Alfred V. Aho is Lawrence Gussman Professor of Computer Science and Vice Chair for Undergraduate Education in the Computer Science Department. He served as Chair of the department from 1995 to 1997 and in the spring of 2003. He was elected to the U.S. National Academy of Engineering for contributions to the fields of algorithms and programming tools.

Professor Aho received his B.A.Sc in Engineering Physics from the University of Toronto and a Ph.D. in Electrical Engineering/Computer Science from Princeton University.

At this year’s computer science department “Hello Meeting,” members of the faculty recalled the early days of the department when the same meeting could be held on the first floor of Mudd, in the small area that is now used for making sandwiches in the Carleton Lounge. Since then, the computer science department has grown both in size and prominence. The faculty now includes six members of the National Academies: Al Aho, Steven M. Bellovin, Zvi Galil, Ted Shortliffe, Joe Traub, and Vladimir Vapnik. Here is a brief overview of their achievements and contributions.
Cover Story (continued)

Professor Aho has won the IEEE John von Neumann Medal and has received honorary doctorates from an NFC study group on science versus terrorism. He was a member of the Internet Architecture Board from 1996-2002, and was a member of the Security Area of the IETF from 2002 through 2004. He sits on the Networks and Security Board, the Computer Science and Technology Advisory Board.

Steven M. Bellovin is a Professor of Computer Science. He was elected to the National Academy of Engineering for contributions to network systems trustworthiness. He was a member of the information technology and cybersecurity research working group where he contributed to the National Research Council study committees, on many National Research Service Advisory Committees, and for leadership in computer science and engineering. He formulated in 1995: “When you really need but not a step. Try to get the answer that you really need but not a more general one.”

Zvi Galil is the Dean of the School of Engineering and Applied Science. He is the Julian Clarence Levi Professor of Mathematical Methods in Computer Science, and served as Chairman of the Department of Computer Science from 1989 until July 1995. In 2004, Dean Galil was elected to the National Academy of Engineering for his contributions to the design and analysis of algorithms and their applications in computer science and engineering. Dean Galil began his studies in applied mathematics; he soon turned to the emerging field of computer science, which he saw to be an extension of his interests in mathematics. His main research pursuits are in the design and analysis of algorithms, computational complexity, and cryptography. He is a world leader in the design and analysis of computer algorithms, and has developed a number of techniques to improve their efficiency. Many of his algorithms remain the fastest at solving a particular problem, and can be used to reduce the smallest amount of memory to find a solution. He has collaborated with scientists in diverse fields, including biology, mathematics, and statistics, to devise novel ways to attack difficult problems.

Dean Galil has been a professor at Columbia University since 1982, having previously served at Tel Aviv University. He is the author of more than 200 research papers in refereed journals and is the editor of the book “Computational Algorithms on Strings.” Although he served in Engineering and Applied Science has taken up most of his time for the last 12 years, he reports, with great mirth, that he still tries to find the most efficient way to solve problems.

Edward H. Shortliffe is Robert Schachter Professor and Chair of the Department of Biomedical Informatics at Columbia College of Physicians and Surgeons. He is an elected member of the Institute of Medicine of the National Academy of Sciences. After receiving an A. B. in Applied Mathematics from Harvard College in 1970, he moved to Stanford University where he could pursue interests in medicine and computer science, and was appointed to the Stanford Biomedical Informations in 1976 and as Dean of the Faculty in 1980. Subsequently, he was served as Professor of Medicine and of Computer Science at Stanford University. In January 2000 he assumed his new post at Columbia where he is also Deputy Vice President (Columbia University Medical Center); Senior Associate Dean for Strategic Information Resources (College of Physicians and Surgeons); Professor of Medicine; Professor of Computer Science; and Director of Medical Informatics Services for the New York Presbyterian University Hospital. During the early 1970s, Professor Shortliffe was the principal developer of the medical expert system known as MYCIN. After a pause for internal medicine house-staff training at Massachusetts General Hospital and Stanford Hospital between 1976 and 1978, he joined the Stanford internal medicine faculty where he served as Chief of General Internal Medicine, Associate Chair of Medicine for Primary Care, while directing an active research program in clinical information systems and decision support, and spearheaded the formation of a Stanford graduate degree program in biomedical informatics and divided his time between clinical medicine and biomedical informatics research.

Professor Shortliffe continues to be closely involved with biomedical informatics and graduate training where he works to create a new breed of health professional. His research interests include the broad range of issues related to integrated decision-support systems, their effective implementation, and the role of the Internet in health care.

Joseph F. Traub is the Edwin Howard Armstrong Professor of Computer Science. He came to Columbia in 1979 as the 7th holder of the Computer Science Department Chair, and became Professor of Computer Science at Stanford University. In January 2000 he assumed his new post at Columbia where he is also Deputy Vice President (Columbia University Medical Center); Senior Associate Dean for Strategic Information Resources (College of Physicians and Surgeons); Professor of Medicine; Professor of Computer Science; and Director of Medical Informatics Services for the New York Presbyterian Hospital. During the early 1970s, Professor Shortliffe was the principal developer of the medical expert system known as MYCIN. After a pause for internal medicine house-staff training at Massachusetts General Hospital and Stanford Hospital between 1976 and 1978, he joined the Stanford internal medicine faculty where he served as Chief of General Internal Medicine, Associate Chair of Medicine for Primary Care, while directing an active research program in clinical information systems and decision support, and spearheaded the formation of a Stanford graduate degree program in biomedical informatics and divided his time between clinical medicine and biomedical informatics research.

Professor Shortliffe continues to be closely involved with biomedical informatics and graduate training where he works to create a new breed of health professional. His research interests include the broad range of issues related to integrated decision-support systems, their effective implementation, and the role of the Internet in health care.

Vladimir Vapnik is Professor of Computer Science. He is currently at the Center for Computational Learning Systems. The National Academy of Engineering elected Professor Vapnik as a new member in 2006 for “insights into the fundamental complexities of learning and for inventing practical and widely applied machine-learning algorithms.”

Professor Vapnik obtained his Masters Degree in Mathematics in 1958 at Uzbek State University, Samarkand, USSR and his Ph.D at the Institute of Control Sciences in the Academy of Sciences, Moscow, in 1964. From 1964 to 1980 he worked at the Institute, where he became Head of the Computer Science Research Department. He then joined AT&T Bell Laboratories in Holmdel, NJ, as a Consultant, and was appointed Professor of Computer Science and Statistics at Royal Holloway in 1995. Professor Vapnik has taught and researched in computer science and theoretical and applied statistics for over thirty years. He has published six monographs and more than one hundred research papers. His major achievements have been the development of a general theory for machine learning, including the expected risk of predictors using empirical data, and a new type of learning machine (the Support Vector Machine) that can learn and make predictions with the best possible generalization ability. These techniques have been used to solve complex tasks in image and text recognition and regression estimation problems and have been applied to the problems of dependency estimation, forecasting, and constructing intelligent machines. Professor Vapnik’s research combines mathematical analysis of the problem of empirical inference in complex (high dimensional) data sets with philosophical interpretation of such inferences based on the imperative which he formulated in 1995: “When solving a problem of interest, do not solve a more general problem first. Try to get the answer that you really need but not a more general one.”

These six National Academy members represent the growth and prominence of the Columbia Computer Science Department since its modest beginnings. Undergraduates and graduate students alike are fortunate to have such giants to stand on.

Feature Article

Professor Shree Nayar Named as 2006 Great Teacher

T.C. Chang of Professor Computer Science Shree Nayar.

Professor Shree Nayar was named as the recipient of the 2006 Great Teacher Award for the School of Engineering at Columbia University. The award was bestowed by the Society of Columbia Graduates, its Board of Directors named Professor Nayar 2006 Great Teacher because it feels that he exemplifies the greatest traditions of teaching at Columbia and has earned the recognition of his students and his peers as a dedicated and inspired undergraduate teacher and mentor. As one of the Society’s Great Teachers, Shree is one of Columbia’s finest and most beloved professors, such as Mark Van Doren, Lionel Trilling, Mario Salvadori, Morton Friedman, Rene Testa, and the others. The Great Teacher of the Year Award Dinner was held in Low Library on the evening of Thursday, October 19, 2006. The Society was formed in 1909; it will soon be celebrating its 100th anniversary. Throughout much of its existence, the Society’s principal mission has been to recognize great service to Columbia by its alumni and by its faculty. Congratulations to Shree on this well-deserved recognition of his outstanding teaching.
New Computer Science Courses

Debbie Cook and Moti Yung offered COMS 6998-1, Practical Cryptography, in the Fall 2006 semester. The course focuses on practical aspects of cryptography to complement the topics covered in “Introduction to Cryptography” and in “Network Security.” The course consists of a mixture of lectures and student presentations of current research papers. Topics include the design of algorithms used in practice and cryptanalysis, hash functions, elliptic curves, electronic cash, threshold cryptography, forward security, key insulated security, and trusted computing. The course requires a semester-long project on a topic of the student’s choice. The projects are practical in nature and include implementations of attacks on or the statistical analysis of algorithms, and implementations of electronic cash and electronic voting systems.

Donna Dillenberger offered a new 6000 level graduate course, Advanced Computer Design, in the Fall 2006 semester. The goal of the course is to train students to design systems holistically, from
- The hardware level (chip level, I/O subsystems, packaging, cooling considerations),
- The virtualization layer (hardware, software hypervisors, paravirtualization),
- The operating system (load balancing, parallelism, reliability, security, and performance),
- Middleware scalability and clustering issues.

Student learn how tradeoffs in these layers lead to different optimizations for different types of workloads. The course uses real case studies in the design of mainframes, desktops, PDAs, and set top boxes to illustrate architectural concepts. The final assignment is for students to write a paper on how they would design their own ideal system for a particular workload they are targeting, describing their ideal hardware, operating system, virtualization, and I/O layers. At the end of class, students pitch their designs to a fictitious CEO (our class) and the class votes for which system to “bet” the company on.

Prabhakar Kudva offered CSEE 4340 (formerly listed as ELEN 4340) for the first time as a Computer Science department course in the Spring 2006 semester. This course is a hands-on, laboratory-based introduction to the design of digital systems, culminating in the design of a complete working computer. Students learn computer architecture, register transfer-level specification, describe designs with and simulate using VHDL, and implement a computer using FPGAs. This is a practical course, not a theoretical one. The goal is to teach the art and practice of design system design by working on a real design; as such, most of the focus of the course is the laboratory and the project. The course includes a lot of hands-on experience with industrial design tools and techniques. Students extensively use the industry standard language VHDL as well as use the Cadence tool suite also widely used in industry to develop their design.

Professor Itstik Pe’er taught COMS 6998-3, Computational Human Genetics, in the Fall 2006 semester. This course is intended to introduce students of both computational and bio-medical skill sets to current quantitative understanding of human genetics and prepare them for computational research in the field. Topics include: genetics of a single site, coalescence with recombination, history of humans, mapping rare mutations through linkage, mapping common variants through association, isolated and admixed populations, natural selection, copy number changes, model organisms, and genotyping technologies. The computational toolbox discussed includes parameter inference, likelihood analysis, hidden Markov and other graphical models, eigenvector decompositions, and classification problems.

Professor Henning Schulzrinne offered COMS 4995-01, Special Topics in Computer: VoIP Security, in the Fall 2006 semester. This is a seminar and lab course on VoIP (voice-over-IP) and VoIP security. The course defines VoIP broadly to include real-time voice, video and instant messaging. Topics covered include:
- basic VoIP technology: audio and video coding, RTP, SIP;
- IM and presence: proprietary systems, XMPP and SIMPLE;
- basic IP network security issues: impersonation and authentication, privacy;
- VoIP transport (“spit,” “spim”) and prevention mechanisms;
- VoIP denial-of-service attacks;
- Skype;
- security for peer-to-peer VoIP;
- and
emergency calling.

As part of the course, students
- learn about VoIP protocols and technology;
- install, test and measure a complete VoIP system;
- conduct a team project implementing, as open-source software, an aspect of VoIP; and
- prepare a survey talk on a topic related to VoIP.

Adjunct Professors Michael Theobald and Franjo Ivanovic taught CSEE 6832, Formal Verification of Hardware and Software Systems, in the Fall 2006 semester. The course introduces the theory and practice of formal methods for the analysis of concurrent and embedded systems. The focus of the course is on model checking, which is an algorithmic approach to verification. Topics include temporal logics, Binary Decision Diagrams, and SAT solvers. Students also learn how to use the popular SPIN and SMV model checkers, and hear about practical experiences of applying formal verification to hardware and software projects in industry. Students may elect to do a research project that is geared towards their background.

Bernard Yee offered COMS 4995-2, Video Game Design and Development, in the Spring 2006 semester. This course covers the process of creating and developing video game designs; the theme is to allow creative vision and technical limitations (time, budget, team size, technology choices, etc.) to constrain and inform each other. Students learn to design and prototype game designs, acquire tools for the critical analysis of the gameplay experience, and apply these critical skills as they play, dissect, tune, and replace games in a rigorous iterative game development process. The course also touches upon related subjects including interactive narrative, character design, multiplayer dynamics, AI design, production processes, and the business of games. The approach uses both academic lecture/discussion class time and practical, guided workshop sessions. Students are expected to work individually and collaborate as part of a small team. The theme is to replicate a real world pre-production process.
Professor Dana Pe’er uses Bayesian networks and other statistical approaches to develop models that integrate heterogeneous types of biological data to uncover interactions between biological components and elucidate how these components work together at a systems level to compute and execute decisions. These models can represent stochastic nonlinear relationships among multiple interacting molecules and accommodate the noise inherent to biologically derived data. Additionally, the models are inherently flexible enough to integrate diverse data types and handle unobserved components. These network models provide the means to address questions such as how dysregulation causes disease and where are the optimal points for corrective intervention and cure.

Dana Pe’er received a B.Sc. in mathematics (1986), an M.Sc. in theoretical computer science (1989) and a Ph.D. for her work in machine learning approaches for reconstruction of molecular networks (2003), all from the Hebrew University of Jerusalem: Professor Pe’er’s graduate research pioneered the use of probabilistic graphical models (Bayesian Networks) for the reconstruction of molecular networks. Her influential adaptation of Bayesian networks has more than 600 citations and is taught in computational biology courses worldwide. In 2003, Dr. Pe’er joined the laboratory of Professor George Church at the Department of Computer Science at Harvard Medical School. During her postdoctoral fellowship, she gained biological expertise, while continuing to do research of significant impact and influence on both the computational and biological sciences.

Professor Dana Pe’er has received numerous awards, most notably the Burroughs Welcome Fund Career Award at the Interface of Science. Her work on reconstructing signaling networks in human cells was runner up for Science Magazine’s Breakthrough of the year 2005. At Columbia University, Dana Pe’er intends to continue bridging the gap between Computer Science and Biology to elucidate fundamental understandings in biology and disease.

Professor Itsik Pe’er holds bachelor’s, masters, and doctoral degrees in computer science from Tel Aviv University. He has worked extensively in biomedical research laboratories at the Weizmann Institute and the Broad Institute of Harvard & MIT, where he conducted postdoctoral research. Itsik develops, applies and novel computational methods in human genetics. How is it best to describe, and quantify differences between individual DNA sequences? How does sequence variation affect biological processes? How can we use it to understand and influence human disease? All these questions pose complex analytical challenges, with direct impact on medical research. Professor Itsik Pe’er is specifically interested in genetics of special populations that underwent bottleneck and admixture events, in the development and analysis of whole genome association studies, and in the interplay between somatic and germline variation.

How does sequence variation affect biological processes? How can we use it to understand and influence human disease? All these questions pose complex analytical challenges, with direct impact on medical research. Professor Itsik Pe’er is specifically interested in genetics of special populations that underwent bottleneck and admixture events, in the development and analysis of whole genome association studies, and in the interplay between somatic and germline variation.
Kevin Egan graduated from Brown University in 2003, then worked on the software renderer at Rhythm and Hues Studios in Los Angeles for two years. He started as a Ph.D. student at Columbia in Fall 2005. He is working with Professor Ravi Ramamoorthy’s Computer Graphics group.

David Harmon graduated from Wofford College with a B.S. in Computer Science and Mathematics in May 2005. After he was awarded a National Science Foundation Graduate Fellowship, he enrolled in the computer science graduate program at Columbia in the fall of 2005. He is currently working with Professor Elian Grinspun as part of the Columbia Computer Graphics Group. His research interests include physical simulation, computer animation, and computer graphics.

Steve Henderson started his Ph.D. studies at Columbia in Fall 2006. He is working with Professor Feiner in the Computer Graphics and User Interfaces Lab. Steve joins Columbia from the faculty of the University of Amsterdam, where he has been an Assistant Professor of Computer Science. Steve is an officer in the US Army, and holds an MS in Systems Engineering from the University of Arizona, and a BS in Computer Science from the United States Military Academy. His research interests are mobile augmented reality, information visualization, intelligent systems engineering, social networks, large-scale database design, and optimization.

Marina Johnson graduated from the University of California at San Diego with a B.A. in Computer Science in May 2005. Her research interests are user-centered design, professional vs. student compute interaction, and security. She is currently working with Professor Steven Bellon in designing a method of evaluating a financial institution’s ability to authentically represent itself to users in an online environment. In the past she has been active in various cryptography in computer science community and plans to continue doing so while at Columbia.

Tie Hu is a postdoc in the Robotics lab working with Professor Peter Allen. He received a Ph.D. in Mechanical Engineering from Brandeis University in 2006, and worked as an automation engineer in China from Fall 2000 to 2001. His research interests are robots, haptics, soft tissue modeling, electro-mechanical system design and control, and instrumentation design. He is currently working on the project ‘in vivo imaging system for minimally invasive surgery’ in the Robotics Lab.

Aly A. Khan graduated from Carnegie Mellon University with a S.B. in Computational Biology in May 2005. He completed his MS thesis under Professor Neil Schwarz on modeling a novel biological mechanism for splicing of very large introns. He was a Howard Hughes Medical Institute post-doctoral fellow while working in the lab of Professor Chris Brey-Kellogg at Purdue. He is interested in biological networks and is currently working with Dr. Christina Leslie and Professor Chris Wiggins, in collaboration with Dr. Chris Sandes lab at Memorial Sloan-Kettering, on studying microRNA regulation.

Bert C. Chung graduated from Brandeis University with a B.S. in Computer Science and a B.A. in Philosophy in May 2005. He completed an M.S. at Columbia in February 2006. Bert’s research interests are in the field of Machine Learning. Currently, he is working in the Center for Computer Learning Systems and in the Machine Learning Laboratory.

Jae Woo Lee graduated from Brandeis University with a B.A. in Computer Science in 1999. After graduating, he worked for a number of financial software/service companies, he co-founded MyRisk.com in 1999, which was acquired by DirectAdvice.com in 2001. He came back to Columbia in 2004, became a full-time Student in computer science department and plans to continue doing so while at Columbia.

Ohan Oda received his B.S. from University of Wisconsin Madison with double degrees in Computer Science and Computer Engineering in May 2005. He has worked at GE Healthcare for two summers as an intern. He started as a M.S student in Computer Science at Columbia in Fall 2005, and moved to the Ph.D program in Fall 2006.

Kevin Egan graduated from Brown University in 2003, then worked on the software renderer at Rhythm and Hues Studios in Los Angeles for two years. He started as a Ph.D. student at Columbia in Fall 2005. He is working with Professor Ravi Ramamoorthy’s Computer Graphics group.

Dana Glasner graduated from Yeshiva University with double B.A.s in Computer Science and Psychology in 2001. She graduated with a B.A. in Computer Science at Columbia in Fall 2006. She is currently working with Professor Kenneth Ross’s group in Database systems.

Richard Neill holds a B.A. and M.S. in Electrical Engineering and has held a technology position at Cablevision Systems Corporation. He started his Ph.D. in Fall 2003 working with Professor Carolin’s research group. Richard’s main research interests are in distributed embedded systems, parallel architectures, and grid computing.


Hans Schulzrinne is a professor with the Computer Science Department at Columbia University. He received his B.S. in Electrical Engineering from Wofford College in 1974. He received his M.S. in Computer Science in 1976 from UIUC. He received his Ph.D. in Computer Science from the University of California, Berkeley in 1980.

New Faces (continued)
Department News & Awards

Alp Atel (Math) received the E. M. Gold Award for ALT 2006 (the 17th International Conference on Algorithmic Learning Theory, held in Barcelona in October). This award is given to the best student paper at the conference. The paper "Learning Unions of $\Omega(m)$-Dimensional Range Spaces" was co-authored by Professor Rocco Servedio.

Professors Steven Bellville and Henning Schulzrinne will participate in the International Technology Alliance. The alliance will perform research in the four areas of network, information, computer, and systems, sensor information processing and delivery and distributed coalition planning and decision making.

Professors Steven Bellville and Sal Stoilo, together with Professor Sean Smith of Dartmouth, received a grant from ARO to organize and run a 'Networks in Cyberspace Threat Research. The workshop will be held in Washington, DC in June 2007.

Professor Steven Bellville was interviewed by Robert Lang from AIP's popular program "All Things Considered" in May. Steven discussed the National Intelligence, and his efforts to analyze the huge data produced by electronic phone calls turned over to the NSA by phone companies.

Professor Steven Bellville was interviewed by a Colorado 7TV article on security and terrorism.

Matthew Burnside, Jae Woo Lee, Christopher Murphy and Professor Rocco Servedio were named "extraordinary TAs" for the Fall 2005 semester, based on the evaluation of students in their classes. Congratulations to our outstanding TAs.

Professor Tom Katsoulis, along with Professor Ken Shepard of the Electrical Engineering, and T. several years, was awarded a three-year NSF grant to study the design of low-power scalable communication networks for multi-core, multi-system on-chip. The conference had a 28% acceptance rate, only two awards were given to the conference, and submissions they had this year.

Professor Stephen Edwards was awarded an NSF grant for NTSS, an effort in collaboration with R. Embedded Systems with Deterministic Communication. The SHIM model of computation provides deterministic concurrency with reliable communication, simplifying validation because behavior is reproducible. Based on asynchronous concurrent processes that communicate through rendezvous channels, SHIM can handle control, communication and dataflow, and data-dependent decisions.

Members of the department's computer graphics group were extremely successful in the upcoming SIGGRAPH 2006 conference. The group accepted, the most from any single institution in the last five years, four papers at the prestigious conference for computer graphics; the 2006 conference took place in Boston, Massachusetts in August 2006. A total of 86 papers were accepted from 436 submissions. Authors from Columbia were: Prof. Ravi Ramamoorthi; Prof. Peter Petereltz; Ian Gout; Julia Hirschberg; Eitan Grinspun; and Prof. Peter Belhumeur, along with several graduate students.

More information about the Columbia Vision + Graphics research lab is found at http://www.cs.columbia.edu/cvgc/

The paper “To Search or to Crawl? Towards a Query and Crawler Design for Text-Centric Tasks,” by Panos Ipeirotis, Eugene Agichtein, Pranay Jain, and Professor Peter Belhumeur, received the “Best Paper” Award at the SIGMOD 2006 Conference. SIGMOD is one of the two premier database conferences in the world, featured over 270 posters. Alex, who is Professor John Katsoulis, was the first author of his paper reporting on research he did as part of his IBM internships with the IBM Multimedia Retrieval group. The paper addresses the following key application areas:

- Matching authors: permutable clustering methods and permutationally invariant kernels are used to compute similarities between authors. The algorithm aims to find a way to cluster similar authors together, given a set of publications.
- Matching text and multimedia documents: permutationally invariant kernels compute similarities between observed text, images and word matchings. The text and multimedia data comes from descriptions of images and text, and the matching is realized by permutationally invariant kernels.

Professor Angelos Keromytis shared the security tracks for the 2007 World Wide Web Conference (WWW) and for the 2007 International Conference on Applications and Systems in Database Computing (ICDSIC). Ph.D. students Homin Lee and Andrew Wan received the Mark Fulken Best Student Paper award at the 19th Annual Conference on Learning Theory (COLT 2006), held in Pittsburgh, PA, in July. The award is for the paper titled "DNNs are Teachable in the Average Case," which is joint work with Professor Rocco Servedio. COLT is the top conference in computational learning theory, with more than 100 papers submitted per year for the last five years. Ph.D. student Jason Liscombe won one of three "best student" paper awards at the INTERSPEECH 2006 conference. The paper was co-authored bypostdoctoral researcher Angelos Keromytis and Professor Julia Hirschberg. The paper was titled "Detecting and Reranking Mutually Invariant Features Using the Multispectral Detector." It is a significant contribution to the understanding of their prosody. This research addresses the "prosody translation" problem for Mandarin Chinese and English second-language learners by identifying correspondences between prosodic phenomena in each language that convey similar meanings.

Professor Julia Hirschberg was awarded the 2007 Governor's Award of Excellence in Research, which is a competitive award from the State of New York. She is the first female scientist awarded this honor. In addition, she was awarded the 2007 Provost's Award for Excellence in Research.

Professor Tony Jebara received his third KDD Award, titled "Leveraging This year, an annual conference on machine learning dates back to 1988. SIGMOD, which has about 1500 members. The conference carries on in a large scale format each year in Pittsburgh. There are usually about 3000 attendees and several hundred tutorials, courses and workshops. ISCS is one of the major sources of funding for the international community.

Professor Vishal Misra (Computer Science), Dan Farkas (Electrical Engineering and CS) and Ed Coffman (EE and CS), together with Professor Predrag Jelenkovic and Professor Mor-Haol Bitter 1st CMU won a
highly-competitive NSF grant to study resource sharing and allocation on large server farms. A wide variety of systems, including web farms, virtual machines, multi-tasking OSes, GRID computing systems, and sensor networks improve their accessibility, availability, resilience and fairness by "sharing" resources across the consumers they support. However, research that explores how to share resources generally point solutions, where different resources/consumer configurations require separately designed sharing mechanisms. This project seeks to develop and analyze Adaptive Sharing Mechanisms (ASMs) in which the mechanisms used to share resources adapt dynamically to both the set of available resources and the current needs of the consumers, such that the system is truly autonomic. The grant extends over three years and is part of the NSF CosyTrust Systems Research (CSR) program. Only approximately 10% of all grant applications were funded.

Professors Vishal Misra and Dan Rubenstein won a three-year NSF CyberTrust grant on network computing security. Their proposal seeks to further development of a methodology for measuring the inherent security of the control plane component of existing and future routing protocols. The approach has a significant theoretical component: it involves looking at general classes of routing protocols and showing how they can be analyzed for their ability to monitor themselves.

Professor Shree Nayar’s work was profiled in an August 2006 IEEE Computer Magazine article (cover feature) on Computational Photography. Professor Kenneth Ross was awarded an NSF grant to study the design of database systems software on modern multicore and multithreaded processors. This project, titled “Cache-Aware Database Systems on Modern Multithreading Processors,” studies how to best utilize the resources available in modern processors in the development of database system software. A primary objective is avoiding the cache interference between threads in multithreaded and multi-core processors, so that performance scales well as the number of cores/threads increases. A variety of techniques are considered, including multi-threaded algorithm design, threads explicitly devoted to resource management, and scheduling algorithms that are aware of thread interference patterns. Simulations and implementations on real hardware are used to measure the effectiveness of each approach. Project-related information can be found at http://www.cs.columbia.edu/~krr/tacpaper.pdf.

This project was one of only eleven funded in the Database Management Systems program in 2006 and lasts through August 2008. Professor Henning Schulzrinne and Professor Ram Dantu (U. North Texas) have received a two-year grant from the National Science Foundation, as part of the CyberTrust program, to study methods to prevent unlicensed calls in VoIP systems. The research will focus on using trust paths to determine whether unknown callers are likely to be telemarketers or other spammers. Trust paths capture transitive trust in a friend-of-a-friend model, with trust established by having a person send email or call another person. Such trust paths are suitable for low-risk decisions, such as whether to accept an email or phone call, rather than high-risk decisions such as whether to loan money or reveal private information.

Professors Sal Stolfi and Angelos Keromytis were awarded a DARPA grant titled “Behavior-based Access Control and Communication in MANETS.” Through this grant, they will develop a new, behavior-based mechanism for authenticating and authorizing new nodes in MANETS. Rather than only granting access to a network, or to services on a network by means of an authenticated identity or a qualified role, they propose to require nodes to also exchange a model of their behavior to grant access and to assess the legitimacy of their subsequent communication. When a node requests access, it provides its pre-computed egress behavior model to another node who may grant it access to some service. The receiver compares the requestor’s egress model to its own ingress model to determine whether the new device conforms to its expected behavior. Access rights are thus granted or denied based upon the level of agreement between the two models, and the level of risk the recipient is willing to manage. The second use of the exchanged models is to validate active communication after access has been granted. As a result, MANET nodes will have greater confidence that a new node is not malicious; if an already admitted node starts misbehaving, other MANET nodes will quickly detect and evict it.

Professors Joseph Traub and Henryk Wozniakowski were awarded a three-year grant from the NSF titled “Quantum and Classical Complexity of Multivariate Problems.” This marks 36 years in which NSF has funded every proposal Professor Traub has submitted; this may be a record for the field of Computer Science. Professors Traub and Wozniakowski also received additional funds from DARPA’s QuIST (Quantum Information Science and Technology) even though that program has ended.

Marc Eaddy spent his summer as an intern at Microsoft Research in Redmond, WA. He worked with Manuel Fadrich on an important problem for Microsoft: porting software libraries, like NET, to different platforms such as watches, cell phones, PDAs, and non-Windows operating systems. He developed automated pruning algorithms that tailor the library to the target platform by removing extraneous and unsupported features. In the process, he learned a lot about software dependencies, graph theory, and Open Classes. He took advantage of Seattle’s gorgeous summer weather (no joke!) to enjoy tennis, volleyball, Bill Gates’s lawn, and a cruise sponsored by MSR.

Ph.D. student David Elton spent the summer at the Institute for Creative Technologies, a center for virtual reality and computer simulation research at the University of Southern California. At ICT, David worked under Dr. Mark Riedl and Julia Kim, who are experts in his research area of computational narrative reasoning. As part of a larger project aiming to automate the creation of animated narrative cinema for case-based leadership training, David built a “robotic movie director” capable of realizing movie scripts as aesthetically coherent films by intelligently blocking virtual actors and placing cameras in a 3D world.

Joseph Kaptur (SEAS ’08) spent his summer in Denver, Colorado. In addition to camping and hiking, he worked for a company named Decisionengineering, which makes Crystal Ball; an add-in for Microsoft Excel that allows a user to turn any spreadsheet into a Monte Carlo simulation. Joseph spent most of his time there writing code that would allow any user of Microsoft Project (a scheduling tool for creating project plans, Gantt charts, and resource schedules) to use the power of Monte Carlo simulation to apply probability distributions to task durations, and be able to say precisely how likely a project is to be on time, two weeks late, etc. The tool he built is now in use at Fortum 500 companies and is being marketed to the Navy.

After a busy year in classrooms and labs at Columbia, CUCS students branched out to a wide range of interesting jobs and internships over the summer. Here are brief snapshots of what a few computer science students did with their summers away from campus.
Alumni News

Regina Barzilay (Ph.D. ’02), now an assistant professor of electrical engineering and computer science at Massachusetts Institute of Technology, was one of five recipients of the highly prestigious Microsoft Research New Faculty Fellowships. Because new faculty members are essential to the future of academic computing, Microsoft Research honors early-career professors who demonstrate the drive and creativity to develop original research while continuing the state of the art in computing. Regina is focusing her research on computational modeling of linguistic phenomena. She is exploring the ability of a computer to summarize information found in multiple documents that contain related information, such as news articles covering the same event. This will help readers find meaning in the ever-increasing body of information. Today, Regina graduated from Columbia University in 2002, where she was advised by Professor Cathy McKeown. Deborah Cook (Ph.D. ’06) finished her dissertation under Prof. Angelos Keromytis. She has taken a position with American Express Corporation. Blair MacIntyre (Ph.D. ’99) was awarded tenure in the College of Computing at Georgia Tech.

Smaranda Muresan (Ph.D. ’06) finished her dissertation on “Learning Constraint-based Grammars from Representative Examples.” She was advised by Professors Judith Klavans and Kathy McKeown and Dr. Owen Rambow. Smaranda is now working as a postdoc with Phil Resnik at the University of Maryland.

Ani Nenkova (Ph.D. ’05) started a tenure-track faculty position at the University of Pennsylvania. Dragomir R. Radev (Ph.D. ’99) has received tenure at the University of Michigan, where he is now an Associate Professor of Information and Computer Science and Engineering. He works on text mining with applications in areas like computer science, bioinformatics, and political science. He recently shared the 2006 Gossen Prize for Excellence in Political Methodology; the prize is awarded for the best work in political methodology presented at any political science conference during the preceding year.

Debra Cook
Advisor: Angelos Keromytis

Elastic Block Ciphers

Abstract: Standard block ciphers are designed around a small number of block sizes. From both a practical and a theoretical perspective, the question of how to efficiently support a range of block sizes is of interest. In applications, the length of the data to be encrypted is often not a multiple of the supported block size. This results in the use of plaintext-paddings schemes that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.

The focus of our research is the design and analysis of a method for creating variable-length block ciphers from existing fixed-length block ciphers. As the heart of the method, we introduce the concept of an elastic block cipher, which refers to stretching the supported block size of a block cipher to any length up to twice the original block size while incurring a computational workload that is proportional to the block size. We create a structure, referred to as the elastic network, that uses the round function from any existing block cipher in a manner that allows the properties of the round function to be maintained and results in the security of the elastic version of a block cipher being directly related to that of the original version. By forming a reduction between the elastic and original versions, we prove that the elastic version of a cipher is secure against round-key recovery attacks. Furthermore, the original cipher is secure against such attacks. We illustrate the method by creating elastic versions of four existing block ciphers. In addition, the elastic network provides a new primitive structure for use in symmetric-key cipher design. It allows for the creation of variable-length pseudorandom permutations and strong pseudorandom permutations in the range of 2b bits from round functions that are 2b bits from round functions that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.

Recent & Upcoming PhD Defenses

Debra Cook
Advisor: Angelos Keromytis

Elastic Block Ciphers

Abstract: Standard block ciphers are designed around a small number of block sizes. From both a practical and a theoretical perspective, the question of how to efficiently support a range of block sizes is of interest. In applications, the length of the data to be encrypted is often not a multiple of the supported block size. This results in the use of plaintext-paddings schemes that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.

The focus of our research is the design and analysis of a method for creating variable-length block ciphers from existing fixed-length block ciphers. As the heart of the method, we introduce the concept of an elastic block cipher, which refers to stretching the supported block size of a block cipher to any length up to twice the original block size while incurring a computational workload that is proportional to the block size. We create a structure, referred to as the elastic network, that uses the round function from any existing block cipher in a manner that allows the properties of the round function to be maintained and results in the security of the elastic version of a block cipher being directly related to that of the original version. By forming a reduction between the elastic and original versions, we prove that the elastic version of a cipher is secure against round-key recovery attacks. Furthermore, the original cipher is secure against such attacks. We illustrate the method by creating elastic versions of four existing block ciphers. In addition, the elastic network provides a new primitive structure for use in symmetric-key cipher design. It allows for the creation of variable-length pseudorandom permutations and strong pseudorandom permutations in the range of 2b bits from round functions that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.

The focus of our research is the design and analysis of a method for creating variable-length block ciphers from existing fixed-length block ciphers. As the heart of the method, we introduce the concept of an elastic block cipher, which refers to stretching the supported block size of a block cipher to any length up to twice the original block size while incurring a computational workload that is proportional to the block size. We create a structure, referred to as the elastic network, that uses the round function from any existing block cipher in a manner that allows the properties of the round function to be maintained and results in the security of the elastic version of a block cipher being directly related to that of the original version. By forming a reduction between the elastic and original versions, we prove that the elastic version of a cipher is secure against round-key recovery attacks. Furthermore, the original cipher is secure against such attacks. We illustrate the method by creating elastic versions of four existing block ciphers. In addition, the elastic network provides a new primitive structure for use in symmetric-key cipher design. It allows for the creation of variable-length pseudorandom permutations and strong pseudorandom permutations in the range of 2b bits from round functions that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.

The focus of our research is the design and analysis of a method for creating variable-length block ciphers from existing fixed-length block ciphers. As the heart of the method, we introduce the concept of an elastic block cipher, which refers to stretching the supported block size of a block cipher to any length up to twice the original block size while incurring a computational workload that is proportional to the block size. We create a structure, referred to as the elastic network, that uses the round function from any existing block cipher in a manner that allows the properties of the round function to be maintained and results in the security of the elastic version of a block cipher being directly related to that of the original version. By forming a reduction between the elastic and original versions, we prove that the elastic version of a cipher is secure against round-key recovery attacks. Furthermore, the original cipher is secure against such attacks. We illustrate the method by creating elastic versions of four existing block ciphers. In addition, the elastic network provides a new primitive structure for use in symmetric-key cipher design. It allows for the creation of variable-length pseudorandom permutations and strong pseudorandom permutations in the range of 2b bits from round functions that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.

The focus of our research is the design and analysis of a method for creating variable-length block ciphers from existing fixed-length block ciphers. As the heart of the method, we introduce the concept of an elastic block cipher, which refers to stretching the supported block size of a block cipher to any length up to twice the original block size while incurring a computational workload that is proportional to the block size. We create a structure, referred to as the elastic network, that uses the round function from any existing block cipher in a manner that allows the properties of the round function to be maintained and results in the security of the elastic version of a block cipher being directly related to that of the original version. By forming a reduction between the elastic and original versions, we prove that the elastic version of a cipher is secure against round-key recovery attacks. Furthermore, the original cipher is secure against such attacks. We illustrate the method by creating elastic versions of four existing block ciphers. In addition, the elastic network provides a new primitive structure for use in symmetric-key cipher design. It allows for the creation of variable-length pseudorandom permutations and strong pseudorandom permutations in the range of 2b bits from round functions that impose computational and space overheads. Furthermore, a variable-length block cipher ideally provides a variable-length pseudorandom permutation and strong pseudorandom permutation, which are the building blocks of practical block ciphers and correspond to ideal properties for a block cipher.
Recent & Upcoming PhD Defenses (continued)

Third, we present a multi-stock automated trading system that includes a machine learning algorithm to detect the formation, a weighting algorithm that combines the experts, and a risk management layer that limits the only the strongest prediction and provides a stop-loss protection history of negative performance. This algorithm was tested with 108 real stock datasets and found to substantially increase transaction costs.

Rui Kuang
Abstract: Christina Leslie
Inferencing Protein Structure with Discriminative Learning and Network Diffusion

Abstract: As the complete genomes of more and more species become available, the large scale protein structure prediction is becoming increasingly important. Given the difficulty of experimental determination of protein structure with X-Ray crystallography or nuclear magnetic resonance (NMR), much effort over the past decade has been devoted to developing new machine learning approaches for tackling the challenging problem of protein structure prediction. In this thesis, we focus on applying several advanced starting techniques, including kernel methods and network diffusion algorithms, to address four fundamental learning problems in protein structure inference:

- Classification of a new protein with its structural class. To detect subtle sequence similarities between remote homologs, we propose a multiple sequence alignment, which models a pairwise alignment method such as PSI-BLAST.
- Predicting functional domains or sequence segments, especially motif-rich regions induced by the propagation are also helpful for discovering conserved structural components in remote homologs.
- Prediction of dihedral torsion angles of protein backbone. We apply kernel methods to accurately predict protein backbone torsion angles, which usually help to substantially improve the modeling of local structures of protein sequence segments, especially the loop conformations, which do not form regular structural motifs.
- Segmentation of multi-domain protein. We predict protein domain labels and boundaries by taking advantage of a new optimization problem, in which we use trained SVM domain recognizers to find the optimal segmentation of a protein giving the largest sum of classification scores.

Our work provides efficient and effective solutions to these four fundamental learning problems in protein structure inference, validated with large scale experiments. The proposed methods achieve high prediction performance and capable of handling huge datasets, which is crucial for developing this increasingly important in this post-genome era.

Smaranda Munteanu

Advisors: Owen Rambow and Judith Klavans
Learning Constraint-based Grammars from Large-Scale Data
Abstract: Computational efficient models for natural language understanding can have a wide variety of applications starting from text mining and question answering, to natural language interface applications. Constraint-based grammar formalisms have been widely used for natural language processing. Yet, one serious obstacle for their successful application is the fact that these formalisms have been overlooked as an important requirement: learning parameters. Here we consider a poor match between these grammar formalisms and existing learning methods.

This dissertation defines a new type of constraint-based grammars, Lexicalized Well-Founded Grammars (LWFGs), which allow deep language understanding and are easy to learn. These grammars model both syntax and semantics and have constraints at the level for semantic composition and semantic interpretation. The interpretation constraints allow access to deep meaning during language processing. They establish links between syntactic aspects and the entities they refer to in the real world. We use an ontology-based knowledge base, populated in a semantic representation that can be conceived as an ontology query language. This representation is sufficiently expressive to represent many aspects of language and yet sufficiently restrictive to support learning and tractable inference.

In this thesis, we propose a new relation model, which achieves high prediction performance and capable of handling huge datasets, which is crucial for developing this increasingly important in this post-genome era.

Alejandro Troccoli

Abstract: Peter Allen
New Methods and tools to manipulate large scale outdoor scenes using range and color images
Abstract: We present two new extensions of photorealistic models using range scans and digital photographs to become increasingly popular in a wide field of ranges, from reverse engineering to cultural heritage preservation. These scenes employ a range finder to acquire the geometry information and a digital camera to measure color detail. But the traditional approach of using range scans and color images to produce photorealistic images has a fundamental problem: the scene model is still an area of research with many unsolved problems.

In this dissertation we present two new approaches for creating digital models from range and color images, with emphasis on large-scale outdoor scenes. First, we address the problem of range and color image registration. In this area, we introduce a semi-automatic tool for range and color image registration that makes use of line-features to solve for the position and orientation of the digital camera. This allows us to efficiently register images of urban scenery. Secondly, we present a registration technique that uses the shadows cast by the sun as a new feature to aid in the correct registration, which we then use to relight images. In this area, we use the technique for image registration that makes use of overlap of a pair of images, which we then use to produce a new image registration technique that computes a model of the environment, which we then use to register images of urban scenery.

Ke Wang

Abstract: Salvatore Stoilo
Network Detection and Anomaly-based Intrusion Detection

Abstract: Every computer on the Internet nowadays is a potential target for a network-based cyber attack moment. The pervasive use of signature-based anti-virus scanners and misuse detection Intrusion Detection Systems have failed to provide adequate protection against a constant barrage of “zero-day” attacks. Such attacks may demand de-service, system crashing and security breaches resulting in the loss of critical information. In this dissertation, we consider the problem of detecting these “zero-day” intrusions quickly and accurately upon their very first appearance.

Most current Network Intrusion Detection Systems (NIDS) use simple features such as protocol, header and packet length and have a high false positive rate. Hence, in this thesis we consider the problem of detecting these “zero-day” intrusions quickly and accurately upon their very first appearance.

An often-cited weakness of anomaly detection systems is that they suffer from the “false positive” problem. Clever adversaries may craft attacks that appear normal to an anomaly detection system and go unnoticed as a false negative. A mimic attack on a system often looks benign. Anomaly detection systems are otherwise very effective at detecting malicious events that indicate a system is under attack. These approaches, however, are blind to the content of the packet stream, and in particular, the packet content delivered to a service or application. Such a system exploits the vulnerable application software. We conjecture that fast and efficient detectors that focus on network packet content anomaly detection will help identify zero-day attacks far more accurately than approaches that are oblivious to the header information.

We therefore present two new methods for network packet content anomaly detection: PAYL and Anagram, for network intrusion detection. We are otherwise very effective at detecting packets that are otherwise packets that are anomalous in some way. We conjecture that these approaches will help identify zero-day attacks far more accurately than approaches that are oblivious to the header information.

We therefore present two pay- load-based anomaly detectors: PAYL and Anagram, for intrusion detection. They are designed to detect attacks that are otherwise normal connections except that the packets carry bad (anom- alous) content. These attacks are otherwise anomalous behavior. We conjecture that these approaches will help identify zero-day attacks far more accurately than approaches that are oblivious to the header information.

We therefore present two new methods for network packet content anomaly detection: PAYL and Anagram, for intrusion detection. They are designed to detect attacks that are otherwise normal connections except that the packets carry bad (anomalous) content. These attacks are otherwise anomalous behavior. We conjecture that these approaches will help identify zero-day attacks far more accurately than approaches that are oblivious to the header information.

We therefore present two new methods for network packet content anomaly detection: PAYL and Anagram, for intrusion detection. They are designed to detect attacks that are otherwise normal connections except that the packets carry bad (anomalous) content. These attacks are otherwise anomalous behavior. We conjecture that these approaches will help identify zero-day attacks far more accurately than approaches that are oblivious to the header information.
Generating robust signatures.

requiring new approaches for shown to have limited value, individual packet content delivery these attacks, correlating the anomaly detection techniques filter newly discovered zero-day “string-based” signatures to automatically generating A new wave of cleverly crafted against zero-day attacks. producing accurate mitigation reducing incorrect decisions and from the set of anomaly alerts, alerts can identify true positives into characterizing traffic of short-term web attacks. This algorithm provides insight into identifying malicious applications and helps to prevent web server overloading, and is useful for server web operators.

Besides building stand-alone anomaly sensors, we also demonstrate a collaborative security strategy whereby different hosts may exchange payload alerts to increase the accuracy of their local sensors and reduce false positives. We propose and examine several new approaches to enable the sharing of suspicious payloads via privacy-preserving technologies. We detail the work we have done with PAWS and Anagram to support generalized payload correlation and signature generation without releasing identifiable payload data. The important principle demonstrated is that correlation of multiple alerts can identify true positives from the set of anomaly alerts, thus reducing incorrect decisions and producing accurate mitigation against zero-day attacks.

A new wave of cleverly crafted polymorphic attacks has significantly complicated the task of automatically generating “string-based” signatures to detect newly discovered zero-day attacks. Although the payload anomaly detection techniques we present are able to detect these attacks, correlating the individual payload content delivering distinct instances of the same polymorphic attack are shown to have limited value, requiring new approaches for generating robust signatures.

Recent & Upcoming PhD Defenses (continued)
Alfred V. Aho is Lawrence Gussman Professor of Computer Science and Vice Chair for Undergraduate Education in the Computer Science Department. He served as Chair of the department from 1995 to 1997 and in the spring of 2003. He was elected to the U.S. National Academy of Engineering for contributions to the fields of algorithms and programming tools. Professor Aho received his B.A.Sc in Engineering Physics from the University of Toronto and a Ph.D. in Electrical Engineering/Computer Science from Princeton University. Prior to his current position at Columbia, Professor Aho served in many capacities at the Computing Sciences Research Center at Bell Labs, such as Vice President, Director, department head, and member of technical staff. This is the lab that invented UNIX, C and C++. Al was also the General Manager of the Information Sciences and Technologies Research Laboratory at Bellcore (now Telcordia). Professor Aho is the “A” in AWK, a widely used pattern-matching language (you can think of AWK as the initial pure version of perl). “W” is Peter Weinberger and “K” is Brian Kernighan. Al also wrote the initial versions of the string pattern-matching programs egrep and fgrep that first appeared on UNIX. His current research interests include quantum computing, programming languages, compilers, and algorithms.

At this year’s computer science department “Hello Meeting,” members of the faculty recalled the early days of the department when the same meeting could be held on the first floor of Mudd, in the small area that is now used for making sandwiches in the Carleton Lounge. Since then, the computer science department has grown both in size and prominence. The faculty now includes six members of the National Academies: Al Aho, Steven M. Bellovin, Zvi Galil, Ted Shortliffe, Joe Traub, and Vladimir Vapnik. Here is a brief overview of their achievements and contributions.