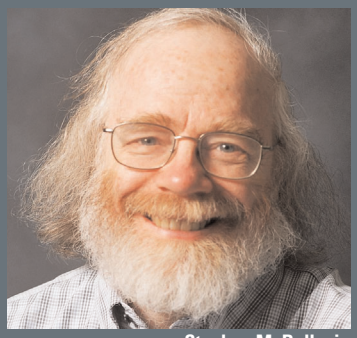
CS@CU

NEWSLETTER OF THE DEPARTMENT OF COMPUTER SCIENCE AT COLUMBIA UNIVERSITY VOL.3 NO.1 FALL 2006


Alfred V. Aho


Stephen M. Bellovin

Columbia Computer Science Faculty in the **National Academies**


Zvi Galil


Edward H. Shortliffe


Joseph F. Traub


Vladimir Vapnik

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.

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.

(continued on next page)

Professor Aho has won the IEEE John von Neumann Medal and has received honorary doctorates from the Universities of Helsinki and Waterloo. He is a Member of the American Academy of Arts and Sciences and a Fellow of the American Association for the Advancement of Science, ACM, Bell Labs, and IEEE. He received the Great Teacher Award for 2003 from the Society of Columbia Graduates.

Steven M. Bellovin is a Professor of Computer Science. He was elected to the National Academy of Engineering for contributions to network applications and security.

Professor Bellovin joined the faculty in 2005 after many years at Bell Labs and AT&T Labs Research, where he was an AT&T Fellow. He was first exposed to computers in 10th grade at Stuyvesant High School and has been working in the field ever since. He received a BA degree from Columbia University, and an MS and PhD in Computer Science from the University of North Carolina at Chapel Hill. As a graduate student, he helped create Netnews; for this, he and his fellow contributors were awarded the 1995 Usenix Lifetime Achievement Award.

Bellovin is the co-author of the book "Firewalls and Internet Security: Repelling the Wily Hacker," and holds several patents on cryptographic and network protocols. His current research interests include security and privacy issues in large scale and distributed networks. He is also interested in human factor aspects of security.

Professor Bellovin has served on many National Research Council study committees, including those on information systems trustworthiness, the privacy implications of authentication technologies, and cybersecurity research needs. He was also a member of the information technology subcommittee of an NRC study group on science versus terrorism. He was a member of the Internet Architecture Board from 1996-2002, and was co-director of the Security Area of the IETF from 2002 through 2004. He sits on the Department of Homeland Security's Science and Technology Advisory Board.

Zvi Galil is the Dean of the School of Engineering and Applied Science. He is the Julian Clarence Levi Professor of Mathematical Methods and 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 for leadership 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 as an extension of his interest 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 kind of problem or use 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 being the Dean of 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 Rolf A. Scholdager 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 awarded a Ph.D. in Medical Information Sciences in 1975 and an M.D. in 1976. Subsequently, he served as Professor of Medicine and of Computer Science at Stanford University. In January 2000 he assumed his new post at Columbia University, 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 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 1979, 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. He 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 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 founding chair of the Computer Science Department. From 1971 to 1979 he was Head of the Computer Science Department at Carnegie Mellon University. He served as founding Chairman of the Computer Science and Telecommunications Board (CSTB) of the National Academies from 1986 to 1992 and is again serving as Chair. CSTB deals with critical national issues in computing, communications, and public policy (see www.cstb.org). He has served as founding Editor-in-Chief of the Journal of Complexity since 1985.

He is the author, co-author, or editor of nine books and has published some one hundred and twenty papers. His latest book, Complexity and Information, co-authored with Professor Arthur G. Werschulz, was published by Cambridge University Press in 1998.

Traub and Professor Henryk Wozniakowski started the field of information-based complexity (IBC), which is now an active research area involving researchers from many countries. IBC studies optimal algorithms and computational complexity for the continuous problems which arise in physical science, mathematical finance, engineering and economics. A major focus of his current work is quantum computing.

He has received numerous honors including election to the National Academy of Engineering in 1985, the Emanuel R. Piore Medal from IEEE in 1991, selection by the Accademia Nazionale dei Lincei in Rome to present the 1993 Lezione Lincei, and the 1999 Mayor's Award for Excellence in Science and Technology in New York City. In 2001, he received an honorary Doctorate of Science from the University of Central Florida.

Vladimir Vapnik is Professor of Computer Science at Columbia and Fellow at NEC Labs. 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 1990 he worked at the Institute, where he became Head of the Computer Science Research Department. He then joined AT&T Bell Laboratories, 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 minimizing the expected risk of predictors using empirical data, and a new type of learning machine (the Support Vector Machine) that possesses a high level of generalization ability. These techniques have been used to solve many pattern 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) worlds 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 as an intermediate step. 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 giant shoulders to stand on.

Feature Article

Professor Shree Nayar Named as **2006 Great Teacher**

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

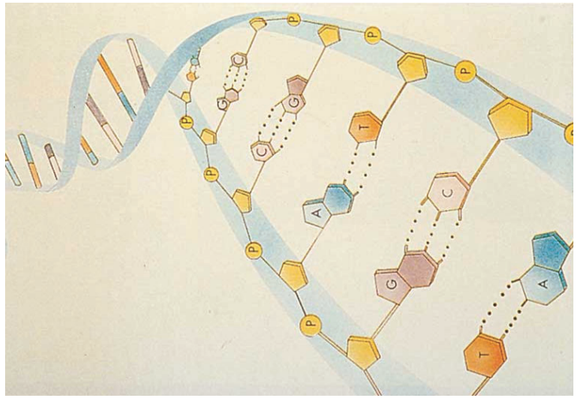
Professor Shree Nayar was named as the recipient of the 2006 Great Teacher Award for the School of Engineering at Columbia University. The award is bestowed by the Society of Columbia Graduates. Its Board of Directors named Professor Nayar for this award because it feels that he exemplifies the greatest 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 joins the ranks of Columbia's finest and most beloved professors, such as Mark Van Doren, Lionel Trilling, Mario Salvadori, Morton Friedman, Rene Testa, and others. The Great Teachers 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 100 year 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), to
- The virtualization layer (hardware, software hypervisors, paravirtualization), to
- The operating system (advanced topics: workload management, recovery, availability, reliability, servicability, metering), to
- 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 digital 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 Itsik 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 biomedical 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, eigenvalue 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 security issues: impersonation and authentication, privacy;
- VoIP spam ("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.

A screenshot from a student project in COMS 4995-2, **Video Game Design and Development**.

Adjunct Professors Michael Theobald and **Franjo Ivancic** 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. The 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 replay 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.

Computer Science Department Welcomes **Professors Dana and Itsik Pe'er**

Professor Dana Pe'er

Professor Itsik Pe'er

The Computer Science department is delighted to welcome Dana and Itsik Pe'er as Assistant Professors of Computer Science.

Professor Dana Pe'er uses machine learning techniques to develop computational tools and models to understand how biological networks process signals and execute complex biological functions. Professor Pe'er is interested in understanding the design, function and evolution of molecular networks. Molecular networks regulate the basics processes of life, for example, a sequence of developmental decisions regulate the formation of an entire human body from a single zygote cell. Understanding the normal function of the molecular network will allow us to understand how its "dysfunction" leads to many diseases including cancer and autoimmunity.

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 combine 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 does dysregulation cause disease and where are the optimal points for corrective interference and cure.

Dana Pe'er received a B.Sc. in mathematics (1995), an M.Sc. in theoretical computer science (1999) 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 world-wide. To strengthen her cross training in biology, following her graduation, Dr. Pe'er joined the laboratory of Professor George Church in the Genetics department 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 bachelors, 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 studies, develops, and applies novel computational methods in human genetics. How is it best to measure, 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 optimal implementation and analysis of whole genome association studies, and in the interplay between somatic and germline variation.

Mansoor Alicherry completed his B.S. in Computer Science from National Institute of Technology, Calicut, India in 1997. He worked in Wipro from 1997-1998. He did his M.S. in Computer Science from Indian Institute of Science (IISc), Bangalore in 2000, where he specialized on operating systems. He has been working in Lucent Technologies' Bell Labs since 2000. His research interests are in network security and routing and design algorithms for optical and mobile networks. He is currently working with Professor Angelos Keromytis on network security.

Stuart Andrews began working with Professor Tony Jebara as a postdoc in the Fall of 2006. He hold B.Sc. and M.Sc. degrees from the University of Toronto and shortly expects to receive his Ph.D. from Brown University. His research focuses on semi-supervised and structured-output learning algorithms. At Columbia, he and Professor Jebara are developing algorithms that learn to predict properties of edges in a network of N entities, based on attributes of the entities as well as the global network structure. When applied to a social network, such as the kind maintained by students in a university setting, this algorithm learns how to assign in a balanced fashion a set of likely friends and/or mentors for the incoming class, from the friendship structure of the existing students. Other applications of these methods include network biology and e-commerce.

Hila Becker graduated from Stony Brook University with a B.S. in Computer Science and Applied Math and Statistics in May 2004. She got her M.S. in Computer Science from Columbia University in 2005 and started as a Ph.D. student in the spring of 2006. Hila's main research interests include theoretical and applied Machine Learning and Information Retrieval. She is currently working in the Center for Computational Learning Systems with Dr. Marta Arias on online-learning techniques for prediction of electricity distribution feeder failures.

Fadi Biadsy graduated from Ben-Gurion University with a B.S. in Computer Science and Mathematics in July 2002. He worked at Dalet Digital Media Systems from 2001 to 2005 as a software architect and developer. He received his M.S. in computer science at Ben-Gurion University with highest honors in 2005; he explored the problem of online handwriting recognition for Arabic script using hidden Markov models. He started as a Ph.D. student in the Department of Computer Science at Columbia University in Spring of 2006. Fadi's main research interests focus on spoken language processing; particularly, identifying charismatic speech, Arabic dialect identification using prosody, and the acoustic/prosodic and intonational structure of Arabic language. He is currently working with Professor Julia Hirschberg's Spoken Language Processing group studying the acoustic/prosodic and lexical characteristics of Arabic and American English Charismatic speech.

Malek Ben-Salem graduated from the University of Hanover, Germany with a B.S. and M.S in Electrical Engineering in 1999 and 2001 respectively. She worked at IBM Corporation from 2001-2006, did an M.S. in computer science at Columbia from 2003-2005, and started as a Ph.D. student in Fall 2006. Malek's main research interests are Computer Security and Machine Learning. She is currently working with Professor Stolfo's group studying Insider Intrusion Detection.

Sasha P. Caskey graduated from Brandeis University with a BA in 1997. After graduation he joined The MITRE Corporation in Bedford, MA where he contributed to research and development in Collaboration Systems (CVS, JCS, SIMP), Natural Language Processing and Spoken Dialog Systems (DARPA Communicator). In 2000 he joined a startup, Speechworks Int., where he worked in product development and research on spoken dialog systems. He joined IBM T.J. Watson Research labs in 2003 and is a working first year Ph.D. Student under the supervision of Professor Kathy McKeown in the NLP group.

Oliver Cossairt graduated with a B.S. from Evergreen State College in June 2001. He did an M.S. in Media Arts and Sciences at MIT from 2001-2003. After graduating from MIT, he worked at Actuality Systems for three years before he began his PhD at Columbia University in the Fall of 2006. Oliver's main research interest is in computation displays and cameras, specifically in Three-Dimensional displays and light fields. He is currently working with Shree Nayar in the Vision and Graphics Group.

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 2006, and is working with Professor Ravi Ramamoorthi's Computer Graphics group.

Ioannis Giannakakis graduated from National Technical University of Athens (NTUA) with a Diploma in Electrical and Computer Engineering in July 2006. He started as a Ph.D. student at Columbia University in Fall 2006. Ioannis's main research interests are Algorithms and Databases. He is currently working with Professor Kenneth Ross's group in Database systems.

Dana Glasner graduated from Yeshiva University with a B.A. in Computer Science and Math in May 2006 and started as a Ph.D. student at Columbia in Fall 2006. Dana's main research interests are Cryptography, Learning Theory, and Complexity Theory. She is currently working with Professor Tal Malkin and Professor Rocco Servedio.

Roni Goldenthal is a Computer Science PhD student at the Hebrew University of Jerusalem, where his advisor is Professor Michel Bercovier. His research interests are geometric modeling and physical simulation. Currently he is part of the computer graphics group at Columbia, where he works with Professor Eitan Grinspun on a new algorithm for physical simulation of cloth.

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 Eitan Grinspun as part of the Columbia Computer Graphics Group. His research interests include physical simulation, computer animation, and computer graphics.

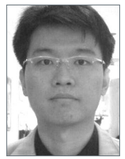
Steve Henderson started his Ph.D. studies at Columbia in Fall 2006. He is working with Professor Feiner in the Computer Graphics and User Interface Lab. Steve joins Columbia from the faculty of the United States Military Academy at West Point, New York where he served as an Assistant Professor of Systems Engineering. 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, social networks, large-scale database design, and optimization.

Tie Hu is a postdoc in the Robotics lab working with Professor Peter Allen. He received a Ph.D in Mechanical Engineering from Drexel University in 2006, and worked as an automation engineer in China from 1999 to 2001. His research interests are medical robotics, 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.

Bert C. Huang graduated from Brandeis University with a B.S. in Computer Science and a B.A. in Philosophy in 2004. 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 Computational Learning Systems and in the Machine Learning Laboratory.

Maritza Johnson graduated from the University of San Diego with a B.A. in Computer Science in May of 2005. Her research interests are user-centered design, human-computer interaction, and security. She is currently working with Professor Steven Bellovin on designing a method of evaluating a financial institute's ability to authenticate itself to users in an online environment. In the past she has been active in the women in computer science community and plans to continue doing so while at Columbia.

Aly A. Khan graduated from Carnegie Mellon with a M.S. in Computational Biology in May 2006. He completed his MS thesis under Professor Russell Schwartz on modeling a novel biological mechanism for splicing of very large introns in DNA. He was a Howard Hughes Medical Institute undergraduate fellow while working in the lab of Professor Chris Bailey-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 Sander's lab at Memorial Sloan-Kettering, on studying microRNA regulation.



Jong Yul Kim came to Columbia University from South Korea after graduating from Seoul National University with a B.S. in Computer Science and Engineering in February 2005. He finished his M.S. in Computer Science in May 2006, and started his Ph.D. program in Fall 2006. Since the summer of 2005, Jong Yul has been working with Professor Schulzrinne on the Next Generation 9-1-1 project.

Jae Woo Lee graduated from Columbia College with a B.A. in Physics in 1994. After working 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, received an M.S. in computer science in 2006, and started as a Ph.D. student in Fall 2006 under Professor Henning Schulzrinne. He is currently investigating adaptive algorithms suitable for transient peer-to-peer communication networks.

Chris Murphy is working with Prof. Gail Kaiser in the Programming Systems Lab and is investigating the software testing and quality assurance of machine learning algorithms. He graduated from Boston University with a BS in Computer Engineering in 1995, and completed his MS in Computer Science at Columbia this past spring. Chris worked as a software architect and consultant for six years; prior to studying at Columbia he taught English essay writing for two years in Seoul, South Korea. Chris enjoys teaching and was the winner of the Andrew P. Kosoresow Memorial Award for Excellence in Teaching and Service in 2006.

Richard Neill holds a B.S and M.S. in Electrical Engineering and has held a technology position at Cablevision Systems Corporation. He started his Ph.D. in Fall 2006 working in Prof. Carloni's research group. Richard's main research interests are in distributed embedded systems, parallel architectures, and grid computing.

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. Ohan's main research interest is Augmented Reality and Virtual Reality. He is currently working with Professor Steven Feiner's group developing a framework that will be useful for creating 3D User Interface and Augmented Reality applications using an existing commercial game engine.

Kumiko Ono graduated from Ochanomizu University, Japan with a B.S. in Mathematics. She has worked as a network research engineer at NTT Labs. for over ten years, and visited Columbia in 2005. She started as a Ph.D. student in Spring 2006. Her main research interests are VoIP network and security. She is currently working with Professor Schulzrinne's group.

Kristen Parton graduated from Stanford University with a B.S. in Computer Science in June 2003. She worked on Intelligent Tutoring Systems at Stottler Henke Associates, Inc. from 2003-2005, then on multilingual web search at Yahoo! in 2006. She started as a Ph.D. student in the Natural Language Processing group in Fall 2006, and will be working with Professor Kathy McKeown and Owen Rambow.

Mariana Raykova graduated from Bard College with a double B.A. in Computer Science and Mathematics in 2006. Her main research interests are in the areas of cryptography and security. She is working with Moti Yung on cryptographic hash functions.

Lily Bao Secora joined the Computer Science Department in April 2006 as the Ph.D. Program Administrator. She manages the day-to-day operations of the PhD program. She has a Masters of Education degree in Counseling and Personnel Services and has worked for five years in higher education. Her hobbies include rollerblading, tennis, volleyball, cooking, and traveling.

Wonsang Song graduated from Seoul National University with a B.S. in Computer Engineering in February 1997. He worked as a system engineer and programmer in Korea from 1997-2004, did an M.S. in Computer Science at Columbia from 2004-2006, and started as a Ph.D. student in Fall 2006. Wonsang's main research interests are multimedia networking, location-based services and distributed systems. He is currently working on the NG911 project with Professor Henning Schulzrinne.

Hang Zhao graduated from National University of Singapore with a B.S. in Computer Science in 2005. She joined the Computer Science Department at Columbia in Fall 2006 as a PhD student. Hang's main research interest is network security. She is currently working with Professor Steven Bellovin on the Policy Based Security Management project collaborating with IBM Research and other universities from the U.S. and the U.K.

Alp Atici (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 award is for the paper "Learning Unions of $\omega(1)$-Dimensional Rectangles" which is co-authored by Professor **Rocco Servedio**.

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

Professors **Steven Bellovin** and **Sal Stolfo**, together with Professor Sean Smith of Dartmouth, received a grant from ARO to organize and run a Workshop on Insider Threat Research. The workshop will be held in Washington, DC in June 2007.

Professor **Steven Bellovin** was interviewed by Robert Siegel on NPR's popular program "All Things Considered" in May. Steven discussed the National Security Agency's efforts to analyze the huge databases of domestic phone calls turned over to the NSA by phone companies.

Professor **Steven Bellovin** was interviewed for an NY1 TV article on security and terrorism.

Matthew Burnside, **Jae Woo Lee**, **Christian Murphy** and **Joshua Reich** and were named as "extraordinary TAs" for the Fall 2005 semester, based on the evaluation of students in their classes. Congratulations to our outstanding TAs!

Professor **Luca Carloni**, along with Professor Ken Shepard of the Electrical Engineering department, was awarded a three-year NSF grant to study the design of low-power scalable communication networks for multi-core systems-on-chip. The project is funded by the NSF Foundations of Computing Processes and Artifacts (CPA) Cluster; in 2005 the NSF CPA cluster received 532 proposals and funded approximately 10% of them.

Professor **Luca Carloni** has become a member of the MARCO-sponsored, multi-institution Gigascale System Research Center (GSRC) where he joins the "Core Design Technology for Complex Heterogeneous Systems" theme. The Microelectronics Advanced Research Corporation (MARCO) funds and operates university-based research centers in microelectronics technology. Its charter initiative, the Focus Center Research Program (FCRP), is designed to expand pre-competitive, cooperative, long-range applied microelectronics research at U.S. universities. Each Focus Center targets research in a particular area of expertise. The GSRC Focus Center brings together 41 faculty from 17 American universities to focus on pertinent problems the semiconductor industry faces in the next decade in the areas of system design, integration, test, and verification. Professor Carloni will work on the development of correct-by-construction methodologies for the integrated design and validation of robust gigascale systems. The ultimate goal of this project is to assist engineering teams coping with the complexity of gigascale system design by addressing two critical challenges: how to reduce the design productivity gap and how to handle the many sources of design variability introduced by nanometer technology processes.

Cristian Soviani, **Olivier Tardieu**, and Professor **Stephen Edwards** won a best paper award at DATE (Design Automation and Test in Europe), one of the major design automation conferences. The conference had a 28% acceptance rate; only two such awards were given out for the 800+ submissions they had this year.

Professor **Stephen Edwards** was awarded an NSF grant titled "SHIM: Developing Embedded Systems with Deterministic Concurrency": 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, multi- and variable-rate dataflow, and data-dependent decisions.

Members of the department's computer graphics group were extremely successful in the upcoming SIGGRAPH 2006 conference, having 10 papers accepted, the most from any single institution in the last five years. SIGGRAPH is the most prestigious conference for computer graphics; the 2006 conference took place in Boston, Massachusetts in August 2006. A total of 86 papers were accepted from 474 submissions. Authors from Columbia University include Prof. **Ravi Ramamoorthi**, Prof. **Shree Nayar**, Prof. **Eitan Grinspun**, and Prof. **Peter Belhumeur**, along with their graduate students. More information about the Columbia Vision + Graphics Center can be found at *http://www.cs.columbia.edu/cvgc/*

The paper "To Search or to Crawl? Towards a Query Optimizer for Text-Centric Tasks," by **Panos Ipeirotis**, **Eugene Agichtein**, **Pranay Jain**, and Professor **Luis Gravano**, received the "Best Paper" Award at the SIGMOD 2006 Conference. SIGMOD is one of the two premier database conferences and is highly selective; the acceptance rate for SIGMOD 2006 was lower than 15%.

Professor **Eitan Grinspun** received one of three "best paper" awards at EUROGRAPHICS 2006 in early September. EUROGRAPHICS is the European cousin of SIGGRAPH and one of the most important computer graphics conferences. The paper, titled "Computing discrete shape operators on general meshes," was co-authored by two NYU students, Yotam Gingold and Jason Reisman, and Prof. Denis Zorin (NYU). Eurographics considered 246 submitted papers, and accepted 42.

Professor **Eitan Grinspun** was awarded an NSF grant together with Professors Karamcheti and Zorin at NYU. The three-year project will develop parallel architectures for interactive scientific computing, with a focus on engineering and biomedical design-space exploration. Specific points of focus include (1) higher-level user control of the overall study (as opposed to individual experiments); (2) reuse of data from prior experiments in carrying forward new computations; (3) dynamic management of system resources by relying on a tighter coupling between application and system software; and (4) software reuse based on common component architecture (CCA) compliance and standardization of a more permeable system-/solver-level interface. The architecture will be evaluated on real-world biomedical applications, with a specific focus on natural incorporation of existing simulation, solver, and domain-specific codes.

Computer Science PhD student **Alex Haubold** won the Best Poster Award for this year's IEEE International Conference and Multimedia Expo (ICME'06). This annual conference, one of the most significant and largest in the area of multimedia, featured over 270 posters. Alex, who is Professor John Kender's student, won for his paper reporting on research he did as part of his IBM internship last summer: "Semantic Multimedia Retrieval using Lexical Query Expansion and Model-Based Reranking".

Professor **Julia Hirschberg**, along with **Martin Jansche** (of the Columbia Center for Computational Learning Systems) and colleagues at UIUC, received an NSF grant to improve the tutoring of the rhythmic and intonational aspect of language (prosody) for those learning Chinese and English as a second language. Prosody is an integral part of human communication, but one that second-language learners rarely learn. Topic shifts, contrastive focus, and even simple question/statement distinctions, cannot be recognized or produced in many languages without an understanding of their prosody. However, 'translating' between the prosody of one language and that of another is a little-studied phenomenon. 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 featured in an article in the International Herald Tribune on text-to-speech technology.

Professors **Julia Hirschberg** and **Jason Nieh**, along with Professor **Cliff Stein** (joint with IEOR), received the prestigious IBM Faculty Award, reflecting their technical achievements and close interaction with IBM researchers.

Professor **Tony Jebara** received his third KDD grant, titled "Learning to Match People, Multimedia and Graphs via Permutation". This proposal undertakes a novel research direction that explores matching and permutation within statistical learning. These research tools have applications in national security as a way to identify and match people from text and multimedia and discover links between them. More specifically, this proposal addresses the following key application areas:

– Matching authors: permutational clustering methods and permutationally invariant kernels are used to compute the likelihood the same person wrote a given publication or text.
– Matching text and multimedia documents: permutational algorithms and permutationally invariant kernels to perform text, image and word matchings of descriptions of people to known individuals in a database.
– Matching social networks and graphs: social network matching tools from permutational algorithms which find a subnetwork in a larger network that has a desired topology.

Professors **Gail Kaiser**, **Angelos Keromytis** and **Sal Stolfo** won a National Science Foundation (NSF) four-year grant as part of the CyberTrust program to study collaborative self-healing systems. Collaborative Self-healing Systems (COSS) is a new paradigm for protecting software systems. Software monocultures are widely used applications that share common vulnerabilities. Hence, any attack that exploits one instance of a vulnerable application provides the means for wide-spread damage. The emerging concept of collaborative security, wherein independent but cooperative entities form a group to improve their individual security, provides the opportunity to exploit the homogeneity of a software monoculture for collective and mutual protection.

Professor **John Kender**, together with **Shih-Fu Chang** of Columbia EE and **Tom Huang** of UIUC, is a co-PI on a three year, ten senior researcher grant headed by PI John Smith of IBM, funded out of the Disruptive Technologies Office as part of their Video Analysis and Content Extraction (VACE) program. The project pursues research on statistical modeling techniques that will characterize video contents in large semantic spaces, using open source international news broadcasts. It emphasizes cross-domain and cross-cultural scalability, faster than real-time performance, and the exploitation of the temporal evolutionary aspects of video contents. It will build a retrieval workbench with video mining, topic tracking, and cross-linking capabilities, along with other video understanding services.

Professor **Angelos Keromytis** chaired the security tracks for the 2007 World Wide Web conference (WWW) and for the 2007 International Conference on Distributed Computing Systems (ICDCS).

Ph.D. students **Homin Lee** and **Andrew Wan** received the Mark Fulk Best Student Paper award at the 19th Annual Conference on Learning Theory (COLT 2006), held in Pittsburgh, PA, in July. The award is for their paper titled "DNF 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 several years.

Ph.D. student **Jackson Liscombe** won one of three "best student paper" awards at the INTERSPEECH 2006 conference. The paper was co-authored by postdoctoral researcher **Jennifer Venditti** and Professor **Julia Hirschberg**. The paper was titled "Detecting Question-Bearing Turns in Spoken Tutorial Dialogues." INTERSPEECH is the annual conference of the International Speech Communication Association *(http://www.isca-speech.org)*, which has about 1500 members. The conference is held annually, this year in Pittsburgh. There are usually about 1100 attendees and approximately 1000 submissions. ISCA is one of the major speech science and technology organizations internationally.

Professors **Vishal Misra** (Computer Science), **Dan Rubenstein** (Electrical Engineering and CS) and **Ed Coffman** (EE and CS), together with Professor Predrag Jelenkovic and Professor Mor Harchol-Balter (of 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 derives point solutions, where different resource/consumer configurations require separately-designed sharing mechanisms. This project seeks to develop and analyze Adaptive Sharing Mechanisms (ASMs) in which the mechanism used to share resources adapts 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 Computer 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 control plane 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 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/~kar/fastqueryproj.html* 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 unsolicited 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 Stolfo** 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 wireless 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.

"What I did with my **summer** vacation"

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.

An image from the intelligent movie director, courtesy Institute for Creative Technologies.

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 Elson** 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 Decisioneering, 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 Fortune 500 companies and is being marketed to the Navy.

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 continually advancing the state of the art of 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 available today. Regina graduated from Columbia University in 2002, where she was advised by Professor Kathy McKeown.

Deborah Cook (Ph.D. '06) finished her dissertation under Prof. Angelos Keromytis. She has taken a position with Bell Labs.

German Creamer (Ph.D. '06) finished his dissertation under Prof. Sal Stolfo. He is working in a management position at American Express Corporation.

Min-Yen Kan (SEAS B.S. '96, Ph.D. '02) married Alicia Ang in a civil ceremony at the Raffles Hotel in Singapore on December 23, 2005. Min is currently an assistant professor at the National University of Singapore, in the Department of Computer Science, researching digital libraries.

Rui Kuang (Ph.D. '06) finished his dissertation under Dr. Christina Leslie of the Center for Computational Learning Systems. He is now an Assistant Professor at the University of Minnesota.

Jonathan Liu (CC '04) writes: 'Currently I am working at PricewaterhouseCoopers doing consulting work. My work is primarily accounting-based consulting and usually involves a technology aspect. Still keeping in touch with **Jonathan Tse** (CS, SEAS '04), **Tim Soohoo** (CS, SEAS '04), and **Jeff Eng** (CS, SEAS '04)!'

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 Gosnell 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. He was elected as secretary of the ACL, the international organization for Computational Linguistics and Natural Language Processing, and will be the program chair for the upcoming US Olympiad in Computational Linguistics. He is on leave in the 2006-7 academic year and is spending some of that time at Columbia.

Jonathan Rosenberg (Ph.D. '01) was named a Cisco Fellow. The Cisco Fellow program was created to honor Cisco's most distinguished engineers; the title of 'Cisco Fellow' is the company's highest honor for engineering excellence. Only ten Cisco employees have been named as Cisco fellows out of nearly 40,000 employees worldwide.

Ke Wang (Ph.D. '06) finished her dissertation under Prof. Sal Stolfo in 2006. She is working in the infrastructure and security group at Google on the West Coast.

Recent & Upcoming **PhD Defenses**

Debra Cook

Advisor: Angelos Keromytis

Elastic Block Ciphers

Abstract: Standard block ciphers are designed around one or 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-padding 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 theoretical counterparts 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 if 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 b to 2b bits from round functions that are independently chosen pseudorandom permutations on b bits.

German Creamer

Advisors:
Sal Stolfo and
Yoav Freund

Using Boosting for Automated Trading and Planning

Abstract: The problem: Much of finance theory is based on the efficient market hypothesis. According to this hypothesis, the prices of financial assets, such as stocks, incorporate all information that may affect their future performance. However, the translation of publicly available information into predictions of future performance is far from trivial. Making such predictions is the livelihood of stock traders, market analysts, and the like. Clearly, the efficient market hypothesis is only an approximation which ignores the cost of producing accurate predictions.

Markets are becoming more efficient and more accessible because of the use of ever faster methods for communicating and analyzing financial data. Algorithms developed in machine learning can be used to automate parts of this translation process. In other words, we can now use machine learning algorithms to analyze vast amounts of information and compile them to predict the performance of companies, stocks, or even market analysts. In financial terms, we would say that such algorithms discover inefficiencies in the current market. These discoveries can be used to make a profit and, in turn, reduce the market inefficiencies or support strategic planning processes.

Relevance: Currently, the major stock exchanges such as NYSE and NASDAQ are transforming their markets into electronic financial markets. Players in these markets must process large amounts of information and make instantaneous investment decisions. Machine learning techniques help investors and corporations recognize new business opportunities or potential corporate problems in these markets. With time, these techniques help the financial market become better regulated and more stable. Also, corporations could save significant amount of resources if they can automate certain corporate finance functions such as planning and trading.

Results: This dissertation offers a novel approach to using boosting as a predictive and interpretative tool for problems in finance. Even more, we demonstrate how boosting can support the automation of strategic planning and trading functions.

Many of the recent bankruptcy scandals in publicly held US companies such as Enron and WorldCom are inextricably linked to the conflict of interest between shareholders (principals) and managers (agents). We evaluate this conflict in the case of Latin American and US companies. In the first part of this dissertation, we use Adaboost to analyze the impact of corporate governance variables on performance. In this respect, we present an algorithm that calculates alternating decision trees (ADTs), ranks variables according to their level of importance, and generates representative ADTs. We develop a board Balanced Scorecard (BSC) based on these representative ADTs which is part of the process to automate the planning functions.

In the second part of this dissertation we present three main algorithms to improve forecasting and automated trading. First, we introduce a link mining algorithm using a mixture of economic and social network indicators to forecast earnings surprises, and cumulative abnormal return. Second, we propose a trading algorithm for short-term technical trading. The algorithm was tested in the context of the Penn-Lehman Automated Trading Project (PLAT) competition using the Microsoft stock. The algorithm was profitable during the competition.

Third, we present a multi-stock automated trading system that includes a machine learning algorithm that makes the prediction, a weighting algorithm that combines the experts, and a risk management layer that selects only the strongest prediction and avoids trading when there is a history of negative performance. This algorithm was tested with 100 randomly selected S&P 500 stocks. We find that even an efficient learning algorithm, such as boosting, still requires powerful control mechanisms in order to reduce unnecessary and unprofitable trades that increase transaction costs.

Rui Kuang

Advisor: Christina Leslie

Inferring Protein Structure with Discriminative Learning and Network Diffusion

Abstract: As the complete genomes of more and more species become available, the large scale study of protein structure 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 the development of new machine learning approaches for tackling the challenging problem of protein structure prediction. In this thesis, we focus on applying several advanced learning techniques, including kernel methods and network diffusion algorithms, to address four fundamental learning problems in protein structure inference:

• Classification of a new protein into its structural class. To detect subtle sequence similarities between remote homologies, we propose inexact matching string kernels and profile-based string kernels for use with the support vector machine (SVM). We also use the kernel-based SVM classifier to extract discriminative sequence motifs, which can correspond to meaningful structural features in the protein data.

• Protein database search for homology detection. We propose a new network diffusion algorithm, called MotifProp, based on a protein-motif network to detect more subtle sequence similarities than a pairwise comparison method such as PSI-BLAST. Furthermore, top-ranked motifs and motif-rich regions induced by the propagation are also helpful for discovering conserved structural components in remote homologies.

• Prediction of dihedral torsion angles of protein backbone. We apply kernel methods to accurately predict protein backbone torsion angles, which helps to substantially improve the modeling of local structures of protein sequence segments, especially the loop conformations, which do not form regular structures.

• Segmentation of multi-domain protein. We predict protein domain labels and boundaries simultaneously by solving an 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 effective and efficient solutions to these four fundamental problems in protein structure inference, validated with large scale experiments. The proposed new techniques, achieving high prediction performance and capable of handling huge datasets, will become increasingly important in this post-genome era.

Smaranda Muresan

Advisors: Owen Rambow and Judith Klavans

Learning Constraint-based Grammars from Representative Examples

Abstract: Computationally efficient models for natural language understanding can have a wide variety of applications starting from text mining and question answering, to natural language interfaces to databases. Constraint-based grammar formalisms have been widely used for deep language understanding. Yet, one serious obstacle for their use in real world applications is that these formalisms have overlooked an important requirement: learnability. Currently, there is 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 learnable. These grammars model both syntax and semantics and have constraints at the rule level for semantic composition and semantic interpretation. The interpretation constraints allow access to meaning during language processing. They establish links between linguistic expressions and the entities they refer to in the real world. We use an ontology-based interpretation, proposing 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 inferences.

In this thesis, we propose a new relational learning model for LWFG induction. The learner is presented with a small set of positive representative examples, which consist of utterances paired with their semantic representations. We have proved that the search space for grammar induction is a complete grammar lattice, which allows the construction and generalization of the hypotheses and guarantees the uniqueness of the solution, regardless of the order of learning. We have proved a learnability theorem and have provided polynomial algorithms for LWFG induction, proving their soundness. The learnability theorem extends significantly the class of problems learnable by Inductive Logic Programming methods.

In this dissertation, we have implemented a system that represents an experimental platform for all the theoretical algorithms. The system has the practical advantage of implementing sound grammar revision and grammar merging, which allow an incremental coverage of natural language fragments. We have provided qualitative evaluations that cover the following issues: coverage of diverse and complex linguistic phenomena; terminological knowledge acquisition from natural language definitions; and question answering, where the questions can be vague or precise, while the answer is always precise at the concept level.

Kundan Singh

Advisor: Henning Schulzrinne

Reliable, Scalable and Interoperable Internet Telephony

Abstract: The public switched telephone network (PSTN) provides ubiquitous availability and very high scalability of more than a million busy hour call attempts per switch. If large carriers are to adopt Internet telephony, then Internet telephony servers should offer at least similar quantifiable guarantees for scalability and reliability using metrics such as call setup latency, server call handling capacity, busy hour call arrivals, mean-time between failures and mean-time to recover. This thesis presents a reliable, scalable and interoperable Internet telephony architecture for user registration, call routing, conferencing and unified messaging using commodity hardware. The results extend beyond Internet telephony to encompass multimedia communication in general.

The architecture presented in this thesis deals with two aspects: at least PSTN-grade reliability and scalability of the Internet telephony servers, and interoperable Internet telephony services such as conferencing and voice mail using existing protocols. We describe the architecture and implementation of our Session Initiation Protocol (SIP)-based enterprise Internet telephony architecture known as Columbia InterNet Extensible Multimedia Architecture (CINEMA). It consists of a SIP registration and proxy server, a multi-party conferencing server, a gateway for interworking SIP with ITU's H.323, an interactive voice response system and a multimedia mail server. CINEMA provides a distributed interoperable architecture for collaboration using synchronous communications like multimedia conferencing, instant messaging, shared web-browsing, and asynchronous communications like discussion forum, shared files, voice and video mails. It allows seamless integration with various communication means like telephone, IP phone, web and electronic mail.

We present two techniques for providing scalability and reliability in SIP: server redundancy and a novel peer-to-peer architecture. For the former, we use DNS-based load sharing among multiple distributed servers that use backend SQL databases to maintain user records. Our two-stage architecture scales linearly with the number of servers. For the latter, we propose a peer-to-peer Internet telephony architecture that supports basic user registration and call setup as well as advanced services such as offline message delivery, voice mail and multi-party conferencing using SIP. It interworks with server-based SIP infrastructures.

Alejandro Troccoli

Advisor: Peter Allen

New methods and tools for 3D-modeling of large scale outdoor scenes using range and color images

Abstract: Systems for the creation of photorealistic models using range scans and digital photographs are becoming increasingly popular in a wide range of fields, from reverse engineering to cultural heritage preservation. These systems employ a range finder to acquire the geometry information and a digital camera to measure color detail. But bringing together a set of range scans and color images to produce an accurate and usable model is still an area of research with many unsolved problems.

In this dissertation we present new tools and methods 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 cues to find the correct camera pose, which we have successfully applied in the creation of a digital model of an archaeological excavation in Monte Polizzo, Sicily. We also address the problem of how to build seamless integrated texture maps from images that were taken under different illumination conditions. To achieve this we present two different solutions. The first one is to align all the images to the same illumination. For this, we have developed a technique that computes a relighting operator over the area of overlap of a pair of images, which we then use to relight the entire image. Our proposed method can handle images with shadows and can effectively remove the shadows from the image, if required. The second technique uses the ratio of two images to factor out the diffuse reflectance of an image from its illumination. We achieve this without any light measuring device. By computing the actual reflectance we remove from the images any effects of the illumination, which then allows us to create new renderings under novel illumination conditions.

Ke Wang

Advisor: Salvatore Stolfo

Network Payload-based Anomaly Detection and Content-based Alert Correlation

Abstract: Every computer on the Internet nowadays is a potential target for a new attack at any 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 cause denial-of-service, system crashes, or information theft resulting in the loss of critical information. In this thesis, 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, like packet headers and derived statistics describing connections and sessions (packet rates, bytes transferred, etc.) to detect unusual events that indicate a system is likely under attack. These approaches, however, are blind to the content of the packet stream, and in particular, the packet content delivered to a service that contains the data and code that exploits the vulnerable application software. We conjecture that fast and efficient detectors that focus on network packet content anomaly detection will improve defenses and identify zero-day attacks far more accurately than approaches that consider only header information.

We therefore present two payload-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 (anomalous) content indicative of a new exploit. These payload-based anomaly sensors can augment other sensors and enrich the view of network traffic to detect malicious events. Both PAYL and Anagram create models of site-specific normal network application payload as n-grams in a fully automatic, unsupervised and very efficient fashion. PAYL computes, during a training phase, a profile of byte (1-gram) frequency distribution and their standard deviation of the application payload flowing to a single host and port. PAYL produces a very fine-grained model that is conditioned on payload length. Anagram models high-order n-grams (n>1) which capture the sequential information between bytes. We experimentally demonstrate that both of these sensors are capable of detecting new attacks with high accuracy and low false positive rates. Furthermore, in order to detect the very early onset of a worm attack, we designed an ingress/egress correlation function that is built in the sensors to quickly identify the worms' initial propagation. The sensors are also designed to generate robust signatures of validated malicious packet content. The technique does not depend upon the detection of probing or scanning behavior or the prevalence of common probe payload, so it is especially useful for the detection of slow and stealthy worms.

An often-cited weakness of anomaly detection systems is that they suffer from "mimicry attack": clever adversaries may craft attacks that appear normal to an anomaly detector and hence will go unnoticed as a false negative. A mimicry attack against a site defended by a content-based anomaly detector can be executed by an attacker by sniffing the target site's traffic flow, modeling the byte distributions of that flow, and blending their exploit with "normal" appearing byte padding. To defend against such attacks, we further propose the techniques of randomized modeling and randomized testing. Under randomized modeling/testing, each sensor will randomly partition the payload into several subsequences, each of whom are modeled/tested separately, thus building a "model/test diversity"

on each host that is unknown to the mimicry attacker. This raises the bar for the attackers as they have no means to know how and where to pad the exploit code to appear normal within each randomly computed partition, even if they have the global knowledge of the target site's content flow.

Finally, PAYL/Anagram's speed and high detection rate makes it valuable not only as a stand-alone network-based sensor, but also as a host-based data-flow classifier in an instrumented, fault-tolerant host-based environment; this enables significant cost amortization and the possibility of a "symbiotic" feedback loop that can improve accuracy and reduce false positive rates over time.

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 the local sensor 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 PAYL 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, reducing incorrect decisions and producing accurate mitigation against zero-day attacks.

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

Weibin Zhao

Advisor: Henning Schulzrinne

Towards Autonomic Computing: Service Discovery and Web Hotspot Rescue

Abstract: Autonomic computing is a vision that addresses the growing complexity of computing systems by enabling them to manage themselves without direct human intervention. This thesis studies two related problems, service discovery and web hotspot rescue, which can serve as a building block and a prototype for autonomic networking and distributed systems, respectively.

Service discovery allows end systems to discover desired services on networks automatically, eliminating administrative configuration. We made four enhancements to the Service Location Protocol (SLP): mesh enhancement, remote service discovery, preference filters, and global attributes. These enhancements improve SLP efficiency and scalability, and enable SLP to better support new and advanced discovery scenarios. The SLP mesh enhancement (mSLP), remote service discovery, and preference filters are now experimental RFCs (Request for Comments). We expect that similar techniques can be applied to other service discovery systems.

During the development of mSLP, we designed selective anti-entropy, a generic mechanism for high availability partial replication. Traditional anti-entropy only supports full replication. We enhanced it to support partial replication by allowing two replicas to selectively reconcile inconsistent data in a session.

Web hotspots are short-term dramatic load spikes. We developed DotSlash, a self-configuring and scalable rescue system for handling web hotspots effectively. DotSlash works autonomously. It uses service discovery to allocate resources dynamically from a server pool distributed globally, and uses adaptive overload control to automate the whole rescue process. As a comprehensive solution, DotSlash enables a web site to build an adaptive distributed web server system on the fly, replicate application programs dynamically, and set up distributed query result caching on demand. DotSlash relieves a spectrum of bottlenecks ranging from access network bandwidth to web servers, application servers, and database servers.

As part of DotSlash, we developed a prediction algorithm for estimating the upper bound of future web traffic volume, which is simple and effective for short-term bursty web traffic. This algorithm provides insight into characterizing traffic of web hotspots, and is useful for web server overload prevention.

2006 Technical Report Series

The following technical reports were published through November 2006.

All 2006 reports are available at *http://www.cs.columbia.edu/research/publications*

CUCS-001-06
Converting from Spherical to Parabolic Coordinates
Aner Ben-Artzi

CUCS-002-06
Binary-level Function Profiling for Intrusion Detection and Smart Error Virtualization
Michael Locasto, Angelos Keromytis

CUCS-003-06
W3Bcrypt: Encryption as a Stylesheet
Angelos Stavrou, Michael Locasto, Angelos D. Keromytis

CUCS-004-06
Grouped Distributed Queues: Distributed Queue, Proportional Share Multiprocessor Scheduling
Bogdan Caprita, Jason Nieh, Clifford Stein

CUCS-005-06
Theoretical Bounds on Control-Plane Self-Monitoring in Routing Protocols
Raj Kumar Rajendran, Vishal Misra, Dan Rubenstein

CUCS-007-06
Using an External DHT as a SIP Location Service
Kundan Singh, Henning Schulzrinne

CUCS-008-06
Rigid Formations with Leader-Follower Architecture
Tolga Eren, Walter Whiteley, Peter N. Belhumeur

CUCS-009-06
A Runtime Adaptation Framework for Native C and Bytecode Applications
Rean Griffith, Gail Kaiser

CUCS-010-06
Evaluating an Evaluation Method: The Pyramid Method Applied to 2003 Document Understanding Conference (DUC) Data
Rebecca Passonneau

CUCS-011-06
Passive Duplicate Address Detection for Dynamic Host Configuration Protocol (DHCP)
Andrea G. Forte, Sangho Shin, Henning Schulzrinne

CUCS-012-06
A Theory of Spherical Harmonic Identities for BRDF/Lighting Transfer and Image Consistency
Dhruv Mahajan, Ravi Ramamoorthi, Brian Curless

CUCS-013-06
Using Angle of Arrival (Bearing) Information in Network Localization
Tolga Eren, Walter Whiteley, Peter N. Belhumeur

CUCS-014-06
PalProtect: A Collaborative Security Approach to Comment Spam
Benny Wong, Michael Locasto, Angelos D. Keromytis

CUCS-015-06
Streak Seeding Automation Using Silicon Tools
Atanas Georgiev, Sergey Vorobiev, William Edstrom, Ting Song, Andrew Laine, John Hunt, Peter Allen

CUCS-016-06
Bloodhound: Searching Out Malicious Input in Network Flows for Automatic Repair Validation
Michael Locasto, Matthew Burnside, Angelos D. Keromytis

CUCS-017-06
Quantifying Application Behavior Space for Detection and Self-Healing
Michael Locasto, Angelos Stavrou, Gabriela G. Cretu, Angelos D. Keromytis, Salvatore J. Stolfo

CUCS-018-06
Seamless Layer-2 Handoff using Two Radios in IEEE 802.11 Wireless Networks
Sangho Shin, Andrea G. Forte, Henning Schulzrinne

CUCS-019-06
A Survey of Security Issues and Solutions in Presence
Vishal Kumar Singh, Henning Schulzrinne

CUCS-020-06
Anagram: A Content Anomaly Detector Resistant to Mimicry Attack
Ke Wang, Janak Parekh, Salvatore Stolfo

CUCS-021-06
A First Order Analysis of Lighting, Shading, and Shadows
Ravi Ramamoorthi, Dhruv Mahajan, Peter Belhumeur

CUCS-022-06
PBS: A Unified Priority-Based CPU Scheduler
Hanhua Feng, Vishal Misra, Dan Rubenstein

CUCS-023-06
SIMPLEstone—Benchmarking Presence Server Performance
Vishal Kumar Singh, Henning Schulzrinne

CUCS-024-06
Speculative Execution as an Operating System Service
Michael Locasto, Angelos Keromytis

CUCS-025-06
Exploiting Temporal Coherence for Pre-computation Based Rendering
Ryan Overbeck

CUCS-026-06
Privacy-Preserving Payload-Based Correlation for Accurate Malicious Traffic Detection
Janak Parekh, Ke Wang, Salvatore Stolfo

CUCS-027-06
Feasibility of Voice over IP on the Internet
Alex Sherman, Jason Nieh, Yoav Freund

CUCS-028-06
Practical Preference Relations for Large Data Sets
Kenneth Ross, Peter Stuckey, Amelie Marian

CUCS-029-06
Measurement and Evaluation of ENUM Server Performance
Charles Shen, Henning Schulzrinne

CUCS-030-06
Projection Volumetric Display using Passive Optical Scatterers
Shree K. Nayar, Vijay N. Anand

CUCS-031-06
Complexity and tractability of the heat equation
Arthur G. Werschulz

CUCS-032-06
Linear Approximation of Optimal Attempt Rate in Random Access Networks
Hoon Chang, Vishal Misra, Dan Rubenstein

CUCS-033-06
Throughput and Fairness in Random Access Networks
Hoon Chang, Vishal Misra, Dan Rubenstein

CUCS-034-06
A Framework for Quality Assurance of Machine Learning Applications
Christian Murphy, Gail Kaiser, Marta Arias

CUCS-035-06
Debugging Woven Code
Marc Eaddy, Alfred Aho, Weiping Hu, Paddy McDonald, Julian Burger

CUCS-037-06
Specifying Confluent Processes
Olivier Tardieu, Stephen A. Edwards

CUCS-038-06
MacShim: Compiling MATLAB to a Scheduling-Independent Concurrent Language
Neesha Subramaniam, Ohan Oda, Stephen A. Edwards

CUCS-039-06
A VoIP Privacy Mechanism and its Application in VoIP Peering for Voice Service Provider Topology and Identity Hiding
Charles Shen, Henning Schulzrinne

CUCS **Resources**

You can subscribe to the CUCS news mailing list at
http://lists.cs.columbia.edu/mailman/listinfo/cucs-news/

CUCS colloquium information is available at
http://www.cs.columbia.edu/lectures

Visit the CUCS alumni portal at **http://alum.cs.columbia.edu**

Columbia University in the City of New York
Department of Computer Science
1214 Amsterdam Avenue
Mailcode: 0401
New York, NY 10027-7003

ADDRESS SERVICE REQUESTED

NON-PROFIT ORG.
U.S. POSTAGE
PAID
NEW YORK, NY
PERMIT 4332