control for unicast traffic.




**This figure shows a typical experiment with three TCP flows and one TFRC flow running concurrently from London to Berkeley, with the bandwidth measured over one-second intervals. In this case, the transmission rate of the TFRC flow is slightly lower, on average, than that of the TCP flows. At the same time, the transmission rate of the TFRC flow is smooth, with a low variance; in contrast, the bandwidth used by each TCP flow varies strongly even over relatively short time periods.**

- **Explicit Congestion Notification:** TCP senses congestion on the Internet by detecting packet drops. However, since the early DEC-bit era there has been interest in delivering Explicit Congestion Notification to end hosts. There is currently an effort in the IETF, led by Sally Floyd and K. K. Ramakrishnan, that advocates the use of a single bit in the packet header to indicate congestion. The basic idea is quite simple, but there are some very subtle issues to deal with. The past year, they have analyzed the interactions between ECN and IPsec [23], and figured out how ECN would work with MPLS [33]. There has also been some work on more general security issues (*e.g.,* how do you prevent a router from removing the congestion indications?). In addition to the technical work involved, considerable time has been devoted to convincing vendors of the utility of ECN.

- **Congestion Control for IP Telephony:** IP will soon be widely used for telephony, and the vagaries of best effort service will inevitably lead to bandwidth variations. How should IP telephony cope with variable bandwidth? That is, given some input from a congestion control algorithm that dictates its rate, how should the application best manage its quality? There are several techniques for managing quality (FEC, changing codecs, etc.) and the goal here is to investigate the speech quality that results from switching between them to adjust rates. The original goal was to find heuristics for optimizing speech quality given mandated changes in the allowed transmission rate. We found that this problem was substantially harder than initially realized, and that further work is needed to clarify the situation. This work is related to earlier work on quality adaptation for video [13] presented at SIGCOMM 99.

- **Other Congestion Control Topics**

**TCP enhancements:**

1) We are working on an extension to TCP SACK in which the receiver reports to the sender the receipt of duplicate packets [24]. This makes TCP significantly more robust in an environment of reordered packets, ACK loss, packet replication, and/or early retransmit timeouts. This is not a conceptually deep change, but it has significant practical importance and it won't get done without ACIRI involvement.

2) We are working on a proposal for adjusting TCP's congestion window after an idle or application-limited period [25]. This will allow TCP to retain some congestion state during an idle period but will avoid congestion-inducing bursts.

3) We have coauthored the IETF's standard documenting TCP congestion control [37], documented known TCP implementation problems [41], and are working on the standard documenting the computation of TCP's retransmit timeout [31].

**Simple Control Transfer Protocol (SCTP):**

This is a reliable transport protocol originally proposed for telephony but now intended for wider use. While not initiated by ACIRI, the development (rewriting) of this protocol has required a substantial amount of ACIRI guidance and support [34].

**ECM:**

We are active in, and were instrumental to the formation of, the IETF's working group on Endpoint Congestion Management, which aims to unify congestion control across an endpoint's different connections and different protocols.

- **Future Directions** While the current focus of our equation-based congestion control work is for unicast, we hope to soon continue our earlier work of extending equation-based congestion control to the multicast case. Moreover, we are investigating the stability of such equation-based congestion control schemes. Much of this work has centered on the end-host algorithms, but we are also quite active in investigating router support for congestion control (in addition to the ongoing work on active queue management and ECN deployment in routers). There are two different known approaches, those that detect ill-behaved flows and those that explicitly manage the bandwidth allocations; ACIRI is pursuing research in both. For the former, we are refining the algorithms presented in [8]. For the latter, we are investigating approaches to providing max-min fair bandwidth allocation that do not require per-flow state in the routers. This work is an extension of the CHOKE work of Prabhaker and Rong (Stanford), and also of the CSFQ work.

**Multicast** : The introduction of multicast is a significant conceptual and technical change to the basic Internet architecture. While research in this arena has been pursued for well over a decade, scalability remains a vexing issue. Multicast will not be widely deployed until the scalability problems have been adequately resolved. Consequently, much of ACIRI's research focus has been devoted to the problem of multicast scalability. This has been pursued from several angles.

- **Multicast Routing Protocols and Address Allocation.** Currently there is no true inter-domain multicast routing protocol deployed on the Internet. We are involved in the design and development of the Border Gateway Multicast Protocol (BGMP) which is a proposed solution for inter-domain multicast routing. To minimize the state needed to join senders and receivers together into a multicast delivery tree, BGMP utilizes a structuring of the multicast address state, and is therefore tied to multicast address allocation. We have been working in the IETF to standardize a scalable mechanism for dynamically assigning multicast address ranges to domains (the structure that BGMP requires) using the Multicast Address Set Claim (MASC) protocol [21]. In addition, we have designed a protocol (AAP) for allocating multicast addresses from these ranges in a dynamic and fault-tolerant manner [27]. This work is expected to be standardized in the IETF in the near future.

There is not universal agreement that BGMP is the right inter-domain multicast routing solution for the Internet. To this end, we are also pursuing an alternative track, by devising a bi-directional variant of Sparse-mode PIM which is the prevalent intra-domain multicast routing solution currently deployed [20]. There are two hopes for this work: that as an intra-domain solution it will provide a better match with BGMP; and that it could also provide an alternative to BGMP. Whilst we believe BGMP is the better inter-domain solution, it seems best to consider alternative solutions given the difficulty of getting any wide-area routing protocol deployed on the Internet.

- **Multicast Address Aggregation** Although we do not yet know which it will be, eventually a scalable inter-domain multicast routing protocol will be deployed on the

Internet. This leads to the additional question of how much multicast forwarding state we are likely to see in multicast routers. The issue is that current and foreseeable multicast routing protocols require at least per-multicast-group state in routers. One option is to aggregate this state in router forwarding tables, and we have investigated the gains that can be achieved using this technique for different classes of multicast routing protocol. Prior work had incorrectly stated that multicast forwarding state could not be aggregated. We show this is incorrect, and that the amount of aggregation that can be achieved without loss of performance is approximately an order of magnitude in backbone routers. This work will be presented at INFOCOM 2000 [17].

- **Reliable Multicast** For multicast to be widely useful, standards for reliable multicast protocols must be devised. While there is no shortage of ideas, the stumbling block has been the lack of suitable congestion control mechanisms. We have been actively working on a technique for multicast congestion control based on the unicast congestion control work described above. This work is currently taking place in the Internet Research Task Force (IRTF) Reliable Multicast Research Group (chaired by Mark Handley), but we expect it to be solid enough to move to the IETF for standardization in the next year. See [45, 26, 35] for summaries of the progress to date.

A new working group (Reliable Multicast Transport) has recently been formed in the IETF to take on this standardization. We have been actively involved in the form this standardization process has taken, in an attempt to avoid the standardization of a large number of different and incompatible protocols. As no single RM protocol will suit all applications we are attempting to standardize protocol building blocks that are common to many such protocols.

- **Future Directions** We are initiating a project to investigate the role Dynamic Packet State (DPS) could play in making multicast, and reliable multicast, more scalable. DPS, which was developed in joint work with Ion Stoica and Hui Zhang (CMU), appears to be quite useful in reducing router state for congestion control and QoS issues. However, its effectiveness in other arenas has not yet been adequately investigated.

**Web Caching Architectures**: From a purely layering point of view one might argue that web caching is not a fundamental architectural issue, living as it does substantially above the IP layer. However, the web dominates today's Internet traffic and providing an effective web caching architecture is crucial for the long-term viability of the Internet. ACIRI's work on web caching complements the extensive work being done at Florham Park.

- **Caching for Streaming Media** Most web caching research focuses on traditional file retrieval. However, streaming media is becoming an important presence on the Internet, and we must find ways to effectively cache this content. Of particular interest is how caching should be done for quality-adaptive streaming applications. The approach advocated in this project has additional layers pre-fetched as the popularity of a stream increases, so that the cache can deliver high quality versions of popular streams. A paper describing this work will appear in Infocom 2000 [14].

- **Consistency Architectures for Web Caching.** While web caching greatly reduces network load, it comes at the price of occasionally delivering stale files; that is, the server may have updated a page but the cache, not knowing about the update, may deliver the old

version of the file in response to a client request. This project investigates what architectural support would be needed to ensure that no pages older than a small staleness bound would be delivered. The consistency architecture uses invalidations to preserve consistency (guaranteeing a small staleness bound) and cache-to-cache heartbeats to aggregate the overhead of keep-alives. Current research is focusing on the role of the cache organization with particular focus on the tradeoffs between mesh and hierarchy configurations. This paper was presented at SIGCOMM '99 [18].

**Measurement Infrastructure** Historically, the Internet has been woefully under-measured and under-instrumented. The problem is only getting worse with the network's ever-increasing size. The National Internet Measurement Infrastructure project (NIMI), for which ACIRI provides technical leadership, aims to develop a scalable architecture for deploying and operating measurement infrastructures, i.e., a collection of measurement "platforms" that can cooperatively measure the properties of Internet paths and clouds by transmitting test traffic among themselves. The architecture emphasizes decentralized control of measurements; strong authentication and security; mechanisms for both maintaining tight administrative control over who can perform what measurements using which platforms, and delegation of some forms of measurement as a site's measurement policy permits; and simple configuration and maintenance of platforms. NIMI currently has 23 hosts, and plans to expand to roughly 35 in the near future.

**Measurement-Based Admission Control** Over the past decade, much work has been devoted to extending the network to allow flows to reserve bandwidth. While the fate of these *Integrated Services* proposals remains in doubt, work continues on some of the basic technical questions. Admission control is at the heart of the Integrated Services architecture. Measurement-based admission control (MBAC) uses measurements of recent network performance to guide the admission control decisions. This approach yields much higher utilizations than the traditional parameter-based approach, and many MBAC algorithms have been proposed in the literature. AT&T in particular has been investigating the use of such algorithms in their network. In this project we evaluated many of the proposed algorithms, and found that they all had essentially the same performance, leading us to conclude that there are underlying limits to MBAC behavior that are independent of the details of the admission control equation. This work will be presented at Infocom 2000 [6]. We have also had extensive discussions with Matt Grossglauer (AT&T) on this topic.

**Future Directions** We think that the work on router-based MBACs is essentially complete. In collaboration with Ion Stoica and Hui Zhang of CMU, we are now pursuing the question, first posed by Frank Kelly and coauthors, of whether measurement-based admission control can be adequately performed at the edges of the network rather than in transit routers. If possible, this could be another step towards a *stateless core*, at least in terms of congestion control and QoS related state. Moreover, it would lead to significantly easier deployment of more advanced services, such as the proposed *controlled load* service, since the interior routers would not have to be modified.

### 1.3 Research on Internet Operations

In the days when the Internet was a protected research playground, no one cared greatly keeping the network secure and stable. However, now that the Internet is a bustling center

of commercial activity, security and stability are of the utmost importance. ACIRI is addressing security through its work on network intrusion detection. ACIRI's work on measurement and analysis aims to make networks more understandable, predictable, and stable.

**Network Intrusion Detection** While it would be desirable to have impenetrable computer systems, achieving this is immensely difficult, as it requires airtight mechanisms to relate authentication and enforcement with widely varying local policies. Consequently, in the operational world it becomes paramount to be able to detect misuse in real time so that one can take appropriate countermeasures. ACIRI has several related projects on network intrusion detection.

**Bro** The Bro project (LBNL and ACIRI) is a network intrusion detection system. It sniffs packets coming across a network link such as DMZ or a sensitive LAN and uses an event engine to analyze the traffic and extract from it events at different levels (*e.g.,* connection attempted; user authenticated; FTP file retrieve request; new line of Telnet output). It then determines whether the traffic is consistent with the site's policy by running the series of events as input to a script that expresses the policy in a domain-specific language. The script can maintain and modify global state, record information to stable storage, synthesize new events, generate real-time alerts, and invoke shell scripts as a form of ``reactive response".

Bro is currently operational at several sites (ICSI, LBNL, UCB, NERSC, ESNET, JGI) and is in the process of being deployed at Florham Park. The code, significantly extended over this past year, is freely available.

**Traffic Normalization** There are many ways one can perturb traffic to make it difficult for a network intrusion detection system to unambiguously interpret the packet stream; for instance, a retransmission can contain different data than the original packet. This is a fundamental problem for network intrusion detection, in the sense that if it is not solved, network intrusion detection is doomed---because just as attackers today use a wide variety of toolkits to automate their attacks, so too will they use a wide variety of evasion toolkits to escape detection. Indeed, such toolkits already exists, and Phrack magazine, the effective journal of record for the attacker underground, has already published an article on the topic.

A traffic normalizer is a box that takes a packet stream and attempts to reshape the traffic flowing through it in order to eliminate ambiguities that can facilitate evasion, but while leaving the end-to-end semantics of the stream unperturbed. A draft paper has been written [46] and further work is continuing.

**Stepping Stone Detection** Intruders often use s*tepping stones* to hide the original source of an attack. This project seeks to identify correlated connections that are serving as stepping-stones, so the attack can be traced back to its source. The initial approach based on looking for coincidences in connection idle periods, appears to work quite well for interactive traffic monitored at a single point (such as a site's ingress/egress link), even if the traffic is encrypted. This work was presented in the New Research session at Sigcomm.

**Future Directions** Our stepping stone research has led to interest in investigating an important generalization of the problem: detecting correlated traffic streams when measuring traffic at different, possibly widely separated points in a network. For example, an attacker may make a connection from host $A$ to $B$, then traverse a large number of intermediary sites, and then finally launch an attack from $Y$ to $Z$. What we want is for site $Z$, which has

detected an attack but has no idea that it originated at $A$, to somehow coordinate with other cooperating sites such that site $A$ can detect that in fact the attack traffic traversed its path to site $B$. To achieve this requires devising robust, unique timing signatures of traffic such that $Z$ can publish the signature of the attack and $A$, either in real-time or retrospectively, can determine that it has hosted traffic with a very similar signature (though not identical, due to network noise and other effects). We are pursuing this ``Internet Trap and Trace" research in a multi-institution collaboration between ACIRI, UC Davis, and the University of North Carolina [47].

A second line of research we have begun concerns detecting ``back doors." That is, services running on ports other than those specified by a site's policy. One of the most important examples is an interactive login back door (such as a telnet, rlogin, or ssh server) left behind by an attacker after a successful attack to provide easy future access. Another example is an unauthorized FTP server running on a non-standard port, being used as a repository for pirated software. We are pursuing this work looking at both protocol signatures and timing signatures.

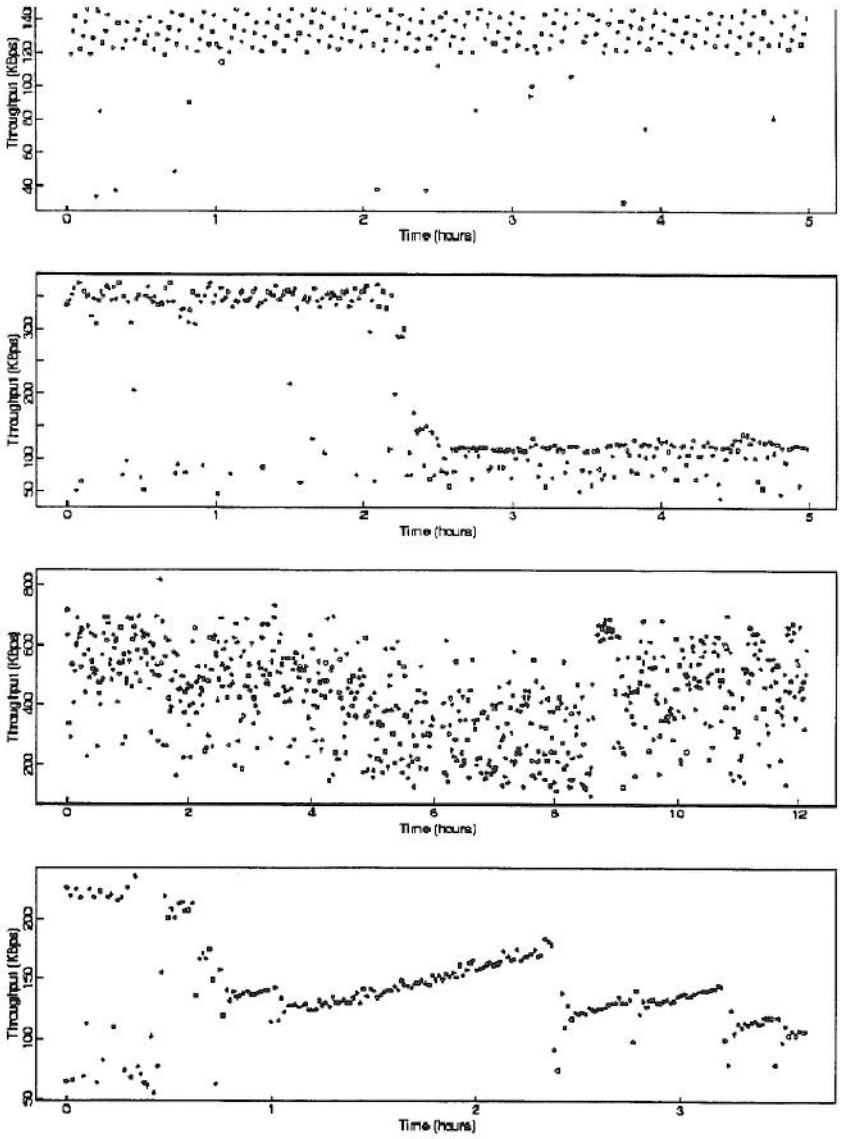
**Measurement and Analysis** Sound, on-going measurements of network link and path properties are fundamental for both trouble-shooting and efficiently operating a large network. Such measurements are also fundamental for developing robust characterizations of how network traffic behaves and evolves, if a sufficiently rich set of measurements can be made. ACIRI has several projects in this arena, in addition to the NIMI measurement infrastructure discussed previously.

**MINC** This project seeks to infer the properties of individual network links from the end-to-end drop and delay characteristics of multicast traffic. Such ``tomography" approaches offer the intriguing possibility of pinpointing network performance problems without requiring ubiquitous measurement access to individual routers.

**Network Path Characterization** The initial focus of this project has been twofold: (1) estimating the appropriate retransmission timer (RTO) for a reliable protocol, and (2) estimating the bandwidth available to a connection as it begins. This is ongoing work that resulted in a SIGCOMM paper this year [3]. Further refinements and plans for live experiments are in progress.

**Future Directions** We have initiated a project with Lee Breslau (AT&T) investigating the extent to which network properties are well-modeled as stationary.. There are several time scales over which measurements are used, all the way from round-trip times for adaptive applications to monthly measurements for inter-ISP contractual agreements. The question is: to what extent are these measurements reasonable predictors of the future? We are using the NIMI measurement infrastructure and have gathered some initial data.

Different patterns of Internet throughput: this figure is from a study of how the performance of Internet paths varies with time, for time scales ranging from seconds to hours. The figure shows how quickly (in kilobytes per second) a large file could be transferred between a pair of Internet hosts, for measurements made across several hours. The top plot shows a steady pattern indicating a completely unloaded network; the next shows an abrupt shift in the network's transfer capacity; the third shows throughput that varies by a factor of three from minute to minute; and the last shows a puzzling, slow increase in throughput that is periodically reset to a lower value.




Another in-progress project is looking at the dynamics of DNS traffic. This effort aims to update the groundbreaking DNS study by Danzig and colleagues, published in Proc. SIGCOMM '92. Working with Ramon Caceres (AT&T), we have begun gathering traces

from a wide variety of perspectives within the DNS, including dialup end-users, campus name servers, top-level domain servers, regional and backbone perspectives, and root name servers.

### 1.4 Research on Internet Theory

One of the initial goals of ACIRI was to bring about closer interactions between the theoretical computer science community and the Internet community. While there had been some algorithmic work related to Internet problems (notably Edith Cohen's work on various web-related problems), there had been little use of traditional theoretic techniques in addressing basic architectural problems.

**Scaling of Multicast Trees** One of the advantages of multicast is that it reduces the overall network load. Despite the years of work on multicast, this performance advantage had never been quantified. In our work, following the lead of Chuang and Sirbu, we studied the number of link traversals $L(n)$ (the number of times a packet traverses a network link) needed to reach a multicast group of size $n$. This function $L(n)$, for large $n$, describes the way multicast trees scale. We found that for a wide variety of networks $L(n) \approx n(c - d \log \frac{n}{M})$ for some constants $c$ and $d$ (with $M$ denoting the number of sites in the network). This result is to be contrasted with the linear growth in link traversals that holds for unicast transmissions. The result is based on analytical approximations for simple solvable cases, and empirical evidence for more general networks. Our work was presented at SIGCOMM '99 [12]. Continuing research (joint with Ramesh Govindan and Deborah Estrin of USC/ISI) is now focusing on the nature of network topologies and the characteristics of current network topology generators.

**Congestion Probing** One can think of TCP's window increase/decrease algorithm as probing the network to detect the amount of available bandwidth. The question we address here is what is the optimal way to do this kind of congestion probing. We considered a simplified model in which a fixed amount of bandwidth $u$ is available to a flow, and the flow must probe to determine this bandwidth. We assume that time is divided into periods (equal in length to the timeout parameter for the flow) and that in each period the sender transmits $x$ packets and finds out whether it has suffered any packet drops. If so it assumes that $x > u$, otherwise it assumes that $x \leq u$. We considered two loss models. In both models, the loss when $x \leq u$ is $u - x$. When $x > u$ the loss is $u$ in the first model and $\alpha(x - u)$ in the second. We focused on the expected loss (assuming $u$ is uniformly distributed in some range) and the worst-case loss given an initial upper bound on $u$. Our results include:

- A simple algorithm for the second model that (almost exactly) minimizes expected loss.
- An efficient dynamic programming algorithm for optimizing expected loss in the first model; the technique applies also to very general cost functions, and can accommodate time discounting.
- Matching upper and lower bounds of $u \log \log u$ on the worst-case loss in the first cost model.
- An exact game-theoretic analysis of the single-period problem.

Much work remains to be done. We'd like to analyze the more realistic setting where $u$ varies from period to period, as well as cost functions reflecting more complex retransmit/acknowledgement protocols.

The general goal of this research program is to put on a scientific basis the study of how flows should voluntarily adjust their sending rate according to feedback about competing traffic.

**Complexity of Cost Sharing Algorithms** Theoretical computer science has typically focused on the complexity of algorithms. Game theory and economics have traditionally focused on the incentive aspect of algorithms. In this work we are interested in both incentives and complexity. The work is part of a larger research agenda to characterize the communication complexity of game-theoretic network mechanisms; that is, how complex do network mechanism need to be to deal with selfish agents. Our current specific target is the communication complexity of strategy-proof cost sharing algorithms for multicast flows. Game-theoretic considerations point to two reasonable ways to share costs which, from the incentive perspective, are roughly equivalent. We find that the communication complexity inherent in implementing these mechanisms are strikingly different; one method involves the minimal amount of communication, and the other involves the maximal amount. This is one of the few attempts to characterize the *network complexity* of fairly realistic network protocols, and as such we expect the resulting methodology to have wide applicability. This work has been submitted to STOC 2000 [7].

**Future Directions** We are investigating, along with Mike Luby of ICSI, ways to share requests in the ``digital rainbow" protocols due to McCanne *et al.* in which data is being multicast to a group using the digital fountain technique. Each receiver generates requests for packets and adjusts the size of its window of in-progress requests by a TCP-like protocol in which the packets it receives play the role of ACKs. Sharing is achieved by allowing a router to combine requests; *i.e.,* when a request is received it can be satisfied by the reply to an earlier request currently in process. The main goal is to analyze the performance improvements realized by such request sharing.

We are also investigating (with C. Schindelhauer and D. Voecking of ICSI) how to best spread ephemeral information (rumors) in a network. Different assumptions about network connectivity and agent knowledge lead to very different bounds on the performance of such mechanisms.

### 1.5 Community Activities

ACIRI members engage in a wide range of community activities. While many of these activities are quite standard (program committees, editorial boards, journal reviewing, etc.) and won't be listed here, this report will focus on the three activities that are more central to the Internet community.

**IETF** The Internet Engineering Task Force (IETF) is the most relevant standards body for the Internet. Vern Paxson is co-director of the Transport Area of the IETF, and is lead Area Director (or co-chair) of 17 IETF Working Groups. This year he wasthe principal behind chartering the RMT (reliable multicast transport), PILC (performance implications of link characteristics), ECM (endpoint congestion management), SPIRITS (signalling from the PSTN to the Internet), and ENUM (mapping E.164 telephone numbers to sets of identifiers

such as URLs) working groups. Sally Floyd and Mark Handley serve on the Transport Directorate, an advisory body for the Transport Area, and are otherwise quite active in the IETF. In particular, Mark Handley has been a long-time chair of the Multiparty Multimedia Session Control WG (he resigned in July).

**SIGCOMM and IRTF** The Internet Research Task Force is the research counterpart of the IETF. Mark Handley is chair of the Reliable Multicast Research Group, and is a member of the Internet Research Steering Group (IRSG). Sally Floyd and Scott Shenker are members of the IRTF End-to-End Research Group, and Vern Paxson and Mark Handley are frequent invitees to its meetings.

SIGCOMM is ACM's special interest group on communications, and is the primary academic research organization for the Internet community. Sally Floyd was Vice-Chair for ACM SIGCOMM from July 1995 to June 1999. Scott Shenker is chair of the SIGCOMM Technical Advisory Committee. Several members of ACIRI serve quite regularly on the SIGCOMM Program Committee.

**VINT** The Virtual Internet Testbed project [4] is based on the original *ns* simulator developed at UCB by Steve McCanne and Sally Floyd. Sally Floyd, Scott Shenker, and Mark Handley have been involved in VINT from the project's inception, and Sally and Mark continue to participate in the weekly VINT meetings and other general oversight activities. We are also actively developing new validation test suites for *ns* [43].

**References:**

**Books**

[1] J. Crowcroft, M. Handley, and I. Wakeman. *Internetworking Multimedia.* Morgan Kaufmann, 1999, ISBN 1-55860-584-3

**Papers**

[2] A. Adams, T. Bu, R. Caceres, N. Duffield, T.Friedman , J. Horowitz, F. Lo Presti, S.B. Moon, V. Paxson, and D. Towsley. The Use of End-to-End Multicast Measurements for Characterizing Internal Network Behavior. Submitted to *IEEE Communications*, special issue on network measurement, September 1999.

[3] M. Allman and V. Paxson. On Estimating End-to-End Network Path Properties. In *Proceedings ACM SIGCOMM '99*, September 1999, Cambridge, MA.

[4] S. Bajaj, L. Breslau, D. Estrin, K. Fall, S. Floyd, P. Haldar, M. Handley, A. Helmy, J. Heidemann, P. Huang, S. Kumar, S. McCanne, R. Rejaie, P. Sharma, K. Varadhan, Y. Xu, H. Yu, and D. Zappala. Improving Simulation for Network Research. Technical Report 99-702, University of Southern California, March, 1999. Under submission.

[5] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proceedings of Infocom '99.*

[6] L. Breslau, S. Jamin, and S. Shenker. Comments on the Performance of Measurement-Based Admission Control Algorithms. To appear in *Proceedings Infocom 2000.*

[7] J. Feigenbaum, C. Papadimitriou, and S. Shenker. Sharing the Cost of a Multicast Transmission. Submitted to *STOC 2000.*

[8] S. Floyd, and K. Fall. Promoting the Use of End-to-End Congestion Control in the Internet. In *IEEE/ACM Transactions on Networking*, August 1999. This is a revised version of sections of Router Mechanisms to Support End-to-End Congestion Control, Technical report, February 1997.

[9] S. Floyd and V. Paxson. Why We Don't Know How To Simulate The Internet. Under review. This is a revised and expanded version of a paper in *Proceedings of the 1997 Winter Simulation Conference*, December 1997.

[10] A. Greenwald, E. Friedman, and S. Shenker. Responsive Learning in Network Games. To appear in *Games and Economic Behavior.*

[11] H. Moulin and S. Shenker. Distributive and Additive Cost Sharing of an Homogeneous Good. To appear in *Games and Economic Behavior.*

[12] G. Phillips, H. Tangmunarunkit, and S. Shenker. Scaling of Multicast Trees: Comments on the Chuang-Sirbu Scaling Law. In *Proceedings ACM SIGCOMM '99*, September 1999, Cambridge, MA.

[13] R. Rejaie, M. Handley, D. Estrin. Quality Adaptation for Congestion Controlled Playback Video over the Internet. In *Proceedings ACM SIGCOMM '99*, September 1999, Cambridge, MA.

[14] R. Rejaie, H. Yu, M. Handley, D. Estrin. Multimedia Proxy Caching for Quality Adaptive Streaming Applications in the Internet. To appear in *Proceedings Infocom 2000.*

[15] V. Paxson. End-to-End Internet Packet Dynamics. *IEEE/ACM Transactions on Networking*, Vol.7, No.3, pp. 277-292, June 1999.

[16] V. Paxson. Bro: A System for Detecting Network Intruders in Real-Time. To appear in Compu*ter Networks and ISDN Systems*, special issue on intrusion detection, 1999.

[17] D. Thaler and M. Handley. On the Aggregatability of Multicast Forwarding State. To appear in *Proceedings of Infocom 2000.*

[18] H. Yu, L. Breslau, and S. Shenker. A Scalable Web Cache Consistency Architecture. In *Proceedings ACM SIGCOMM '99*, September 1999, Cambridge, MA.

**Internet Drafts**

[19] S. Bradner and V. Paxson. IANA Allocation Guidelines For Values In the Internet Protocol and Related Headers. Work in progress, Internet-Draft draft-bradner-iana-allocation-02.txt, October 1999.

[20] D. Estrin, M. Handley, I. Kouvelas, and L. Vicisano. A New Proposal for Bi-directional PIM. draft-kouvelas-pim-bidir-new-00.txt, 10/20/1999.

[21] D. Estrin, M. Handley, and D. Thaler. The Internet Multicast Address Allocation Architecture. draft-ietf-malloc-arch-03.txt, 10/21/1999.

[22] S. Floyd. Congestion Control Principles. Internet draft draft-floyd-cong-00.txt, October 1999.

[23] S. Floyd, David Black, and K.K. Ramakrishnan. IPsec Interactions with ECN. Internet-draft draft-ietf-ipsec-ecn-01.txt, October 1999. This is a revised version of draft-ietf-ipsec-ecn-00.txt, April 1999.

[24] S. Floyd, J. Mahdavi, M. Mathis, M. Podolsky, and Allyn Romanow. An Extension to the Selective Acknowledgement (SACK) Option for TCP. Internet draft draft-floyd-sack-00.txt, August 1999.

[25] M. Handley, S. Floyd, J. Padhye TCP Congestion Window Validation. draft-handley-tcp-cwv-00.txt, 10/20/1999.

[26] M. Handley, S. Floyd, B. Whetten, R. Kermode, and L. Vicisano. The Reliable Multicast Design Space for Bulk Data Transfer. draft-ietf-rmt-design-space-00.txt, 06/18/1999.

[27] M. Handley and S. Hanna. Multicast Address Allocation Protocol (AAP). draft-ietf-malloc-aap-02.txt, 10/22/1999.

28] M. Handley and C. Perkins. Guidelines for Writers of RTP Payload Format Specifications. draft-ietf-avt-rtp-format-guidelines-04.txt, 10/18/1999.

[29] M. Handley, D. Thaler, and R. Kermode. Multicast-Scope Zone Announcement Protocol (MZAP). draft-ietf-mboned-mzap-04.txt, 07/01/1999.

[30] M. Handley, B. Whetten, R. Kermode, S. Floyd, and L. Vicisano. The Reliable Multicast Design Space for Bulk Data Transfer. Internet draft draft-ietf-rmt-design-space-00.txt, June 1999.

[31] V. Paxson and M. Allman. Computing TCP's Retransmission Timer. Work in progress, Internet-Draft draft-paxson-tcp-rto-00.txt, October 1999.

[32] C. Perkins, M. Handley, and E. Whelan. Session Announcement Protocol. draft-ietf-mmusic-sap-v2-03.txt, 10/26/1999.

[33] K.K. Ramakrishnan, S. Floyd, and B. Davie. A Proposal to Incorporate ECN in MPLS. Internet-draft draft-mpls-ecn-00.txt, June, 1999.

[34] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, and V. Paxson. Simple Control Transmission Protocol. Work in progress, Internet-Draft draft-ietf-sigtran-sctp-03.txt, October 1999.

[35] B. Whetten, L. Vicisano, R. Kermode, M. Handley, S. Floyd, and M. Luby. Reliable Multicast Transport Building Blocks for One-to-Many Bulk-Data Transfer. draft-ietf-rmt-buildingblocks-01.txt.

[36] X. Xiao, E. Crabbe, A. Hannan, and V. Paxson. TCP Processing of the IP Precedence Field. Work in progress, Internet-Draft draft-xiao-tcp-prec-00.txt, September 1999.

**RFCs**

[37] M. Allman, V. Paxson, and W. Stevens. TCP Congestion Control. RFC 2581, Proposed Standard, April 1999.

[38] S. Floyd and T. Henderson. The NewReno Modification to TCP's Fast Recovery Algorithm. RFC 2582, Experimental, April 1999. This is a revised version of an internet-draft from December 1998.

[39] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. SIP: Session Initiation Protocol. RFC 2543, Proposed Standard, March 1999.

[40] J. Mahdavi and V. Paxson. IPPM Metrics for Measuring Connectivity. RFC 2678, Proposed Standard, September 1999.

[41] V. Paxson, M. Allman, S. Dawson, W. Fenner, J. Griner, I. Heavens, K. Lahey, J. Semke, and B. Volz. Known TCP Implementation Problems. RFC 2525, Informational, March 1999.

**Work in Progress**

[42] P. Barford, A. Feldmann, S. Jamin, V. Paxson, and C. Williamson. The state of the art of Internet measurement. To appear in Report of the Dagstuhl International Workshop on Internet Performance Modeling, 1999.

[43] S. Floyd. Validation Experiences with the NS Simulator. White paper, DARPA Workshop on Validation of Large Scale Network Simulation Models, April 1999

[44] M. Handley and J. Crowcroft. Internet Multicast Today. http://www.aciri.org/mjh/ipj-cast-1.ps

[45] M. Handley, S. Floyd, and B. Whetten. Strawman Specification for TCP Friendly (Reliable) Multicast Congestion Control. http://www.aciri.org/mjh/rmcc.ps (June 1999).

[46] Mark Handley and Vern Paxson. Network Intrusion Detection: Evasion, Traffic Normalization, and End-to-end Protocol Semantics. http://www.aciri.org/mjh/norm.ps (Aug 1999).

[47] S. Staniford-Chen, V. Paxson, and F. Wu. Internet Trap-and-Trace. Proposal submitted to DARPA, September 1999.

## 2 Speech Processing: Signals and Systems (the Realization Group)

The Realization Group [1]has focused for some time on a range of topics related to speech processing. The current goal is robust recognition and understanding of spoken language (as well as speaker detection) for realistic speaking styles and acoustic conditions. This goal requires a range of individual studies and projects that are grouped into four clusters below. In each case, a strong local program is significantly augmented by strategic collaborations outside of ICSI, both at UC Berkeley and elsewhere. The local efforts have been headed by group PIs Ellis, Greenberg, Hermansky, and Morgan, with significant contributions from students, postdocs, and visitors.

### 2.1 Signal Modeling

The group has a number of projects focused on processing and modeling speech signals in order to provide fundamental information that can be incorporated in systems for recognition of spoken language and other related goals.

- **Automatic Prosodic Stress Recognition:** Using acoustic features of the speech signal to label prosodic stress in the speech.

We are studying the prosodic properties of spontaneous speech, with a special focus on the Switchboard database. A team of local linguists and computer scientists previously worked to generate phonetic transcriptions for a significant subset of Switchboard. This effort was called the Switchboard Transcription Project, or STP. Currently, we are working on a new project to annotate spontaneous speech for stress, and to develop automatic or semi-automatic methods that will speed this process in the future. In 1999 we made significant progress in this area, using acoustic features of the speech signal to label prosodic stress in the speech [22]. We investigate the role of duration, amplitude, and fundamental frequency of syllabic vocalic nuclei for marking prosodic stress in spontaneous American English discourse. Local maxima of different evidence variables, implemented as combinations of the three basic parameters, duration, amplitude, and pitch, are supposed to be related with prosodic stress. As reference, two different subsets from the OGI English stories database were manually marked in terms of prosodic stress by two different trained linguists (see figure 1). The ROC curves, built on the training examples, show that both transcribers grant a major role to the amplitude and duration rather than to the pitch of the vocalic nuclei. More complex evidence variables, involving a product of the three basic parameters, allow around 80%primary stressed and 77% unstressed syllables to be correctly recognized in the test files of both transcribers' datasets. The agreement between the two transcribers on a set of common files supplies only slightly higher percentages. Such experiments have to be extended to new data sets and improved. New evidence variables are to be explored and a more accurate algorithm for the estimation of the fundamental frequency will be developed.

[1]The Realization Group was originally named for its focus on the Realization of Massively Parallel (typically connectionist) Systems. While maintaining an interest in designing and implementing large scale connectionist systems, the group objective has shifted in recent years to issues of signal processing and pattern recognition relevant for the realization of speech processing systems.

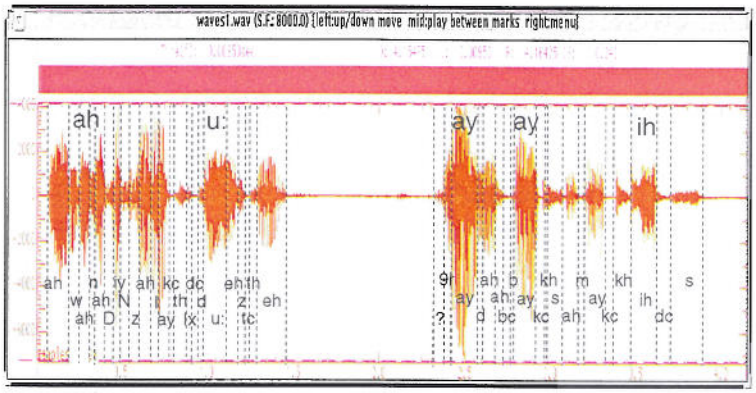



**Figure 1: Example hand transcription of stressed vocalic nuclei in "uh one of the things I like to do is ride is [pause] ride bikes with my kids".**

- **Automatic Phonetic Transcription:** Using spectral-temporal profiles of speech and articulatory features to phonetically transcribe a spontaneous speech corpus.

It has long been a conjecture in automatic speech recognition that substantial improvement of phonetic classification and segmentation could dramatically improve recognition performance. We are developing a system for automatic phonetic transcription based on (1) spectro-temporal profiles (STePs) of phonetic segments; (2) an intermediate representational tier derived from articulatory features; and (3) the temporal flow model (TFM), a form of neural network. Articulatory features offer a number of advantages over direct phone classification. In particular, such features are better at modeling coarticulation affects, dealing with degraded acoustic conditions, and speaker-dependent departures from canonical phonetic realizations. STePs are three-dimensional (time, frequency, and power) representation of average log critical band energy over large temporal spans, and their visualizations are color-coded by the variance of energy. The STePs representation is used to constrain and inform the architecture of our neural network models. A TFM is a neural network model that admits both traditional feedforward multi-layer networks and recurrent networks with lateral and feedback connections and variable propagation delays. The flexibility of TFMs provides special temporal and spectral processing ability particularly suitable for speech processing. Preliminary results show that most acoustic features are classified with an accuracy of 75-90%, and the most frequently occurring phones are classified with an accuracy of 70-90% giving a mean frame accuracy of ca. 83% [20].

- **Confidence Estimation:** Attaching a confidence score to each recognized word to indicate how likely it is that the word is correctly recognized.

Given the often errorful nature of automatic speech recognition (ASR) systems, it is generally desirable to equip recognizers with some sense of how likely it is that a recognized speech segment is correct. The automatic computation of such likelihoods, or "confidence measures," can greatly extend the robustness of ASR systems as well as the number of applications to which ASR can be applied [15]. One application where confidence measures may yield great benefit is in systems that have a low tolerance for recognition errors, (e.g., secure applications). However, confidence measures may be applied in several more far-reaching ways as well. For instance, accurate confidence measures may allow recognizers to bootstrap their own learning process by training to speech that has been automatically

transcribed with a high degree of confidence. This approach could substantially reduce the need for manual transcription of ASR training data--a costly and time-consuming task.
We are currently studying the estimation of word- and phone-level confidence measures in conversational speech. So far, our experiments have demonstrated the effectiveness of employing frame-level posterior phone probabilities derived from a neural network as a basis for determining phone-level and word-level confidence. However, we are still in the beginning stages of our model development and are investigating a number of approaches for improving our current system. One area that we are studying is the improved estimation of sub-word probabilities for the purposes of estimating word-level confidence. We also plan to conduct a detailed survey of potential sources of uncertainty in the recognition process.

- **Multi-Stream Features:** Generating multiple "snapshots" of an input utterance using multiple temporal resolutions and containing some new, novel, or robust type of information.

Automatic Speech Recognition (ASR) still poses a problem to researchers. In particular, most ASR systems have residual difficulties handling the differences between training and testing environments. A large number of modifications have been proposed and implemented in an effort to alleviate this. Although these modifications have resulted in increased levels of performance, ASR systems still fall short of human recognition ability in a large number of environments. In the past decade researchers have experimented with incorporating more speech information into the ASR process. At ICSI, we do this with a multi-stream system in which many feature representations of the speech signal are computed in parallel and many probability information streams are folded into the ASR decoding process. Such an approach facilitates the use of additional knowledge sources and can improve robustness to noise conditions.
Deciding the types of information streams and feature extraction methods to incorporate into the ASR system is not an obvious task. For robustness to various corrupting acoustic conditions, we have been experimenting with sets of condition-specific feature extraction methods. Current research involves adjusting our base signal processing to improve classification performance on each of several types of acoustic conditions. We employ discriminant analysis to obtain temporal modulation filters that are optimized for our set of subword units under a number of different acoustic conditions [21]. This method gives an array of task-specific basis filters. When placed in turn into the feature extraction process, we obtain an array of class probability estimate streams, each of which should be optimized for a particular type of acoustic corruption; for example, near-clean conditions, reverberant conditions, and additive noise. The probability streams provide some overlapping and complimentary evidence for the speech classes. When intelligently merged in the decoding, the performance of the ASR system will hopefully exhibit robustness when presented with speech from an unknown acoustic condition, assuming the unknown condition resembles at least one of the particular cases.

- **Acoustic Change Detection and Clustering:** Enabling adaptive speech models for recognition and providing speaker labels for speech/information retrieval.

Through the use of iterative clustering and segmentation procedures, we have reached 20% Equal Error Rate on three hours of speech from the Hub4 1997 evaluation data of Broadcast News. This is significantly better than other published results for this same data set. Also, the acoustic change detection subsystem has been optimized for speed, with recent

improvements reducing CPU time from about an hour to 2 minutes, using full covariance Gaussian models for the speaker change detection algorithm.

## 2.2 Recognition

In this category, there are a number of projects focused on the recognition process itself, incorporating both the signals and models described in the previous section and the necessary higher level models and methods for speech recognition, including adaptation and search.

- **Dynamic Pronunciation Models:** Changing probabilities of alternate pronunciations based on various factors, including word context, speaking rate, and the predictability of what is being said.

Accurate speaker-independent recognition of large-vocabulary speech by computers remains an unattained goal, particularly for spontaneous speech. High transcription error rate is caused in part by poor modeling of pronunciations within spontaneous speech. A PhD dissertation [6]examines how speaking rate and word predictability can be used to estimate when greater pronunciation variation can be expected; speaking rate and word predictability are also correlated with speech recognition errors. The results of these studies suggest that for spontaneous speech, it may be appropriate to build models for syllables and words that dynamically change the pronunciations used in the speech recognizer based on the context [1][7][8]. Implementation of new pronunciation models automatically derived from data within the ICSI speech recognition system has shown a 4-5% relative improvement in transcribing television and radio news reports and interviews. Roughly two thirds of these gains can be attributed to static baseform improvements; adding the ability to dynamically adjust pronunciations within the recognizer provides the remaining of the improvement. The corpus also allows for comparison of performance on different styles of speech: the new pronunciation models do not provide an improvement for speech categorized as pre-planned, but they provide a significant gain for spontaneous speech. Not only do the automatically learned pronunciation models capture some of the linguistic variation due to the speaking style, but the models also represent variation in the acoustic model due to channel effects. The largest improvement was seen in the telephone speech condition, in which 12% of the errors produced by the baseline system were corrected [6][9].

- **Statistical Language Modeling:** Learning topic based language models using an EM model of word co-occurrences.

Most speech recognition systems combine acoustic information with a statistical language model that assigns prior probabilities to strings of words. Conventional n-gram language models calculate statistics based on the previous few words recognized; this project aims to incorporate longer-distance information within language modeling.

A probabilistic model is trained on the co-occurrences of words in a collection of documents. Hidden sources, which are intended to model document topics, are assigned to each word occurrence using an EM-based clustering algorithm. This can be thought of as a probabilistic version of the technique of latent semantic analysis used in information retrieval. Using this method to model language generation has been shown to substantially reduce perplexity on the TDT1 corpus of broadcast news stories [11].

- **Multi-Stream Combination:** Combining information generated from multiple experts trained on multi-stream features to improve word recognition and increase robustness to corrupting environmental conditions.

Multi-band speech recognition is an approach in which the input speech is divided into disjoint frequency bands and treated as separate sources of information. They can then be merged to determine the most likely words spoken. One potential advantage of the multi-band approach is the ability to combine sub-band information asynchronously. Several researchers have found that phone transitions can occur at different times in different frequency bands. For instance, inspecting the spectrogram of the phone transition /jh/ to /ow/ in figure 2, you will notice that the transition occurs later in the second formant than in the first formant. If a multi-band speech recognizer allows phone transitions to occur in an asynchronous manner, then it may be better able to model the correct utterance.

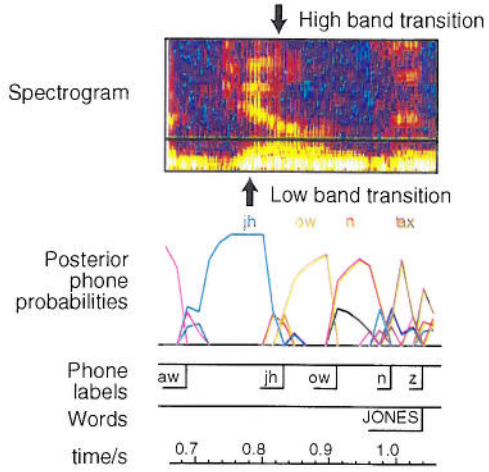



**Figure 2: Spectrogram showing transition asynchrony between /jh/ and /ow/, along with acoustic model and recognizer outputs.**

Our work has shown, however, that complete relaxation of synchrony constraints does not consistently reduce the word error rate [16]. Rather, the optimal permissible asynchrony should reflect the asynchrony in the training data. We are currently experimenting with the constraints on asynchrony by building multi-band word models from training data broken into two sub-frequency bands. These word models explicitly model the average duration of asynchronous states (states in which sub-frequency bands predict different phones) and transition probabilities between asynchronous and synchronous states. Our current experiments explore the performance of these asynchronous word models under various training and testing conditions.

Aside from this work on multi-band recognition, we also have applied multi-stream approaches to the combination of feature streams with differing temporal properties, such as RASTA-PLP (designed at ICSI in the early 90's) and Modulation-filtered spectrograms (MSG, designed at ICSI in the late 90's and described in detail in Brian Kingsbury's 1998 thesis). As noted in [13], this resulted in significant improvements for our Broadcast News recognition system [4].

Finally, multi-stream combination methods were employed for an international evaluation on noise robustness in speech recognition, coordinated by the European Telecommunications Standards Institute (ETSI). The "Aurora" evaluation is oriented

towards defining a standard feature set to be computed within cellular telephones (for ultimate recognition at distributed servers).

In the Aurora task, we tested two systems built up from approaches developed in this project:

1. Discriminant filter design described earlier. Other key features of this system were decimation that made possible by the constrained modulation spectrum, dimensionality reduction via PCA, and a nonlinear transformation of the feature vector using a neural network trained for phonetic discrimination. While the outputs of the net approximate posterior class probabilities, the pre-nonlinearity (softmax) outputs are more like log probabilities and have an approximately Gaussian distribution. These values are then transformed to have a diagonal covariance matrix by using the Karhunen-Loeve (KL) transform.

2. Combined design including the above features, plus PLP (designed in the mid-80's by Hermansky), MSG (designed at ICSI in the 90's), and a multi-band system using critical bandwidths and 1 second temporal windows, all combined in a merging neural network. System (a) was the simpler one (computationally), and reduced word error rate over the range of noise conditions in the task by 40% (less than a mel cepstral system). System (b) reduced the word error rate by 60%. The best system from a competing lab reduced the error rate by 42%. Thus, our efforts in this project led to the best results in the world on this task.

- **Speaker Adaptation:**Improving recognition accuracy by adapting Artificial Neural Networks to particular speakers.

One approach to speaker adaptation for our neural-network acoustic models is to adapt a speaker-independent network by performing a small amount of additional training. By using data only from the target speaker, the result is an acoustic model specifically tuned to that speaker. This adapted model might be useful for *speaker recognition* too, especially since state-of-the-art speaker recognition typically performs a speech-recognition labelling of the input speech as a first stage. However, in order to exploit the discriminant nature of the neural nets, it is better to train a single model to discriminate both between the different phone classes (as in conventional speech recognition)*and* between the target speaker and the 'rest of the world' (a common approach to speaker recognition). We investigated this approach for a set of 12 speakers selected from the DARPA/NIST Broadcast News corpus. The speaker-adapted nets showed a 17% relative improvement in word-error rate on their target speakers, and were able to identify among the 12 speakers with an average equal-error rate of 6.6%.

- **Word Hypothesis Search Algorithms:**Developing word hypothesis search algorithms that run efficiently on vector architectures, especially for hand-held applications.

The typical large vocabulary decoder consists of a highly pruned (and therefore irregular) search through all possible utterances. The primary focus of our current efforts is to vectorize this portion of the speech system, to allow efficient execution on next-generation vector microprocessors. Among the optimizations we are investigating is the efficient generation of word recognition hypotheses. Our first step has been to group words of the same length together to form regular vectorizable computations. We are also examining ways to vectorize the traditional methods of reducing the search space, including search beam-widths and least-upper-bound path calculations.

## 2.3 Perception and Understanding

This category consists of a number of projects which take the human interpretation of spoken language as a central concern.

- **Semantic Role Assignment:** Automatically assigning semantic roles to sentence constituents based on a data-driven approach.

This project aims to extend statistical parsing techniques for natural language to produce analyses of text that give not only syntactic structure, but also semantic roles for various constituents of a sentence. The system may be used as a first step toward a deeper semantic analysis, or directly in applications such as statistical machine translation, language modeling, or information retrieval. We are using the results of the Berkeley FrameNet project, which has defined numerous semantic "frames" and their semantic roles, or "frame elements," and annotated a corpus of text by labeling instances of frame elements in context.

- **Speech Perception:** Assessing spectral-temporal characteristics of the human auditory system by means of psycho-acoustic experiments.

Two separate perceptual studies have recently been completed and others are in the process of being conducted.

**1.** As a function of spectrally desynchronized information.

The spectrum of spoken sentences was partitioned into quarter-octave channels and the onset of each channel shifted in time relative to the others so as to desynchronize spectral information across the frequency axis. Human listeners are remarkably tolerant of cross-channel spectral asynchrony induced in this fashion. Speech intelligibility remains relatively unimpaired until the average asynchrony spans three or more phonetic segments. Such perceptual robustness is correlated with the magnitude of the low-frequency (3-6 Hz) modulation spectrum and thus highlights the importance of syllabic segmentation and analysis for robust processing of spoken language. High-frequency channels (>1.5 kHz) play a particularly important role when the spectral asynchrony is sufficiently large as to significantly reduce the power in the low-frequency modulation spectrum (analogous to acoustic reverberation) and may thereby account for the deterioration of speech intelligibility among the hearing impaired under conditions of acoustic interference (such as background noise and reverberation) characteristic of the real world.

**2.** Sparse spectral information provides nearly perfect intelligibility

Traditional models of speech assume that a detailed auditory analysis of the short-term acoustic spectrum is essential for understanding spoken language. The validity of this assumption was tested by partitioning the spectrum of spoken sentences into 1/3-octave channels ("slits") and measuring the intelligibility associated with each channel presented alone and in concert with the others. Four spectral channels, distributed over the speech-audio range (as illustrated in figure 3) are sufficient for human listeners to decode sentences with nearly 90% accuracy although more than 70%of the spectrum is missing. Word recognition often remains relatively high (60-83%) when just two or three channels are presented concurrently, despite the fact that the intelligibility of these same slits presented in isolation is less than 9%. Such data suggests that the intelligibility of spoken language is derived from a compound "image" of the modulation spectrum distributed across the frequency spectrum. Because intelligibility seriously degrades when slits are desynchronized

by more than 25 ms this compound image is probably derived from both the amplitude and phase components of the modulation spectrum, and implies that listeners' sensitivity to the modulation phase is generally "masked" by the redundancy contained in full-spectrum speech.[23]

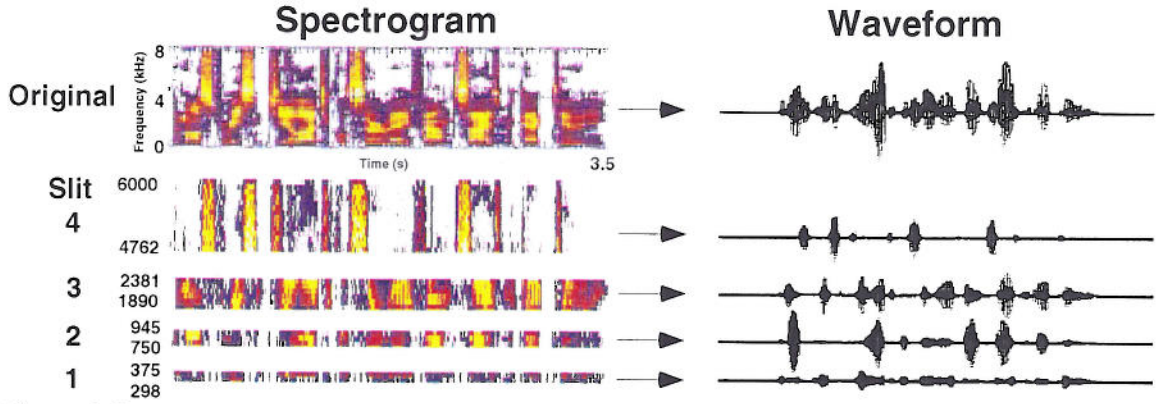



**Figure 3: Illustration of the reduction of full-band speech to four narrow spectral slits.**

## 2.4 Applications

In the Realization Group it has always been important to design complete prototype systems in order to focus research attention on the most relevant problems. Currently we are focusing on the design of several systems for information extraction (IE) and information retrieval (IR). Although the IE and IR components per se are not the major focus of our research, having complete systems that include such components provide us with greater insight into the relevance and efficacy of our design decisions for spoken language processing.

- **Meeting Recorder:** Designing, implementing, and testing a portable digital speech recorder for real-time recording of impromptu meetings.

Automatic speech recognition (ASR) provides a natural interface to small form-factor computers (such as PDAs) since keyboards and large displays are absent on these platforms. However, large vocabulary, robust ASR requires hardware resources far beyond those available on current PDAs. The Vector IRAM processor provides a partial solution by delivering very high performance for relatively little expenditure of power. However, for speech recognition to take advantage of IRAM, the components of the system must be redesigned with vectorization in mind.

We are currently adapting the workstation based ASR system used at ICSI to run efficiently on IRAM. Two out of the three major components of ICSI's speech system, the acoustic front-end and the phoneme probability estimator, contain computational kernels that vectorize easily (FFT and matrix-matrix multiply, respectively).

We have begun recording a corpus of real meetings held within the group to act as training material and other evidence for this task.

- **Speech Enabled Information Systems:** Building speech-enabled kiosks, desk tablets, and personal data assistants to enable users to find and display current information, as partners in the German consortium, SmartKom.

The SmartKom Public scenario is the component of SmartKom that is most appropriate for this kind of effort. In this scenario, users will interact with a system to retrieve information about local events such as film listings. However, since there was no existing system to support the collection of English data for the SmartKom Public scenarios, we are bootstrapping our effort by collecting some general English speech at ICSI; once our baseline modules are working, we will use this system to collect data that is specific to the theater information scenario of SmartKom Public. We hope to leverage the wizard-of-oz work at LMU (Munich) to speed collection of English SmartKom data.

Since the inception of the project, we have finished construction of an audio-recording conference room; the equipment for this room was acquired under US project funding. SmartKom Public will require recognizers to be able to handle far-field microphone input, since most users will not be likely to be using a headset. The conference room supports multiple channels of sound input; this will allow us to collect data similar to the German SmartKom data collected by the microphone array at LMU. The current setup will allow simultaneous (low skew) recording with several types of microphones, including head-mounted, lapel, and table-mounted far-field microphones.

In addition to the assembling the hardware necessary for recording spoken language data, our staff has been developing software to facilitate collecting multi-channel data. Pilot recordings of general English data are now underway.

Our current line of acoustic work in this project is another facet of the combination problem: assuming that we keep the model of probability combination at 10 millisecond frames, what is the appropriate combination function? Traditionally, we have found good performance from multiplying the probability streams together, and then renormalizing so that the distribution sums to 1. However, combining with a multi-layer perceptron often works much better, although for a large number of input sources the input space of the neural network becomes unwieldy. Oracle experiments suggest that if we could choose the best combination of classifiers dynamically, we can more than halve the error rate of our recognizer on a spoken numbers task. Our current efforts are dedicated to finding a way to duplicate the oracle experiment without cheating. We have found that the Kullback-Liebler distance between the subbands correlates well with the combination of classifiers that the oracle selects; we hope to use this measure to improve our probability combination techniques. The statistical techniques we are developing are general enough that we hope that they will carry over to general multi-stream probability combinations, including fusing information from multiple spectral representations and multiple microphones.

The language modeling work for this project is primarily discussed in the AI Group component of this report. The principal result from their effort for the SmartKom project in 1999 is the completion of a prototype system that performs the analysis portion of language understanding. The system integrates diverse information --- including grammatical cues, ontological world knowledge and features of the current discourse and situational context in producing an appropriately parameterized simulation specification. To support such an integrated approach, we have had to develop richly fine-grained semantic representations, not only in our ontology and in the final output simulation specification, but crucially also in the representations of linguistic knowledge itself.

- **Information Retrieval for Audio Documents:** Using transcriptions automatically generated by our large-vocabulary continuous speech recognizer to build indexes that can be queried by standard information retrieval engines for searchable audio archives.
In this effort, speech from audio data sets (such as the American Broadcast News corpus and a similar collection from the BBC) is roughly transcribed and stored for later searching. A spoken query system is then used to search for plausible sections of the audio, using information retrieval techniques; see figure 4 for the screen image of this system. This work has largely been supported by our participation in a European Union research project.

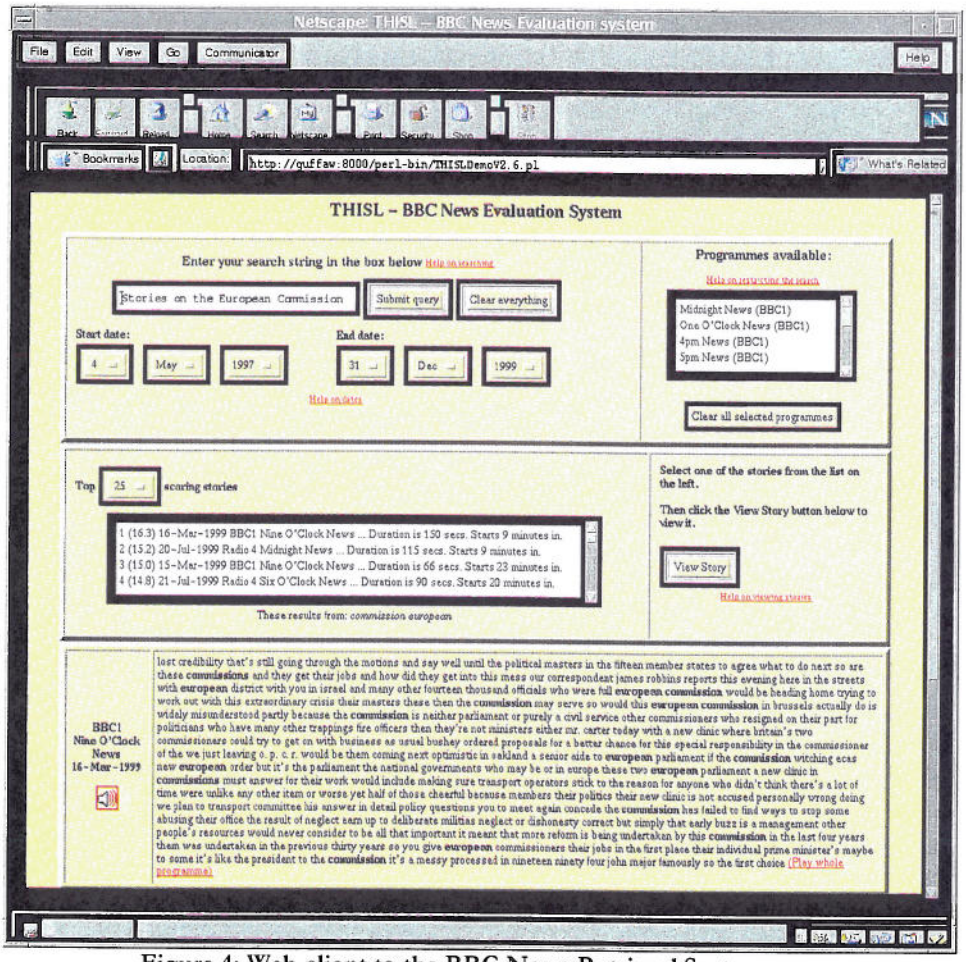



**Figure 4: Web client to the BBC News Retrieval System.**

Aside from providing a tangible application for our recognition methods, this work also exercised a number of other techniques such as the separation of different sound sources, such as speech and music [24]. In particular we investigated the use of Computational Auditory Scene Analysis to separately index individual sound sources present in archive recordings.

**Publications appeared and in press during 1999**

**References**


[1] Alan Bell, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand and Daniel Gildea Forms of English function words - Effects of disfluencies, turn position, age and sex, and predictability (Presented at Int. Cong. of Phonetic Sciences, San Francisco, August 1999, pp. 1:395-398)

[2] Bilmes, J., "Natural Statistical Models for Automatic Speech Recognition," PhD Dissertation, University of California at Berkeley, May 1999.

[3] Bilmes, J., "Buried Markov models for speech recognition,"*Proc. ICASSP-99*, Phoenix, pp. II-713-716.

[4] Cook, G., Christie, J., Ellis, D., Fosler-Lussier, E., Gotoh, Y., Kingsbury, B., Morgan, N., Renals, S., Robinson, A.J., and Williams, G., "An Overview of the SPRACH System for the Transcription of Broadcast News,"*Proc. DARPA Broadcast News Transcription and Understanding Workshop,* Herndon VA, 1999.

[5] Ellis, D., and Morgan, N., "Size matters: An empirical study of neural network training for large vocabulary continuous speech recognition,"*Proc. ICASSP-99*, Phoenix, pp. II-1013-1016.

[6] Fosler-Lussier, E., "Dynamic Pronunciation Models for Automatic Speech Recognition," PhD Dissertation, University of California at Berkeley, May 1999. Reprinted as ICSI Technical Report No. tr-99-015, 1999.

[7] Fosler-Lussier, E., "Contextual word and syllable pronunciation models," ASRU-99, Keystone CO, December 1999.

[8] Fosler-Lussier, E., "Multi-Level Decision Trees for Static and Dynamic Pronunciation Models," Eurospeech-99, Budapest, pp. I-463-466, 1999.

[9] Fosler-Lussier, E., and Williams, G., "Not just what, but also when: Guided automatic pronunciation modeling for Broadcast News,"*Proc. DARPA Broadcast News Transcription and Understanding Workshop,* Herndon VA, 1999.

[10] Genoud, D., Ellis, D., and Morgan, N., "Combined speech and speaker recognition with speaker-adapted connectionist models," IEEE Workshop on Automatic Speech Recognition and Understanding, pp 177-180, Keystone, CO, Dec. 1999.

[11] Gildea, D. and Hofmann, T., "Topic-based language models using EM,"*Proc. Eurospeech,* Budapest, pp. V-2167-2170, 1999.

[12] Gold, B., and Morgan, N.,*Speech and Audio Signal Processing*Wiley Press, New York, 1999.

[13] Janin, A., Ellis, D., and Morgan, N., "Multi-stream speech recognition: Ready for prime time?"*Proc. Eurospeech,* Budapest, pp. II-591-594.

[14] Kirchhoff, K., and Bilmes, J. "Statistical Acoustic Indications of Coarticulation"*Proc. Int. Cong. of Phonetic Sciences,* San Francisco, August 1999, pp. 3:1729-1732/

[15] Kirchhoff, K., and Bilmes, J. "Dynamic classifier combinations in hybrid speech recognition systems using utterance-level confidence values,"*Proc. ICASSP-99*, Phoenix, pp. II-693-696.

[16] Mirghafori, N., and Morgan, N., "Sooner or Later: Exploring Asynchrony in Multi-Band Speech Recognition,"*Proc. Eurospeech*, Budapest, pp. II-595-598, 1999.

[17] Morgan, N., Ellis, D., Fosler-Lussier, E., Janin, A., and Kingsbury, B., "Reducing errors by increasing the error rate: MLP Acoustic Modeling for Broadcast News Transcription,"*Proc. DARPA Broadcast News Transcription and Understanding Workshop,* Herndon VA, 1999.

[18] Morgan, N., "Temporal Signal Processing for ASR," IEEE Workshop on Automatic Speech Recognition and Understanding, pp 9-16, Keystone, CO, Dec. 1999.

[19] Schwenk, H., "Using Boosting to Improve a Hybrid HMM/Neural Network Speech Recognizer,"*Proc. ICASSP-99,* Phoenix, pp. II-1009-1012.

[20] Shastri, L., Chang, C., and Greenberg, S., "Syllable Detection and Segmentation Using Temporal Flow Neural Networks *Proc. Int. Cong. of Phonetic Sciences,* San Francisco, August 1999, pp. 3:1721-1724.

[21] Shire, M., "Data-driven modulation filter design under adverse acoustic conditions and using phonetic and syllabic units,"*Proc. Eurospeech* Budapest, Hungary, pp. III-1123-1126.

[22] Silipo, R., and Greenberg, S., "Automatic Transcription of Prosodic Stress for Spontaneous English Discourse," Int. Cong. of Phonetic Sciences, San Francisco, August 1999, pp. 3:2351-2354

[23] Silipo, R., Greenberg, S., and Arai, T., "Temporal Constraints on Speech Intelligibility as Deduced from Exceedingly Sparse Spectral Representations,"*Proc. Eurospeech* Budapest, Hungary, pp. VI-2687-2690, September 1999.

[24] Williams, G., and Ellis, D., "Speech/music discrimination based on posterior probability features ,"*Proc. Eurospeech,* Budapest, pp. II-687-690.

## 3 Artificial Intelligence and its Applications

The Artificial Intelligence group continues its long term study of language, learning, and connectionist neural modeling. The scientific goal of this effort is to understand how people learn and use language. The practical goal is to develop systems that support human centered computing through natural language and other intelligent systems. There continues to be close cooperation with other groups at ICSI, at UC Berkeley and with external sponsors and other partners. There are three articulating subgroups and this report summarizes their work.

### 3.1 Learning and Language Use

It has long been known that people would prefer to talk with computer systems in natural language if they could. The problem of communicating with machines is becoming increasingly important to society because computers will soon be embedded in nearly every artifact in our environment. But how easy will it be for people of all ages and abilities to use them? In ten years or less, virtually every device in our environment will have a computer in it. This raises the specter of an embedded computing malaise---every device will have its own interface that the user has to learn. Most people today cannot program their VCRs or use more than half the functionality of their answering machines. In the world of embedded computing, there could be thousands of idiosyncratic interfaces to learn. Many people will not be in control of the devices in their own environments. In addition, ever widening aspects of society, from education to employment depend upon everyone interacting with computational systems.

Natural interaction with computerized devices and systems requires a conceptual framework that can communicate about requests specified in ordinary language. Systems may well need to *tell* their human users what is going on, *ask* for their advice about what to do, *suggest* possible courses of action, and so on. The machines, and more especially the interactions among the machines, are getting to be so complicated and autonomous, and yet so intimately involved in the lives of the human users, that they (the machines) have to be able to take part in a kind of social life. The central goal of this project is to provide a conceptual basis and a linguistic framework that is rich enough to support a natural mode of communication for this evolving human/machine society.

While the usefulness of Natural Language Usage (NLU)systems has never been questioned, there have been mixed opinions about their feasibility. Most current research is focused on goals that are valuable, but fall far short of what is needed for the natural interactions outlined above. We believe that recent advances in several areas of linguistics and computational theory and practice now allow for the construction of programs that will allow robust and flexible integrated language interaction(ILI) within restricted domains.

For many years, Jerome Feldman has studied various connectionist computational models of conceptual memory and of language learning and use. George Lakoff and Eve Sweetser have worked on the relation between linguistic form, conceptual meaning, and embodied experience. Over the past dozen years, the group has explored biologically plausible models of early language learning (Bailey et al. 97) and of embodied metaphorical reasoning (Narayan 97). About two years ago, we extended our efforts on modeling child language acquisition from individual words and

phrases (Regier 96, Bailey 97) to complete utterances. This required us to develop a formal notion of what it means to learn the relationship between form and meaning for complete sentences. Many groups, including ours (Feldman 98) have worked on algorithms for learning abstract syntax, but we decided that it was time to look directly at learning form-meaning pairs, generally known as *constructions*. After an intensive effort by the whole research group, we now have an adequate formalization of constructions and are moving ahead with the project of modeling how children learn grammar from experience. We also realize that our formalized notion of linguistic constructions that systematically links form to conceptual meaning is potentially a breakthrough in achieving robust and flexible NLU systems.

The most novel computational feature of the NTL effort is the representation of actions: executing schemas (x-schemas), so named to remind us that they are intended to execute when invoked. We represent x-schemas using an extension of a computational formalism known as Petri nets (Murata, 1989). As discussed below, x-schemas cleanly capture sequentiality, concurrency and event-based asynchronous control. With our extensions they also model the hierarchy and parameterization needed for action semantics.

Our goal is to demonstrate that unifying two powerful linguistic theories, embodied semantics and construction grammar, together with powerful computational techniques, can provide a qualitative improvement in HCI based on NLU. Over the next period we will extend our existing pilot system to moderate-sized applications in real HCI settings and develop the methodology needed for large-scale realization of NLU interaction. This involves formalization and additional research in cognitive linguistics, development of probabilistic best fit algorithms, and significant system integration. Much of the group's effort over the past year has gone into developing these formalisms and to producing a pilot version of the integrated language understanding system(Bergen & Chang 99).

For concreteness, we have chosen a specific task domain for the proof-of-concept demonstration of our research. We will construct a system for understanding and responding to dialog with tourists, initially focused on Heidelberg, Germany. This applied project is being carried out in cooperation with a partner group at the European Media Lab (EML) in Heidelberg, which has built an extensive data base describing their city (www.villa-bosch.de/english/research) and will implement the detailed actions for using it based on our natural language analysis. This cooperation will bring several benefits to the project and provides clear milestones for evaluating our effort.

The core computational question is finding best match of constructions to an utterance in linguistic and conceptual context. One of the attractions of traditional phrase structure grammars is the fact that the time to analyze (parse) a sentence is cubic in the size of the input. If one looks at the comparable problem for our more general construction grammars, context-free parsing becomes NP complete (~exponential) in the size of the input sentence and thus impractical. But people do use larger constructions to analyze language and we believe that we have two insights that seem to render the problem of construction analysis tractable. The general computational point is that our task of finding a best-fit analysis and approximate answers that are not always correct presents a more tractable domain than exact symbolic matching. More importantly, our integrated constructions are decidedly not context-free or purely syntactic. We believe that constraints from both semantics and context will be sufficiently constraining that it will be possible in practice to build best-fit construction matchers of the required scale.

This sequence of operations: surface analysis, construction parse, SimSpec, simulation and inference is repeated for every clause. The current pilot system does not make use of extensive context or

world knowledge, but these are central to our new design. There is currently a great deal of renewed effort to develop ontologies of words and concepts for a wide range of semantic domains (Fikes 1994). After analyzing these efforts, we have decided against committing to any one of the competing formulations and have instead defined an Application Programming Interface (API) that our system can use to access information from any source. A preliminary version of this is used in the pilot system and we will evolve the API as experience requires. The current API has the usual commands for adding information and some special ones for retrieving ordered lists of concepts most likely to fulfill a request.

Our HCI system also requires both situational and discourse context as well as general knowledge. Despite a large literature on context (Hobbs 79, Kehler 93), there is currently no integrated theory or system that meets our needs. Again, we have built an API (which returns ordered lists of potential role fillers) for the pilot system and this remains an active research area. We will also employ one of modern large scale parsers (Charniak 98, Collins 98). Although much needs to be done, we believe that we have identified all the main components of an ILI system and have at least preliminary versions of each subsystem operational and a sound overall design plan.

Thus the NTL group has, over the last year, formalized and significantly extended its work on language learning and use based on deep conceptual semantics. Both the learning sub-task and the performance HCI system are moving ahead in collaboration with other efforts at ICSI and elsewhere.

### 3.2 FrameNet Project

As of February, 2000, the NSF-sponsored FrameNet project (NSF IRI-9618838 - "Tools for Lexicon Building") is continuing to build the FrameNet Database, a lexicon which will contain about 2000 words (to be made available on the World Wide Web by May, 2000), providing (1) a collection of semantically annotated examples for each sense of each word, (2) links to descriptions of the conceptual structures ("frames") which underlie each such sense, (3) details of the syntactic ways in which the semantic roles ("frame elements") in each such conceptual structure are grammatically realized in sentences containing the word (in a table of Frame Element Realizations) and (4) records of the combinations of frame elements and their syntactic realizations expressed in the sample sentences (in a table of Valence Patterns).

The project makes use of the 100M word British National Corpus (provided by Oxford University Press), corpus-management tools developed in IMS-Stuttgart, MITRE-Corp annotation software which allows the introduction of XML tags into the sentences forming the examples database, and various in-house devices for viewing the database. Currently a MySQL-based web interface is used for making the database viewable and searchable by human users.

The manual analysis and annotation activities of the project are extremely labor-intensive, but it is hoped that the resulting data can be used as a training corpus for future automatic operations, these designed to annotate sentences that have not been manually annotated, for frames that have been analyzed, and to discover phrases that are most likely to have semantic relations to given target words, based on syntactic parsing of corpus sentences, in the case of frames that have not been analyzed.

The project is nearing the end of its third year. Although new funding has been assured from NSF for the next phase of Frame Net, various efforts are still under way to gain further funding;

proposals are in to various funding sources and new proposals are being written. Future work is projected to move in various directions: (1) greatly increasing the scope and coverage of the current sample of the English vocabulary, (2) concentrating on automating various steps in the analysis and processing of corpus material, (3) demonstrating applications of FrameNet methods and results to the work of word sense disambiguation, language understanding, and mechanical translation, and (4) fitting FrameNet lexical descriptions into a precise theory of grammatical constructions.

A bilingual spin-off of FrameNet work, sponsored by the German American Academic Council (through the TRANSCOOP initiative), is in its final stages and its findings will be presented shortly. The function of this effort (directed by Ulrich Heid of IMS-Stuttgart) will be to demonstrate the feasibility of a frame-semantics-based bilingual dictionary.

As the final entries are being completed, they are sent to Dan Jurafsky and his collaborators at the University of Colorado to determine relative frequencies of senses and valence patterns in the larger corpus, and to Daniel Gildea (U. C. Computer Science graduate student at ICSI) for his research on automating the search for syntactic constituents that serve as semantic dependents of lexical items.

Various efforts are under way to provide representations of phenomena and relationships discovered in the course of carrying out the FrameNet work for which the project was not initially prepared. Much of this is spear-headed by Srini Narayanan at SRI-International, co-PI of the next phase of FrameNet. Using formalisms being developed by Mark Paskin and Nancy Chang of ICSI (U.C. CS graduate students), these are aimed at providing the means of using FrameNet data in an inference-making context, building in relations of inheritance, blending, context-licensed omissions, and various kinds of frame complexities.

### 3.3 Connectionist Modeling

Lokendra Shastri's work on computational modeling has spanned three different representational and processing tiers of language processing. One tier focuses on high-level reasoning underlying language understanding. The second tier focuses on the formation of episodic memory whereby transient patterns of neural activity representing events and situations are rapidly transformed into persistent neural circuits (memory traces) capable of supporting recognition and recall. The third modeling effort concerns the extraction of syllabic segments from spontaneous and noisy speech. The results of the three efforts are summarized below.

**3.3.1 A neurally motivated model of rapid inference.** We are capable of drawing a variety of inferences effortlessly and with remarkable efficiency. To wit, our ability to understand language in real-time is a task that requires the hearer to draw a number of inferences in order to establish referential and causal coherence, generate expectations, make predictions, and recognize the speaker's intent. The rapid rate of language understanding suggests that we are capable of performing a wide range of inferences rapidly, spontaneously, and without conscious effort --- as though they are a *reflex* response of our cognitive apparatus. In view of this, such reasoning may be described as *reflexive* reasoning (Shastri & Ajjanagadde, 1993). This remarkable human ability poses a challenge for cognitive science and computational neuroscience: How can a system of simple and slow neuron-like elements represent a large body of systematic knowledge and perform a wide range of inferences with such speed?

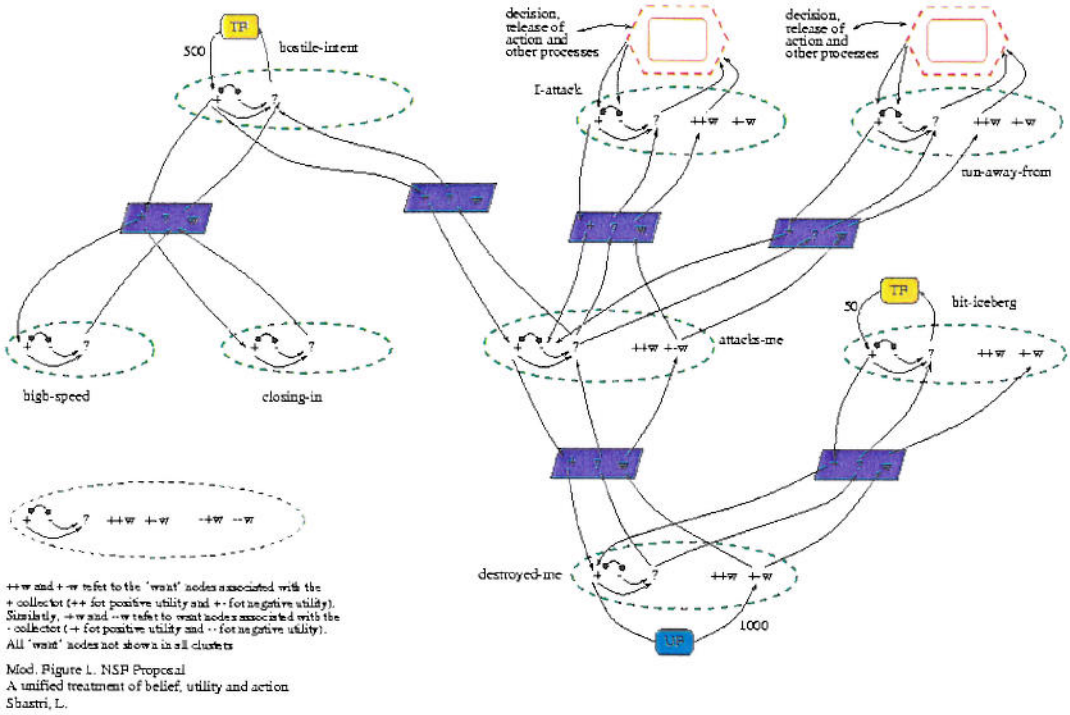
SHRUTI is a neurally motivated model of knowledge representation and inference. It demonstrates how a network of neuron-like elements can encode a large body of evidential knowledge and yet perform a variety of explanatory and predictive inferences within a few hundred milliseconds.

SHRUTI can encode first-order evidential rules involving n- *ary* relations (predicates), variables, and quantification. It can also encode specific facts as well as context-sensitive priors over event types. It expresses dynamic bindings between roles and fillers via the synchronous firing of appropriate node clusters and performs inferences via the propagation of rhythmic activity over node clusters. This propagation amounts to a parallel breadth first activation of the underlying inference graph, and hence, the reasoning in SHRUTI is extremely fast.

SHRUTI is fully implemented to execute on any Solaris and Linux platforms. This implementation has been used to model tactical decision making under stress as part of a research project funded by the ONR and as the inferential engine underlying a research agent for inferential retrieval.

Parallel processor implementations for an earlier version of SHRUTI have been studied extensively on the CM-5 platform (Mani, 1995). The resulting system could encode knowledge bases with over 500,000 randomly generated rules and facts, and respond to queries of inferential depth five in under 250 milliseconds. These performance results (circa 1995) were based on technology that is now obsolete, and much better results can be expected using today's hardware. We have sound technical reasons to believe that the results of this effort will transfer well to a mapping of SHRUTI to a network of workstations, Beowulf cluster, or an array of StrongARM processors.

The representational and inferential abilities of SHRUTI, a structured connectionist model of reflexive inference were extended in the several ways to enable it to draw inferences required for establishing referential and causal coherence. SHRUTI's encoding of rules was extended to support the simultaneous propagation of activation in the forward (predictive) as well as the backward (explanatory) direction. Its representation of rules and types was extended to support the dynamic instantiation of entities during inference, as well as the dynamic unification of entities (via synchronization of node firing). A family of evidence combination functions was derived to allow the coding of probabilistic, evidential, and fuzzy knowledge, the behavior of links and nodes was augmented to exhibit priming effects for entities, types, facts, and rules. Finally it was shown that under a suitable interpretation of link weights, the inferential behavior of SHRUTI can be related to probabilistic inference. All of the above extensions were incorporated into the SHRUTI simulator. An Application Programmer Interface (API) for the SHRUTI simulator was developed to enable its integration with other models. Preliminary work was completed on extending SHRUTI to encode beliefs as well as utilities. The research on SHRUTI described above was done with graduate student Carter Wendelken. The results of the above research are described in (Shastri, 1999a, 2000a; Shastri & Wendelken, 1999; Wendelken & Shastri, 2000).




An example SHRUTI network. Focal-clusters encoding relations are shown in green, mediator-clusters linking antecedent and consequent relations are shown in blue, action schemas are shown in red, taxon facts encoding statistical priors are depicted in yellow, and utility facts encoding values associated with world states are encoded in cyan. The network fragment encodes knowledge from a simplified naval domain. This knowledge can be paraphrased as follows: Something closing in at high-speed may have hostile intent; something having hostile intent may attack me (i.e., my ship) and destroy me; hitting an ice-berg may destroy me; a counter-attack may foil an attack; running away from an attack may prevent an attack; there is a moderately high prior of being attacked; and getting destroyed has extremely negative utility. Note that beliefs and utility propagate along the same structure. Given a set of observations, spreading activation over the network structure enables SHRUTI to draw explanations, make predictions, and identify actions that would tend to maximize expected utility.

### 3.3.2 Biological grounding of recruitment learning and modeling episodic memory formation

Biological neural networks are capable of gradual learning based on observing a large number of exemplars over time as well as rapidly memorizing specific events as a result of a single exposure. The primary focus of research in connectionist modeling has been on gradual learning, but some researchers have also attempted the computational modeling of rapid (one-shot) learning within a framework described variably as *recruitment learning* and *vicinal algorithms*. While general arguments for the neural plausibility of recruitment learning and vicinal algorithms based on notions of neural plasticity have been presented in the past, a specific neural correlate of such learning has not been proposed. Here it is shown that recruitment learning and vicinal algorithms can be firmly grounded in the biological phenomena of long-term potentiation (LTP) and long-term depression (LTD). Toward this end, a computational abstraction of LTP and LTD was developed and "algorithms" for the recruitment of binding-detector cells and binding-detector circuits were developed and evaluated using biologically realistic data. It was shown that binding-detector and binding-error detector cells of distinct bindings exhibit low levels of cross-talk even when the bindings overlap. Given this grounding, the specification of a vicinal algorithm amounts to specifying an appropriate network architecture and suitable parameter values for the induction of LTP and LTD.

A detailed quantitative analysis of SMRITI, a model of episodic memory formation in the hippocampal system, was carried out. The analysis made use of realistic values of the number of cells in various regions and the density of projections between regions. The results indicate that the model has a large memory capacity (about 75,000 events) and is robust against diffuse cell loss.

### 3.3.3 Temporal flow models for speech processing

The syllable provides an intermediate representation capable of binding the phonetic and phonological tiers with the lexical and grammatical levels of spoken language. Work on the processing of spontaneous speech has focused on the development of Temporal Flow neural network models (TFMs) of Watrous and Shastri, for the automatic segmentation of the acoustic stream into syllabic units. TFMs admit feedforward, lateral, as well as recurrent connections, and support variable propagation delays along links. These models achieved an accuracy of ca. 85% (Shastri, Chang, and Greenberg, 1999). TFM networks have also been developed for recognizing articulatory features (AF) with between 60-90% accuracy. These models have been used in turn to drive networks for accurately recognizing phonetic segments. The recognition appears to be highly resilient to the effects of background noise and reverberation. The underlying basis for the recognition of articulatory features and phonetic segments is the use of modulation spectral features across the frequency spectrum. The networks are currently being trained to automatically segment the three hours of the Switchboard Transcription Corpus that has not been hand-segmented into phonetic elements. This work was carried out with Steven Greenberg and graduate student Shawn Chang.

## References


[1] Jagota, A,, Plate, T., Shastri, L., and Sun, R. (eds.), Connectionist Symbol Processing: Dead or Alive?, *Neural Computing Surveys 2*, pp. 1-40, 1999, 220 references. http://www.icsi.berkeley.edu/~jagota/NCS

[2] Mani, D.R. (1995) The Design and Implementation of Massively Parallel Knowledge Representation and Reasoning Systems: A Connectionist Approach, Ph.D. Dissertation, Department of Computer and Information Science, University of Pennsylvania.

[3] Mani, D. and Shastri, L. (1999), Parallel Inference with Very Large Knowledge Bases, In *High Performance Cluster Computing* 2, (Ed.) R. Buyya, Prentice Hall. 1999.

[4] Shastri, L. (1999a), Advances in Shruti - A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony, *Applied Intelligence*, 11, pp. 79-108.

[5] Shastri, L. (1999b), Recruitment of binding and binding-error detector circuits via long-term potentiation, Neurocomputing, 26-27, pp. 865-874, 1999.

[6] Shastri, L. (1999c), Infants learning Algebraic Rules, *Science*, 285 (5434): 1673, September, 1999.

[7] Shastri, L. (1999d), A Biological Grounding of Recruitment Learning and Vicinal Algorithms, Technical Report TR-99-009, International Computer Science Institute, Berkeley, CA, 94704. April, 1999.

[8] Shastri, L. (2000a), Types and Quantifiers in Shruti --- a connectionist model of rapid reasoning and relational processing, *Hybrid Neural Symbolic Integration*, S. Wermter and R. Sun (Eds.), *Lecture Notes in Computer Science*, Springer-Verlag, Berlin. (In Press).

[9] Shastri, L., Chang, S., and Greenberg, S. (1999), Syllable Detection and Segmentation Using Temporal Flow Neural Networks, *Proceedings of the XIVth International Congress of Phonetic Sciences. ICPhS 99*, vol. 3, pp. 1721-1724, San Francisco, CA, August, 1999.

[10] Shastri, L., Grannes, D., Narayanan, S., and Feldman, J. (In Press), A Connectionist Encoding of Parameterized Schemas and Reactive Plans, in *Hybrid Information Processing in Adaptive Autonomous vehicles*, G.K. Kraetzschmar and G. Palm (Eds.), Lecture Notes in Computer Science, Springer-Verlag, Berlin.

[11] Shastri, L. and Wendelken, C. (1999a), Soft Computing in shruti: --- A neurally plausible model of reflexive reasoning and relational information processing," L. Shastri and C. Wendelken. *Proceedings of the Third International Symposium on Soft Computing. SOCO'99*, Genova, Italy. June, 1999, pp. 741-747.

[12] Shastri, L. and Wendelken, C. (1999b), Knowledge Fusion in the Large -- taking a cue from the brain, *Proceedings of the Second International Conference on Information Fusion*, FUSION'99, Sunnyvale, CA, July, 1999, pp. 1262-1269.

[13] Wendelken, C. and Shastri, L. (2000), Probabilistic Inference and Learning in a Connectionist Causal Network, *Proceedings of Neural Computation 2000*, Berlin, May 2000.

# 4 Network Services and Applications

## 4.1 Overview

The *Network Services and Applications (NSA) Group* arose out of the old group for *Very Large Distributed Systems and High-Speed Networking*. Many of the network specific research activities are part of the ACIRI research domain. The NSA group complements the activities of ACIRI and provides a platform for postdocs and visitors from sponsor countries and companies to either collaborate with ACIRI or to pursue new activities focused on network services and applications. Visitors to the NSA group currently include academic researchers from Spain and Germany, and industrial researchers from DaimlerChrysler, KPN (Dutch Telekom), Siemens, and Swisscom. The following sections give an overview of new research activities started in 1999.

## 4.2 Simulation of a Mobile Network Architecture for Vehicles

In this project (a joint endeavor of ICSI and DaimlerChrysler Research Palo Alto), we will investigate the design space for a mobile network architecture to enable electronic devices in vehicles to communicate with an internetwork such as the global Internet.

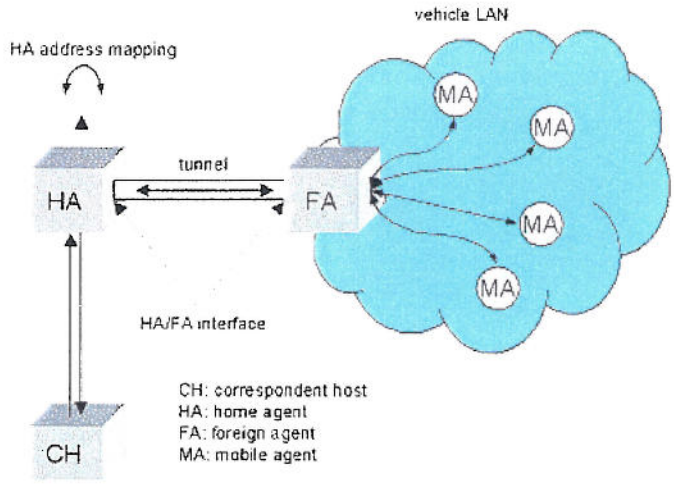



**Figure 5 Mobile IP Architecture for Vehicles**

It has to be determined if Mobile IP can be used as starting point for a network protocol that allows for transparent communication between a small network of interacting mobile hosts (i.e. the different network-enabled devices used in a car) and the Internet. A handover between two wireless communication providers can easily be predicted by using information of the car navigation system about current location, destination, and the anticipated route; and information on coverage areas by the providers. By making this information available to higher level applications and the home agent, the quality of service of the applications can be increased.

The first step toward the design of the architecture is to build a model of this network environment in a network simulator.

## 4.3 Equation Based Congestion Control (joint project with ACIRI)

In contrast to the TCP protocol, the UDP protocol has no back-off mechanism to reduce the bandwidth in case of network congestion. The goal of this project is to develop a UDP-based congestion control protocol that is TCP-friendly in the sense that it achieves the same long-term throughput as a TCP connection over a congested link.

We use a model-based characterization of TCP/IP throughput based on network conditions such as round trip time and loss rate. Together with an increase and decrease policy this allows us to compromise between responsiveness on the one hand and timeliness of continuous media data on the other [3].

Future work will include the adaption of the congestion control protocol to multicast environments.

## 4.4 Projects of the Middleware Technology Research Group at ICSI

- **Joint project with the WOS-Project Group: The Web Operating System (WOS)**

  The goal of the WOS is to provide a platform that allows the user to benefit from the computational potential offered by a wide-area distributed environment. It addresses the problem of making available to all sites of the network, the resources available at that network, to execute computations for which local resources are missing.

  Examples of resources that may be treated by the WOS range from the integration of different printers in a heterogeneous hard- and software environment to the "plug and play" installation

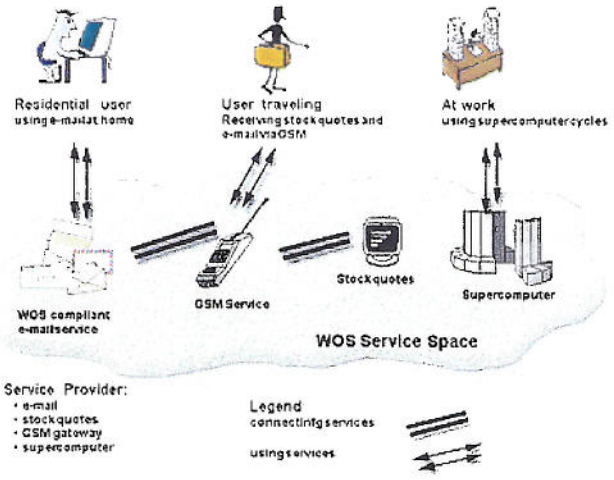



**Figure 6 WOS Architecture**

  of wireless PDAs (Personalized Digital Assistants) to the Internet via ISPs (Internet Service Providers) [4,5].

- **Joint project with Swisscom Research: ComSpace**

- ComSpace is carried out in collaboration with Swisscom Corporate Technology and aims to develop a profound understanding of how different heterogeneous (in terms of devices and bandwidth) communication media can be combined into a single, comprehensive communication framework for residential users.

  This includes the design of an easy-to-use interface employing a functional design approach, the development of new services which take into account typical access scenarios (broad-narrowband) encountered for residential users, and various technical questions such as service handover, and the separation of content and presentation of a ComSpace service [2].

- **Joint project with Eurescom: Jini and Friends @ Work: Towards secured service access**

  Our vision of the next step of the networked information society is that billions of small, intelligent devices will be equipped with spontaneous networking capabilities and have access to any information and provide access to any service "on the net". This will result in completely new security situations that must be solved and requires the fundamental building blocks of state-of-the-art IT technology, like interworking objects, and the corresponding middleware must be turned into Telco-grade technology allowing to offer services in a secure and consumer-friendly way to customers. Because we think that Jini is currently the most promising technology upon which to build this new world of networked computing communities, this project concentrates primarily on Jini and related technologies.

- **Joint project with SUN Microsystems: The Jini Home**

  The next generation of private households will use off-the-shelf consumer electronics connected to the Internet to directly take advantage of the vast amount of services delivered through the Internet. Possible applications include, but are not limited to: a lawn sprinkler talking to the Weather Channel to determine optimal lawn and shrub watering, home monitoring applications such as security services to protect your home, and services which download music and games and store it on your Jini enabled CD-player. SUN's ".com Home" initiative is one of the driving forces behind this development. As of today, the most promising technology to support these types of application is Sun's Jini Technology.

  The goal of this project is to verify the usability and/or to unveil the weakness of Jini as the base technology for such an interconnected home. Two aspects will be of particular concern: the overall usability if Jini, i.e. the complexity of integrating off the shelf devices into a Jini network using a proxy approach, and the secure access of services advertised by Jini.

## 4.5 Anonimity and Unobservability in the Internet

Today, "browsing" anonymously and unobservably on the Internet is an illusion. Insufficient data security such as authenticity and integrity is another serious problem.

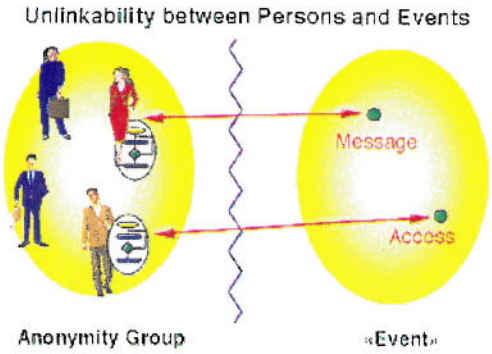



**Figure 7 Anoynimity and Unobservability**

Only the usage of strong cryptographic protocols and algorithms can provide these features by encrypting and signing data [1].

The objectives of our research are

1) to increase the privacy of using the Internet and to avoid the traceability and observability of users,
2) to provide a theoretical background for the efficient implementation of anonymity services in the Internet,
3) to develop a statistical model of user behavior as a basis for the creation (or rather "formation") of the anonymity group. Anonymity services are useful for electronic commerce applications (obviously with the need for strong authenticity and integrity of data) as well as for services where the user wants to remain anonymous (e.g. web based advisory services or consultancy).

For more information on this project contact Hannes Federrath (hannes@icsi.berkeley.edu).

### 4.6 High Speed Cable Access for Internet and Telelearning Applications

Internet and remote access control services are growing by leaps and bounds, major efforts are currently underway to provide greater bandwidth to the user by a multitude of access technologies. The key access technologies currently being developed are: XDSL, cable data networks, and fixed and mobile wireless networks. Satellite-based access to public and private data networks will have the potential to completely change the way we connect to the Internet in the future. There are several primary applications and services driving the demand for high speed access such as: Internet, multimedia communications, distance learning, remote access, etc. The proposal of this project is to study and analyze some relevant aspects of using cable-modems over cable TV systems to provide a high speed access alternative to DSL technologies or fixed and mobile wireless networks in an collaborative and educational environment ( Distance Learning). Several experiences have been developed in the last years. In this scenario, the CATV network can be viewed as just another data link technology that delivers broadband services to the subscriber in a shared medium. It will be simply be add some new components at the subscriber and head-end locations to send and receive packets. What distinguishes a telephone modem from a cable modem is that the former operates on a dedicated switched circuit , while the latter does it over a shared medium, like LAN architecture, where the users don't receive guaranteed bandwidth. Other problems in cable based Internet access are related with security and privacy in this shared medium. Finally, one of the worse effects in the

return path is due to electromagnetic interference between multiple users that share the same bus. There are several standard bodies and industry efforts in developing data over cable specifications that will be compatible with the existing customer terminal equipment, cable network infrastructure and data networking technologies. One of the leaders is MCNS ( Multimedia Cable Network Systems Partners). This group released in 1996 a set of specifications "DOCSIS" (Data Over Cable Service Interface Specifications). Other industry organizations like DAVID, Broadband Link and SCTE pretend the same. On the other way IETF- IPCDN WG is currently examining architectural issues related to Ipv4 and Ipv6 support over CATV networks. IP operating over CATV network is relatively straightforward but some issues could arise. One requirement is to keep out unnecessary traffic to avoid bandwidth consumption. Another issue is how best to support the requested QoS that the subscribers ask for. One of the best candidates for this purpose is the RSVP protocol. The issue of security is of particular importance to many applications that range from Distance Learning to work at home.

**Objectives to be addressed**

- To study several scenarios of Distance Learning systems using Internet/Internet as a medium of connected users.
- To analyze the possibilities of cable networks to gain fast broadband access in Distance Learning applications.

The development of these objectives will be addressed considering the following points:

- To select Distance Learning models and new classroom and working environments.
- To establish the necessities of a hypothetical and experimental high speed computer network and a very large digital media storage system.
- To know different tools and alternatives for protecting intellectual property rights.
- To study the current efforts of different standardization bodies regarding two-way transmission of multiple services over cable TV systems.
- To research among different protocols, for the best possible candidates to offer and guarantee assigned QoS over the shared network bandwidth.
- To propose security mechanisms to access Internet or Corporate Networks that include security services such as access control, authentication, digital signature, confidentiality and not deny of the service.

For more information on this project contact Jose Luis Melus (melus@icsi.berkeley.edu)

### 4.7 Multiuser Detection Receivers in CDMA systems

The work in this area is focused in the physical layer analysis and design of the following environments and research issues

- Quasi-static environment (WLAN and WLL applications)
- Hybrid TD-CDMA system, low spreading gains (<16)

- New signal models considering MUD coming from ISI. New matched front-ends
- Space-time diversity in reception. Vector receivers
- Time Division Duplex systems (TDD). Exploiting channel reciprocity
- Smart utilization of physical resources in TX (TFS)
- Analysis of Vector MC-CDMA
- New waveforms for CDMA
- Analysis of implementation issues (Computational load, Adaptability, etc.)

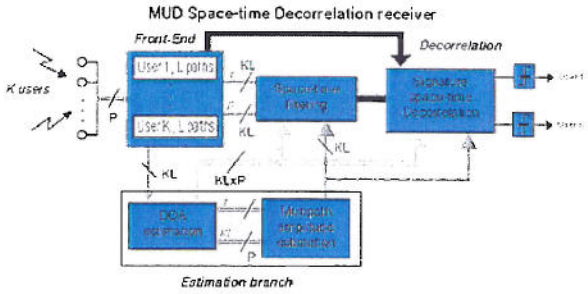



**Figure 8 Combined Space-Time Processing for Multiuser Detection Receivers in CDMA systems**

For more information on this project contact Jose M. Paez-Borrallo (paez@icsi.berkeley.edu)

## References


[1] Oliver Berthold, Hannes Federrath, Marit Köhntopp: Anonymity and Unobservability in the Internet, Workshop on Freedom and Privacy by Design / Conference on Computers, Freedom and Privacy 2000, Toronto/Canada, April 4-7, 2000.

[2] O. Krone, B. Messmer, H. Almiladi, T. Curran: Java Framework for Negotiating Management Agents, IMPACT99, Impact of Agent Technology on Telecommunications, Seattle, USA, December 2-3, 1999

[3] Sally Floyd, Mark Handley, Jitendra Padhye, and Jörg Widmer, "Equation-Based Congestion Control for Unicast Applications: the Extended Version", International Computer Science Institute Technical Report TR-00-03, March 2000

[4] O. Krone, Alex Josef,"Using CORBA in the Web Operating System" *Proceedings Workshop on Distributed Communitites on the Web*, LNCS, Springer Verlag, Quebec, June, 19-21, 2000.

[5] O. Krone, Alex Josef, "Integrating CORBA into the Web Operating System: First Experiences", High Performance Computing (HPC 2000), Washington D.C., April 16-20, 2000.

## Theory of Computation

### Introduction

The Theory Group has been concentrating on the application of algorithms and mathematical analysis to two areas of applied science where revolutionary developments promise to change society: computer networking and molecular biology. We also conduct research on more foundational areas such as computational complexity and combinatorial optimization.

### Mathematics of Computer Networking

**Centralized Control Versus Decentralized Control.** Tim Roughgarden and Eva Tardos are working on understanding the quality of routing obtained if separate agents each make selfish routing decisions in a global network. Our goal is to understand the tradeoffs between introducing a global control mechanism, and the loss in quality obtained by letting each agent make selfish decisions. We use a model when each link in a network has a delay that is an arbitrary monotone and continuous function of the amount of flow on the link. We assume that each agent has a source and a destination, and it controls an infinitesimally small amount of flow. A Nash equilibrium for this problem consists of flows such that each agent's flow is routed on the shortest available path in the network.

We compare the quality of such a Nash equilibrium with a globally planned optimal routing whose goal is to minimize the sum of all delays. In the case of linear delay functions, we show that the global optimum can be at most a factor of 4/3 better than the Nash equilibrium. For more general delay functions the value of the global optimum can be arbitrarily better than the Nash equilibrium even when the delay functions are rather simple (e.g., polynomials). On the other hand, we show that for any monotone and continuous delay function the cost of a Nash equilibrium is at most as much as the optimal cost for the case in which each agent has to send twice as much flow. Roughly speaking this shows that one can eliminate the need of introducing a global control mechanism at the cost of designing a network that can support twice as much flow. The work so far is limited to models where each agent has a given (small) rate of flow. In joint work with Scott Shenker we are working on extending the work to traffic models where agents have incentives to increase their rates.

### Broadcast Problems

- **Broadcasting in Static Networks** One of the simplest models of information flow in a processor network is the *broadcast* model. Here, in every unit of time a processor can only inform one of its neighbors. The fast transmission of information to all processors starting from given sources requires a good strategy. The computational complexity of determining good broadcast strategies is NP-hard. In [6] C. Schindelhauer has solved an open question in a surprising direction by showing that even a near-optimal solution to this problem is hard to compute.

- **Broadcasting in Dynamical Networks.** While traditional databases require transactional consistency, many repositories of information require only the much weaker notion of *eventual* consistency. That is, in such cases we care only whether, and how quickly, the information is disseminated, but do not require global consistency during the dissemination. Rumor-spreading algorithms (or, equivalently, epidemic algorithms) were used to distribute data in the Clearinghouse

nameserver database. In such algorithms, information is spread via randomized pairwise exchanges. A caller can transmit his information or can be informed by the called party. The first mode is called *push*, the other *pull*. How can a distributed algorithm be defined that informs all or a certain fraction of all players without using too many pushes and pulls? R. Karp, C. Schindelhauer, S. Shenker and B. Vöcking have obtained provably optimal algorithms for this problem and some of its generalizations.

**Multicast Versus Unicast** R Karp and S. Shenker have analyzed the relative efficiency of multicast versus unicast, in terms of the number of links that must be traversed to deliver data to a randomly placed set of receivers. In one simple model, where the interconnection structure is a uniform tree, an exact analysis has been carried out, even under models where the receivers are clustered together, or widely scattered across the network.

## 4.8 Computational Molecular Biology

**Universal DNA Tag Systems.** A fundamental property of DNA is the complementarity between the nucleotides A and T and between the nucleotides C and G. Two single-stranded DNA molecules tend to join together to form a double helix if each nucleotide on one strand is complementary to the corresponding nucleotide on the other strand. This joining is called *hybridization*. DNA microarrays take advantage of hybridization to detect particular molecules within a complex mixture of DNA or RNA. One realization of this idea is the DNA microarray, in which thousands of DNA probes are deposited at distinct positions on a small surface such as a microscope slide. When a fluorescently labeled mixture of DNA or RNA is applied to the chip, it is possible to detect the positions where hybridization has taken place. Hybridization to a probe indicates the presence in the mixture of a molecule complementary or nearly complementary to the probe. Usually the set of probes on a microarray is chosen with a particular application in mind, but recently an alternative approach was suggested that uses fixed, universal arrays. The idea is similar to the use of bar codes to identify products in a shop. Molecules in the mixture are marked with arbitrary DNA tags (bar codes), and then recognized through the hybridization of these tags to their complements on the chip. This approach presents an interesting design problem - the array should contain as many probes as possible, while minimizing errors due to unwanted hybridization of tags to noncomplementary probes. The paper [1] solves a version of this design problem based on a simple mathematical model of the hybridization process.

**Analysis of Gene Regulation** A cell in a living organism is said to *express* a gene if the cell manufactures and activates the protein associated with the gene. Although all cells in a living organism contain the same genes, different cell types express different gene, and do so under different conditions. Gene expression is regulated by a complex feedback control network, in which external influences on the cell and the distribution of proteins within the cell influence the expression of different genes. In [2] we present a new method to infer the structure of gene regulatory networks by performing strategically chosen interventions and observing theor effects. Each intervention artificially knocks out or over-expresses some gene. Through such interventions we derive mathematical functions which relate the expression level of a gene to the expression levels of the genes that influence it. The interventions are chosen through an iterative maximum-entropy procedure which, at each iteration, identifies that intervention which is expected to narrow down the space of alternative network hypotheses to the greatest extent.

**Pedigree Analysis** Pedigrees are "family trees" describing relationships within extended families. They are being extensively collected, together with phenotypic and genotypic information, in order to identify the genes responsible for particular functions or diseases and determine their positions in the genome. Most algorithms for solving this problem need to repeatedly evaluate the probability of incomplete data under a given hypothesis on the relative distances between the genes. In [3] we show that this problem is NP-hard even in the case of a single locus in inbreeding-free pedigrees. On the positive side, it is possible to obtain an approximate evaluation of the same quantities in polynomial time, using techniques developed for Bayesian networks.

A similar class of problems arises when trying to reconstruct the gene flow in a pedigree from partial data. In this case the goal is to find the most likely explanation of the data. A more modest goal is to find one consistent explanation. Restrictions on the class of pedigrees and on the number of loci considered can affect the difficulty of these problems. A. Piccolboni and D.Gusfield are currently working to determine which restrictions lead to tractable computational problems.

**Combinatorial Chemistry** Antonio Piccolboni has been investigating problems related to graph theory and combinatorial chemistry. The basic assumption is that the structural formula of a molecule can be viewed as a graph. A variety of descriptors for graphs has been defined. They can range from a single number, to a list, or they can be even more complex. Some of these descriptors are related to chemical and biological properties of interest, like the boiling temperature or the ability to cross membranes. In modern applications, many such descriptors are used simultaneously to predict the activity patterns of candidate drugs. The problem is therefore to find graphs, within a given class, for which a set of descriptors has an optimal value. A classic descriptor is the Wiener index. In [4] an efficient algorithm is given for finding a graph with a given Wiener index. A closely related algorithm solves the more concrete problem that arises when the set of graphs is generated starting from a combinatorial library. The approach can be extended to a class of indexes including the polarity index and to linear combinations of them.

**Protein Folding** For some years Piccolboni has been investigating protein folding, one of the grand challenges of contemporary molecular biology. This problem can be formulated as the optimization of the free energy of a protein. The simplest and best known model of protein folding is the HP or Dill model. Under this model, the problem is known to be NP-hard. In joint work with G. Mauri and G. Pavesi, a class of algorithms for this model has been designed, which are based on a formal grammar description of solutions. It was shown that one of these algorithms achieves 1/4 of the energy optimum in O(n^2) sequential time for the two-dimensional version of this model. Moreover, the 1/4 bound does not appear to be tight, and richer grammars have given better performance, at least experimentally.

### 4.9 Computational Complexity

**Circuit Complexity** In earlier work at the University of Lübeck, C. Schindelhauer and and his colleagues investigated *Average Circuit Complexity Classes*. For some well-known functions they derived matching upper and lower bounds on average complexity for certain classes of probability distributions of the circuit inputs. Joint work with Andreas Jakoby (University of Toronto) on this topic was continued at ICSI. Currently, his work concentrates on how to transfer results to VLSI-circuits.

**Average Complexity** In [7] we show that certain non-recursive problems become feasible under an average-case complexity measure, while some other prominent problems such as the *halting problem* or *Kolmogorov complexity problem* remain highly infeasible.

### 4.10 Combinatorial Optimization

Vardges Melkonian and Eva Tardos are working on approximation algorithms for network design and routing problems. This type of problems have a wide range of applications in telecommunication networks, internet, VLSI design, etc. The problems are all NP-hard and so it is important to find algorithms which will return reasonably good solutions in polynomial time.

The goal in the network design problem is to design a network of minimum cost satisfying connectivity constraints. In the last 10 years there has been significant progress in designing approximation algorithm for network design problems on undirected graphs. Very little is known about directed versions. In [13] we make the first step for network design problems on directed graphs. We give a 2-approximation algorithm for the directed network design problem where the connectivity requirements give rise to a crossing supermodular function. The main tool of their research is the technique of formulating the problems as integer programs and trying to find approximation algorithms exploring the structure of the solution of the linear programming relaxation.

We are working on extending our techniques to other network design problems. Maybe the most important directed network design problem is the Traveling Salesman Problem, where the best known approximation guarantee is $O(\log n)$. Other problems they have been working on include the minimum latency problem, network design with orientation constraints and some variations of Steiner tree problem.

**References:**


[1] A. Ben-Dor, R. Karp, B. Schwikowski and Z. Yakhini, Universal DNA Tag Systems: A Combinatorial Design Scheme, *Proc. RECOMB 2000.*

[2] T. E. Ideker, V. Thorsson and R.M. Karp, Discovery of Regulatory Interactions Through Perturbation: Inference and Experimental Design, *Proc. Pacific Symposium on Biocomputing*, 1999.

[3] Piccolboni, A., Gusfield, D., On the complexity of fundamental computational problems in pedigree analysis. Technical Report CSE-99-8, Computer Science Department, University of California, Davis, 1999.

[4] Goldman, D., Istrail, S., Lancia, G., Piccolboni, A. and Walenz, B., Algorithmic strategies in Combinatorial Chemistry. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 275--284, 2000.

[5] Mauri, G., Pavesi, G., Piccolboni, A., Approximation Algorithms for Protein Folding Prediction. In *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 945--946, 1999.

[6] C. Schindelhauer, Broadcasting Time cannot be Approximated within a Factor of $57/56 + \varepsilon$, submitted for publication (2000)

[7] A. Jakoby and C. Schindelhauer, The Non-Recursive Power of Erroneous Computation, *FST and Theoretical Computer science* (1999).

[8] V. Melkonian and E. Tardos, Approximation Algorithms for a Directed Network Design Problem. In the *Proceedings of the 7th International Integer Programming and Combinatorial Optimization Conference (IPCO'99)*, Graz 1999.