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

CUCS **Resources**

Stay in Touch!

Visit the CUCS Alumni Portal at **https://mice.cs.columbia.edu/alum** to:

- Update your contact information
- Look at recent job postings
- Get departmental news

If you know another alumni who may not be receiving the newsletter, please forward the Alumni Portal link to them.

In the spirit of environmental responsibility, we are moving towards more electronic (and less hard-copy) distribution of the newsletter.

If you'd like to continue receiving paper copies of the newsletter, please visit the Alumni Portal to sign up.

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

Read the CUCS Newsletter online at
http://www.cs.columbia.edu/resources/newsletters

CS@CU

NEWSLETTER OF THE DEPARTMENT OF COMPUTER SCIENCE AT COLUMBIA UNIVERSITY VOL.6 NO.2 SPRING 2010

Young photographers at work on their Bigshot camera

Professor Shree Nayar's **Little Camera is a Big Idea** for Children Around the World

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

Professor **Shree Nayar** has dedicated much of his computer science career to improving the way cameras take pictures. Four years ago, he decided to move in a new direction: to design a camera that could improve the way children learn about science and one another.

Nayar, the T. C. Chang Professor of Computer Science, came up with a prototype as sleek as an iPod and as tactile as a Lego set: the Bigshot digital camera. It comes as a kit, allowing children as young as eight to assemble a device as sophisticated as the kind grown-ups use—with a flash and standard, 3-D and panoramic lenses—only cooler. Its color palette is inspired by M&Ms, a hand crank provides power even when there are no batteries and a transparent back panel shows the camera's inner workings.

Nayar also worked with a group of engineering students, led by Guru Krishnan, An Tran and Brian Smith, to create a website, bigshotcamera.org, that walks children, teachers and parents through the assembly process. Eventually, it will serve as a kind of Flickr for kids, with young photographers from around the world sharing their pictures. "The idea here was not to create a device that was an inexpensive toy," says Nayar. "The idea was to create something that could be used as a platform for education across many societies."

Nayar, chair of the Columbia University Computer Science Department, worked on Bigshot for two years. The project is an extension of his work as director of the Computer Science Department's Computer Vision Lab, where he has expertise in highly sensitive cameras. Among his inventions is the Omnicam, a video camera that shoots seamless 360-degree images, and a technology—recently

T. C. Chang Professor Shree Nayar
with his Bigshot digital camera

developed in collaboration with Sony—that extends the range of brightness and color that cameras can capture.

But, as the father of two young children, he wanted to have an impact beyond the high-tech sector on a humanitarian level. He was inspired by the 2005 Oscar-winning documentary *Born Into Brothels*, which depicts the lives of children growing up in Calcutta's red-light district. The filmmaker, British photographer Zana Briski, gave 35 mm film cameras to eight children and watched as those cameras transformed their lives.

"The film reaffirmed something I've believed for a long time, which is that the camera, as a piece of technology, has a very special place in society," says Nayar, who grew up in New Delhi. "It allows us to express ourselves and to communicate with each other in a very powerful way."

With the Bigshot, Nayar wants to not only empower children and encourage their creative vision, but also get them excited about science. Each building block of the camera is designed to teach a basic concept of physics: why light bends when it passes through a transparent object, how mechanical energy is converted into electrical energy, how a gear train works.

Nayar would like to roll out the camera, now in prototype form, along the lines of the One Laptop Per Child campaign: For each one sold at the full price of around $100, several would be donated to underprivileged schools in the United States and abroad. He will soon begin looking for a partner—a company or nonprofit—to help put Bigshot into production.

In the meantime, Nayar, Krishnan, Tran and Smith have been field-testing the camera with children around the world. Over the summer, Krishnan and Tran took several Bigshot prototypes to their hometowns: Bangalore, India, and Vung Tau, Vietnam, respectively. Nayar also brought the camera to two New York City Schools, the private School at Columbia and Mott Hall in Harlem.

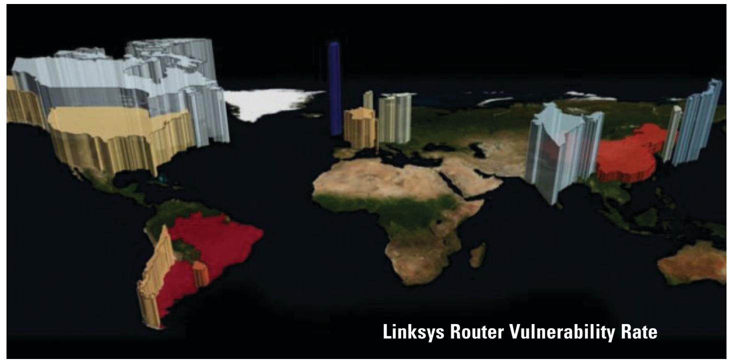
Each spent a morning teaching several small groups of children how to assemble the cameras; after lunch, their charges went out to take pictures. The response from the kids was one of overwhelming enthusiasm. "They were ready to buy the camera then and there," says Krishnan. "One offered me 10,000 rupees ($200)." More importantly, tests that Nayar and his team gave out two days later showed that the students had retained the science concepts that Bigshot was expected to teach.

For Nayar, the best part of this experience has been looking at the pictures. "I am addicted to the pictures; I can't get enough of them," he says. "The fact that some of the kids were using a camera for the first time, and they were able to frame what they thought was important and capture that moment so beautifully, was really remarkable."

It's an experience he hopes to bring to many more children, locally and globally.

*This article by Anna Kuchment appeared in the November 4, 2009 issue of the **Columbia Record** and is reprinted with permission.*

Scan of Internet Uncovers **Thousands of Vulnerable Embedded Devices**



Linksys Router Vulnerability Rate

Researchers scanning the internet for vulnerable embedded devices have found nearly 21,000 routers, webcams and VoIP products open to remote attack.

Professor
Sal Stolfo

Ph.D. student
Ang Cui

Their administrative interfaces are viewable from anywhere on the internet and their owners have failed to change the manufacturer's default password.

Linksys routers had the highest percent of vulnerable devices found in the United States—45 percent of 2,729 routers that were publicly accessible still had a default password in place. Polycom VoIP units came in second, with default passwords lingering on about 29 percent of 585 devices accessible over the internet.

"You can reflash the firmware or install any software you wish on vulnerable devices," said **Salvatore Stolfo**, a Columbia University computer science professor who is overseeing the research project aimed at uncovering vulnerable appliances on the internet. "These devices will be owned and used by bot herders and other miscreants."

Hackers can use vulnerable routers to conduct click fraud or DNS cache poisoning attacks or to launch attacks on other systems. (See Wired's recent Threat Level story about vulnerable routers used by Time Warner customers.) Someone with remote access to the administrative interface of a VoIP system would also be able to install firmware to record conversations.

The research project, devised by Columbia University grad student **Ang Cui** at the university's Intrusion Detection Systems Laboratory, involves scanning networks belonging to the largest internet service providers in North America, Europe and Asia. The lab is sponsored by the Defense Advance Research Projects Agency (Darpa), the Department of Homeland Security and other federal agencies.

"Vulnerable devices can be found in significant numbers in all parts of the world covered by our scan," the researchers wrote in a summary of their initial findings presented at a symposium in June. "The double digit vulnerability rates suggest that a large botnet can be created by constituting only embedded network devices."

Since initiating the project last December, the Intrusion Detection researchers have scanned 130 million IP addresses and found nearly 300,000 devices whose administrative interfaces were remotely accessible from anywhere on the internet. The 21,000 devices with default passwords are the most vulnerable, but the rest are theoretically vulnerable to brute-force password-cracking attacks, Stolfo said. Extrapolating from the numbers they've gathered, the researchers estimate that 6 million vulnerable devices are likely connected to the internet.

The group has so far focused on residential routers and devices but is now looking at scanning more sensitive networks to search for vulnerable devices inside large corporations and government networks.

"People tend to buy stuff and bring them to work and just plug them in," Stolfo said. "So we think we'll be able to find vulnerable devices in highly sensitive places."

The researchers didn't attempt to explore the administrative interfaces or tamper with the devices they found, so they believe their work isn't illegal.

"The scan script sends the public password for the product, and if the device responds with the 'command prompt' for that product interface, then the machine is obviously open," Stolfo said. "We do not access the machine. We break the connection at that point and move on."

ISPs can easily detect the scanning, and the researchers embedded a URL in their probes for a webpage explaining the project that gives network providers a chance to opt out. Stolfo says a couple of universities, a security company and government agency have so far asked to be exempt from the scan.

The researchers have provided ISPs with their findings in the hope that they will do something to protect vulnerable customers.

"It's not clear how an ISP is going to do a general announcement, but we hope there will be some way to communicate to the home user in particular about what they have to do to reconfigure their device," Stolfo said.

But Stolfo says product makers are the real culprits and need to hide their administrative interfaces by default and provide clear instructions for users who want to alter that configuration. Vendors should also be more forceful in communicating to users that default passwords need to be changed to robust alphanumeric passwords that include special characters to thwart brute force attacks.

"This is not a password you're going to need every day, so setting a very hard password and recording it at home on a piece of paper is probably a safe thing to do," Stolfo says.

The group plans to run the scan for a few more months, then wait before re-running it to see if the number of vulnerable devices has fallen after they've notified ISPs about the vulnerabilities.

*Originally published in **Wired.com**.
Reprinted by permission.*

Addressing the **Insider Threat**

by **Shari Lawrence Pfleeger** *(RAND Corporation)*
and **Salvatore J. Stolfo** *(Columbia University)*

Professor
Sal Stolfo

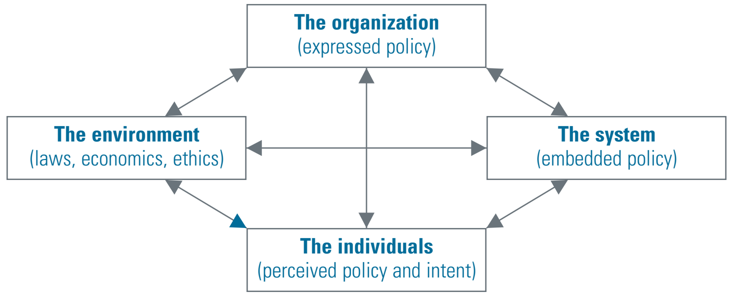


Figure 1. Framework for taxonomy of insiders and their actions. This taxonomy provides a consistent vocabulary for describing which aspects of the insider threat are being addressed, and takes into account the roles of organizations, individuals, IT systems, and the environment in enabling insider threat behavior.[9]

	DETECTION	PREVENTION	MITIGATION	PUNISHMENT	REMEDIATION
ORGANIZATION: NO EXPRESSED POLICY		Create organizational policy	Update related policies		Update related policies
SYSTEM: NO EMBEDDED POLICY	Embedded decoys; watchful monitoring	Embed organizational policy			
INDIVIDUAL: NO MALICIOUS INTENT		User training, incentives, reminders, access control			
ENVIRONMENT: LAW, ETHICS APPLY		Remind users of legal implications of their actions and of costs to organization		Apply legal punishments	

Table 1. Applying a framework for responding to insider threats.

As users, managers, researchers, or administrators, we often worry about outsiders attacking our systems and networks, breaking through the perimeter defenses we have established to keep out bad actors. But we must also worry about the *insider threat*—

people with legitimate access who behave in ways that put our data, our systems, our organizations, and even our businesses' viability at risk. Such behavior might not be malicious; it might be well-intended but still have unwelcome consequences.

Considerable research has been done to examine the nature of inappropriate insider activity, with the goal that eventually organizations can reduce the threat. Beginning in 1999, RAND conducted a series of workshops to generate a research agenda for addressing this problem.[1-3] In parallel, the US Department of Defense (DoD) outlined a set of policy changes and research directions for reducing the insider threat.[4] And the Software Engineering Institute's Computer Emergency Response Team (CERT) has been working with the US Secret Service to understand convicted insiders' motivations.[5] From these and other efforts, a rich literature illuminating various aspects of the insider threat problem is emerging.

The Scope of the Insider Threat

But how real is the insider threat? Many recent anecdotes about cybercrime suggest that often the threat to an organization's computer-based assets is greater from those within the organization than from without. In a 2007 Computer Security Institute survey about computer crime and security[6], 59 percent of respondents thought they had experienced insider abuse of network resources. About one in four respondents said that more than 40 percent of their total financial losses from cyber attack were due to insider activities. However, the 2008 survey had significantly different results:

> As noted in last year's report, a great deal is made of the insider threat, particularly by vendors selling solutions to stop insider security infractions. It's certainly true that some insiders are particularly well-placed to do enormous damage to an organization, but this survey's respondents seem to indicate that talk of the prevalence of insider criminals may be overblown. On the other hand, we're speaking here of financial losses to the organization, and in many cases significant insider crimes, such as leaking customer data, may not be detected by the victimized organization and no direct costs may be associated with the theft.[7]

Credible data describing the scope and impact of unwelcome insider actions are hard to come by, for two reasons. First, many organizations are loathe to reveal the nature and magnitude of the cyber incidents they've experienced for fear of reputational harm. Second, most cyber surveys are convenience surveys; it's impossible to know what population the results represent. This paucity of data is challenging for insider threat researchers, who need good data with which to build models, make predictions, and support good decision-making. The large-scale, carefully sampled National Computer Security Survey[8] suggests that the threat is real and the consequences significant:

- Forty percent of all incidents reported by the 7,818 respondents (representing 36,000 US businesses) were attributed to insiders.
- Seventy-four percent of all cyber theft was attributed to insiders, including 93 percent of embezzlement incidents and 84 percent of intellectual property thefts.

For the past few years, the Institute for Information Infrastructure Protection (I3P) at Dartmouth College (with funds from the US Department of Homeland Security) has supported a project exploring ways to understand and address the insider threat. With researchers at Columbia University, Cornell University, Dartmouth College, MITRE, Indiana University, Purdue University, and RAND, we've examined not only how technology can reveal the threat's nature and magnitude, but also how it's influenced by the environment in which the insiders operate. The overall goal is to suggest appropriate responses to the insider threat; after all, the response to an insider accidentally selecting the wrong menu entry should be different from the response to an ex-employee trying to exact revenge.

In our project's early days, it became clear that many ideas exist about what "insider" means and what unwelcome behavior constitutes an "insider threat." For example, insiders are more than just employees or ex-employees—they can be business partners, auditors, consultants, or other people and systems who receive short- or long-term access to an organization's systems. Without a unifying framework, we have difficulty recognizing emerging insider problems, comparing incidents, or dealing with them appropriately. For this reason, Joel Predd, Shari Lawrence Pfleeger, Jeffrey Hunker and Carla Bulford[9] developed a taxonomy of insiders and their actions, which Figure 1 shows. The taxonomy doesn't prescribe a uniform definition; rather, it provides a consistent vocabulary for describing clearly which aspects of the insider threat problem the research and practice address. It also provides the basis for discussing, comparing, and contrasting the various possible responses to different kinds of insider behaviors. The framework takes into account the roles of the organization, the individual, the IT system, and the environment in enabling unwelcome behavior.

Because not all insiders are alike, we must distinguish among the different types of insider threat, differentiate problems we can address from those we can't, and determine the roles technology and policy play in crafting responses. Policy is a particularly thorny issue because the stated policy (called the *de jure* policy in the taxonomy) isn't always the same as its interpretation and enforcement (the *de facto* policy). For example, most organizations forbid the use of their computer systems for personal use: the *de jure* policy. But in actuality, most organizations look the other way for some personal uses: the *de facto* policy. Sometimes, the *de facto* policy is stronger than the *de jure*, as when a security guard challenges a worker in the office at 2 a.m., even though the worker is wearing a proper badge.

The goal of much insider threat research is to make more effective the prevention, detection, mitigation, remediation, and punishment of unwelcome action by the people and systems that have legitimate access to our networks. It's not enough to expect technology to prevent insider misdeeds. Instead, we need a multifaceted set of strategies that address all elements of the taxonomy: the organization (including its culture and goals), the system (including the completeness and correctness of its implementation of *de jure* policy and its ability to learn *de facto* behavior), the environment (including legal restrictions on monitoring and analysis), and the individual (including motivation and intent). Table 1 illustrates how the taxonomy, coupled with goals of prevention, detection, mitigation, remediation, and punishment, can suggest sensible and effective response options.

As technologists, we often hope to use our skills to monitor behavior and predict the new threats our systems will face. But the dynamic threat environment, coupled with continuing technological advancement, make it impossible to predict with certainty what our systems will look like and what features and functions they'll provide. That same uncertainty makes it difficult to predict what insiders will do and when and how they'll do it. However, the substantial literature on "workplace deviance"[10] tells us with certainty that insiders will continue to behave badly, using our computer systems as a means or as a target. Thus, insider threat detection and mitigation will continue to be a vexing and persistent security—and very human—problem.

Acknowledgments

This material is based on work supported by the US Department of Homeland Security under grant award numbers 2006-CS-001-00000 and 2006-CS-001-000001-02 and the Army Research Office under grant ARO DA W911NF-06-10151. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US Department of Homeland Security or the Army Research Office.

References

1. R.H. Anderson, *Research and Development Initiatives Focused on Preventing, Detecting, and Responding to Insider Misuse of Critical Defense Information Systems: Results of a Three-Day Workshop*, research report RAND CF-151-OSD, RAND Corp., 1999.
2. R.H. Anderson et al., *Research on Mitigating the Insider Threat to Information Systems #2, Proc. Workshop Held August 2000*, research report RAND CF-163-DARPA, RAND Corp., 2000.
3. R.C Brackney and R.H. Anderson, *Understanding the Insider Threat: Proceedings of a March 2004 Workshop*, research report RAND CF-196-ARDA, RAND Corp., 2004.
4. *Final Report of the Insider Threat Integrated Process Team*, Office of the Assistant Secretary of Defense (Command, Control, Communications, and Intelligence), US Dept. Defense, 24 Apr. 2000.
5. M. Keeney et al., "Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors," US Secret Service and CERT Coordination Center/SEI, May 2005.
6. R. Richardson, "2007 Computer Crime and Security Survey," Computer Security Inst., 2007, pp. 12-13, 15; http://i.cmpnet.com/v2.gocsi.com/pdf/CSISurvey2007.pdf.
7. R. Richardson, "2008 CSI Computer Crime and Security Survey," Computer Security Inst., 2008; www.gocsi.com/forms/csi_survey.jhtml.
8. R. Rantala, *Cybercrime against Businesses*, 2005, special report NCJ221943, US Bureau of Justice Statistics, Sept. 2008; www.ojp.usdoj.gov/bjs/pub/pdf/cb05.pdf.
9. J. Predd et al., "Insiders Behaving Badly," *IEEE Security and Privacy*, vol. 6, no. 4, 2008, pp. 66-70.
10. S.L. Robinson and J. Greenberg, "Employees Behaving Badly: Dimensions, Determinants, and Dilemmas in the Study of Workplace Deviance," *Trends in Organizational Behavior*, C.L. Cooper, and D.M. Rousseau eds., vol. 5, Wiley, 1998, pp. 1-30.

This article is excerpted from the Guest Editor's Introduction of the November/December 2009 ***IEEE Security and Privacy*** *Magazine's Special Issue on Insider Threats, and is reprinted with permission of the magazine. For the full article see http://www.computer.org/portal/web/security/home*

2009-10 **Distinguished** Lecture Series

The Computer Science Department enjoyed another successful Distinguished Lecture Series during the 2009-10 academic year. This yearly series brings world-renowned scientists and pioneers in academia and industry from across the country to Columbia's campus.

These distinguished speakers give talks about their research, field questions from the audience, and meet with faculty and students. The lectures are open to the public, and videos of all talks are available at the Computer Science department website *(http://www.cs.columbia.edu/lectures)*. The Distinguished Lecture Series ensures that members of the department and the community at large see first-hand the latest and greatest computer science research emerging outside the Columbia campus.

This year four lecturers visited and discussed a broad and exciting range of topics. These speakers helped make this year an exciting and inspirational one for computer science at Columbia. Stay tuned next year when more of the world's greatest researchers will visit our department and tell us about their work.

Professor **Andrew Odlyzko** of the University of Minnesota visited on October 5 and spoke about **Network Evolution, Network Economics, and Network Innovation**.

Professor Odlyzko has written over 150 technical papers in computational complexity, cryptography, number theory, combinatorics, coding theory, analysis, probability theory, and related fields. In recent years he has also been working in electronic commerce, economics of data networks, and economic history, especially on diffusion of technological innovation. Professor Odlyzko had a long career in research and research management at Bell Labs and AT&T Labs, and then built an interdisciplinary research center in Minnesota.

Technology is opening up exciting new opportunities in networking, especially through the convergence of wireline and wireless communications. But technology is just one element, and, as in the past, and perhaps more than in the past, economics and regulation will have major influences on what is deployed, and how it is used. In his talk Professor Odlyzko surveyed the past and present, and offered some speculations about the future. He presented some of the key constraints on technological dreams, including the many false dogmas that are hobbling progress.

Professor **Yoky Matsuoka** of the University of Washington visited on October 26 and spoke about **Understanding Humans with Neurobotics**.

Professor Matsuoka is the Torode Family Endowed Career Development Professor in Computer Science and Engineering at the University of Washington. She received her Ph.D. at MIT in Electrical Engineering and Computer Science in the fields of Artificial Intelligence and Computational Neuroscience. Her work has been recognized with a MacArthur Fellowship, and she has been acclaimed as one of "The Brilliant Ten" in Popular Science Magazine and one of the "Power 25" in Seattle Magazine. She has received a Presidential Early Career Award for Scientists and Engineers (PECASE), an Anna Loomis McCandless Chair from Carnegie Mellon University, and the IEEE Robotics and Automation Society Early Academic Career Award.

Neurobotics is a field that lies at the intersection of Robotics and Neuroscience. In Neurobotics, robotic models and environments are used to understand the neuromuscular control and biomechanics of human limbs. In parallel, robotic systems are developed to augment, replace and rehabilitate damaged sensorimotor functions. Professor Matsuoka presented an overview of this work, with a detailed example of how to understand and restore human level dexterity.

Professor **Barbara Liskov** of the Massachusetts Institute of Technology visited on December 7 and spoke about **The Power of Abstraction**.

Professor Liskov is an Institute Professor at MIT and also Associate Provost for Faculty Equity. She is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, and a fellow of the ACM. She received the ACM Turing Award in 2009, the ACM SIGPLAN Programming Language Achievement Award in 2008, the IEEE Von Neumann medal in 2004, a lifetime achievement award from the Society of Women Engineers in 1996, and in 2003 was named one of the 50 most important women in science by Discover Magazine. Her research interests include distributed systems, replication algorithms to provide fault-tolerance, programming methodology, and programming languages. Her current research projects include Byzantine-fault-tolerant storage systems, peer-to-peer computing, and support for automatic deployment of software upgrades in large-scale distributed systems.

Abstraction is at the center of much work in Computer Science. It encompasses finding the right interface for a system as well as finding an effective design for a system implementation. Furthermore, abstraction is the basis for program construction, allowing programs to be built in a modular fashion. In her talk, Professor Liskov discussed how the abstraction mechanisms we use today came to be, how they are supported in programming languages, and some possible areas for future research.

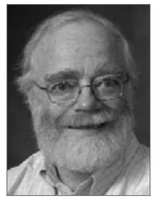
Professor **John Lafferty** of Carnegie Mellon University visited on March 1 and spoke about **Three Rivers in Machine Learning: Data, Computation and Risk**.

Professor Lafferty is a professor in the Computer Science Department at Carnegie Mellon University, with joint appointments in the Machine Learning Department and the Department of Statistics. His research interests include machine learning, statistical learning theory, natural language processing, information theory, and information retrieval. Prof. Lafferty received the Ph.D. in Mathematics from Princeton University, where he was also a member of the Program in Applied and Computational Mathematics. He is an IEEE Fellow, has served as co-director of CMU's new Machine Learning Ph.D. Program, and currently serves as associate editor of the Journal of Machine Learning Research and the Electronic Journal of Statistics.

Machine learning is a confluence of computer science and statistics that is empowering technologies such as search engines, robotics, and personalized medicine. Fundamentally, the goal of machine learning is to develop computer programs that predict well, according to some measure of risk or accuracy. The predictions should get better as more historical data become available. The field is developing interesting and useful frameworks for building such programs, which often demand large computational resources. Theoretical analyses are also being advanced to help understand the tradeoffs between computation, data, and risk that are inherent in statistical learning. Two types of results have been studied: the consistency and scaling behavior of specific convex optimization procedures, which have polynomial computational efficiency, and lower bounds on any statistically efficient procedure, without regard to computational cost. Professor Lafferty's talk gave a survey of some of these developments, with a focus on structured learning problems for graphs and shared learning tasks in high dimensions.

Further details of the lecture series are available online at: http://www.cs.columbia.edu/lectures

New Computer Science Courses



Professor
Steven Bellovin

Professor **Steven Bellovin** taught a new course, COMS 3995, **Computers and Society**, in the Spring 2010 semester.

Computers are heavily entwined in almost everything we do. Their use is not an unmixed blessing. For example, we can get our news from many free, online sources—but their existence is threatening the existence of the newspapers that employ the reporters who gather the news. Social media are a great way to interact—but they can threaten personal privacy. This course explores such issues and more. Specific topics covered include privacy, risks of computer systems, ethical issues for practitioners, employment, freedom of speech, social networks, and intellectual property. The course is structured as a combination of lecture and seminar; discussion and opinions are strongly encouraged.

Adjunct Professor
Sameer Maskey

Adjunct Professor **Sameer Maskey** taught a new course, COMS 6998-007, **Statistical Methods for Natural Language Processing**, in the Spring 2010 semester. Prof. Maskey is a Research Staff Member at the IBM T.J. Watson Research Center.

The course explores topics in Statistical Methods/Machine Learning for real-world Natural Language Processing (NLP) problems. Students study ML topics that are commonly used in NLP such as Maximum Entropy Models, Hidden Markov Models, Clustering techniques, Conditional Random Fields, Expectation-Maximization algorithm, Active Learning and Support Vector Machines. The course explains how these methods are applied to real world NLP problems such as information extraction, stochastic parsing, text segmentation and classification, topic/document clustering and word sense disambiguation. Students also study the details of inference algorithms such as Viterbi, Synchronous Chart Parsing and Beam Search. Students get hands-on experience by implementing some of these ML techniques for classification, clustering and a complex NLP task of machine translation.

Professor
Vishal Misra

Professor **Vishal Misra** taught a new course, COMS 6998-002, **Internet Economics**, in the Spring 2010 semester.

This course introduces modern topics in Internet Economics. The course first addresses the theoretical foundations from a game theoretic perspective, covering topics from cooperative and non-cooperative game theory. Topics include Information, Auctions, Bargaining and Coalitions and Mechanism Design. The course then moves on to application areas such as ISP settlements, Computational Advertising, P2P incentive mechanisms and Spectrum Auctions. Both real life case studies as well as research papers are covered in class.

Adjunct Professor
Erich Nahum

Adjunct Professor **Erich Nahum** taught a new course, COMS 6998-005, **Network Systems Design and Implementation**, in the Spring 2010 semester. Prof. Nahum is a Research Staff Member at the IBM T.J. Watson Research Center.

This class takes a deep look at how network protocols are designed and implemented using the Linux kernel as a case study. The goal is to understand how this important subsystem works in detail in order to conduct experimental research using the Linux kernel. The class includes detailed code walkthroughs of a recent Linux kernel (most likely 2.6.31), a class project, and relevant research paper readings and presentations. Topics include:

- Specific network protocol implementations such as sockets, TCP, IP, Ethernet, ARP, network device drivers, routing and bridging;
- Support mechanisms such as buffer management, packet queuing, timers, hash tables, interrupts, and synchronization;
- Network-related security mechanisms such as packet capture, filtering, firewalling, iptables, and netfilter;
- Other kernel support facilities such as profiling, tracing, and debugging.

Grading is based on class participation, an in-class presentation and a project involving modifying the Linux kernel.

Adjunct Professor
Pablo Rodriguez

Adjunct Professor **Pablo Rodriguez** taught a new course, COMS 6998-003, **Next Generation Network Architectures**, in the Spring 2010 semester. Prof. Rodriguez is the Internet Scientific Director at Telefonica Barcelona, where he leads the systems and networking research team.

Students in this course learn the network architectures behind some of the most successful Internet services (e.g. Facebook, Twitter, Spotify); learn about the latest modern Internet architectures (greener networks, networking technologies for emerging regions, Data Center networking, clean slate designs); become equipped for researching modern networks and distributed systems; and gain exposure to practical, real world network architectures. Each topic provides background on existing mechanisms and includes discussions of current publications and ongoing research. Students are graded based on paper reviews, contributions to in-class discussions, and a final project. The final project is an opportunity for hands-on research. It involves literature survey, programming, running experiments or modeling, analyzing results and writing a report.

Professor
Yechiam Yemini

Professor **Yechiam Yemini** taught a new course, COMS 4995-001, **Principles of Innovation and Entrepreneurship**, in the Spring 2010 semester.

Dell, Yahoo, Google and Facebook were founded by college students. Could you be building the next star technology startup? While the future will tell, this course provides the basic knowledge and skills to help you answer this question. Among other topics, the course studies how changes in technology paradigms give rise to novel opportunities; how to identify, analyze and exploit these opportunities; how to design innovative products and business models to create sustainable competitive edge; how to transform loose hypotheses—on technologies, opportunities, customers, markets, etc—to effective strategic business plans; and how to translate these plans into focused execution while avoiding common 'bugs'. Classes include discussions of the basic principles of innovation; brief workshops in applying them; presentations by founders/CEOs of high-tech startups; and assignments and team projects to create early stage startups. Teams of 3-4 students will pursue an incremental sequence of assignments, emulating early development of a startup. These assignments will culminate in a business plan and proof-of-concept prototypes. The course is intended for graduate and advanced undergraduate CS students, but is suitable for other engineering, science and business school students.

CUCS Faculty Receive **Google Research Awards**

Professor
Luis Gravano

Professor
Tony Jebara

Professor
Tal Malkin

Professor
Vishal Misra

Professors **Luis Gravano**, **Tony Jebara**, **Tal Malkin**, and **Vishal Misra** have recently received Google research awards for projects on web search for event media, on collaborative filtering, on secure routing protocols using oblivious nodes, and on incentivizing managed peer-to-peer systems.

Professor **Luis Gravano** won a Google Research Award for his project titled "Finding and Characterizing the World's Event Media." Social media sites make it easy for users to publish content that is captured or produced in association with real-world events. These events range from widely known ones, such as a presidential inauguration, to smaller, community-specific events, such as an annual convention or a local gathering. Unfortunately, the existing tools to find, organize, and search the social media content associated with events are extremely limited. This project aims to transform the way in which people search and consume social media content from real-world events, a key information-seeking task, by addressing two important problems: identification of events and their associated social media content, and event search.

Professor **Tony Jebara** won a Google Research Award for his project titled "NetTrailMix." The goal of this project is to set up a collaborative filtering problem much like the Netflix challenge where recommendations are provided to users based on large amounts of unsupervised human social activity (as opposed to more standard rating data).

Professor **Tal Malkin** won a Google Research Award for her project titled "Efficient Routing by Oblivious Nodes." The research will enhance routing protocols such that they can compute high-performance routes in a computationally efficient manner without revealing information that might reveal the location of participating nodes. This allows users to send and receive high-bandwidth, low-latency transmissions such as video and audio feeds without revealing their location. Potential applications include celebrity multimedia twitter-like feeds, and network-supported action gaming.

Professor **Vishal Misra** won a Google Research Award for his project titled "Incentivizing Managed Peer to Peer Systems: A Fluid Shapley Value Approach." With the spread of technology today, users everywhere have some resources that can aid providers in increasing revenue or reducing costs. Examples range from users contributing popular viral videos on YouTube to cellphone customers installing femtocells at their residences to reduce network load. This project aims to study the right incentive mechanisms for both providers as well as customers to participate in the system based on some recent analytical techniques developed by Prof. Misra and his collaborators.

Spryridon Antonakopoulos

Buy-at-bulk-related Problems in Network Design

Advisor: Professor Mihalis Yannakakis

Abstract: We study a number of graph-theoretic optimization problems arising in network design. In general, the objective is to determine a network with minimum total cost that can support given traffic demands, possibly subject to additional constraints. Due to economies of scale, the cost of network components is typically sub-additive as a function of the capacity they provide, so these problems have a so-called buy-at-bulk character. Network dimensioning, which represents a high-level phase within the long process of telecommunications network planning, is a primary motivation for this work. We remark that although buy-at-bulk network design has received a lot of attention from researchers in theoretical computer science for more than a decade, these efforts were concentrated on investigating a rather simple version of the problem with minimal constraints, under a broad range of cost models that fall within the buy-at-bulk framework. By contrast, herein we initiate the study of several more involved problem variants, inspired by real-life situations. First, we investigate directed buy-at-bulk network design, which includes cases where network links allow only one-way traffic, or more generally where the cost of installing capacity on a link is asymmetric with respect to the direction of the traffic. Furthermore, in buy-at-bulk network design with protection, we are required to ensure robustness against the failure of any single component, by allocating capacity for each traffic demand along two disjoint routes in the network. Finally, we also consider the traffic grooming problem, which arises when demands requesting small amounts of bandwidth must be accommodated in an optical network offering high-capacity transmission channels. Our main contributions are approximation algorithms with non-trivial worst-case guarantees for each of the above problems; moreover, from the experimental perspective, we present efficient heuristics that perform very well on realistic problem instances.

Matei Ciocarlie

Low-Dimensional Robotic Grasping: Eigengrasp Subspaces and Optimized Underactuation

Advisor: Professor Peter Allen

Abstract: This thesis introduces new methods for enabling the effective use of highly dexterous robotic hands, interfacing with the upcoming generation of neurally controlled hand prostheses, and designing a new class of simple yet effective grasping devices based on underactuation and mechanical adaptation. These methods share a common goal: reducing the complexity that has traditionally been associated, at both computational and mechanical levels, with robotic grasping in unstructured environments.

A key prerequisite for robot operation in human settings is versatility, which, in terms of autonomous grasping, translates into the ability to reliably acquire and interact with a wide range of objects. In an attempt to match the abilities of the most versatile end-effector known, the human hand, many anthropomorphic robotic models have been proposed, with the number of degrees of freedom starting to approach that of their human counterpart. However, these models have proven difficult to use in practice, as the high dimensionality of the posture space means that finding adequate grasps for a target object is often an intractable problem.

In this thesis, we propose using low-dimensional posture subspaces for dexterous or anthropomorphic hands. Human user studies have shown that most of the variance in hand posture for a wide range of grasping tasks is contained in relatively few dimensions. We extend these results to a range of robotic designs, and introduce the concept of eigengrasps as the bases of a low-dimensional, linear hand posture subspace. We then show that a grasp synthesis algorithm that optimizes hand posture in eigengrasp space is both computationally efficient and likely to yield stable grasps.

The emerging field of neurally controlled hand prosthetics faces a similar challenge when using dexterous hand models: bridging the gap between incomplete or noisy neural recordings and the complete set of variables needed to execute a grasping task. We propose using an automated grasp planning component as an interface, accepting real-time operator input and using it to assist in the synthesis of stable grasps. Computational rates needed for direct interaction can be achieved by combining operation in eigengrasp space with online operator input. Furthermore, the eigengrasp planning space can also act as an interaction space, allowing the operator to provide meaningful input for the hand posture using few channels of communication.

Algorithmic approaches to low-dimensional grasping can enable computationally effective algorithms and interaction models. Hardware implementations have the potential to reduce the mechanical complexity and construction costs of a hand design, using concepts such as underactuation and passive mechanical adaptation. Instead of complex run-time algorithms, hand models in this class use design-time analysis to improve performance over a spectrum of tasks. Along these directions, we present a set of analysis and optimization tools for the design of low-dimensional, underactuated hands. We focus on tendon-based mechanisms featuring adaptive joints and compliant fingertips, and show how a number of design parameters, such as tendon routes or joint stiffnesses, can be optimized to enable a wide range of stable grasps.

The ability to effect change on the environment through object acquisition (grasping) and manipulation has the potential to enable many robotic applications with high social impact, including effective neural prostheses, robots for house care or personal assistance, etc. We believe that the methods presented in this thesis represent a number of steps in this direction, advancing towards a proven solution for reliable autonomous grasping in human environments.

Jinwei Gu

Measurement, Modeling, and Synthesis of Time-Varying Appearance of Natural Phenomena

Advisors: Professors Shree Nayar, Peter Belhumeur, Ravi Ramamoorthi

Abstract: Many natural phenomena evolve over time—often coupled with a change in their reflectance and geometry—and give rise to dramatic effects in their visual appearance. In computer graphics, such time-varying appearance phenomena are critical for synthesizing photo-realistic images. In computer vision, understanding the formation of time-varying appearance is important for image enhancement and photometric-based reconstruction. This thesis studies time-varying appearance for a variety of natural phenomena—opaque surfaces, transparent surfaces, and participating media—using captured data.

We have two main goals: (1) to design efficient measurement methods for acquiring time-varying appearance from the real world, and (2) to build compact models for synthesizing or reversing the appearance effects in a controllable way.

We started with time-varying appearance for opaque surfaces. Using a computer-controlled dome equipped with 16 cameras and 160 light sources, we acquired the first database (with 28 samples) of time-and-space-varying reflectance, including a variety of natural processes—burning, drying, decay and corrosion. We also proposed a space time appearance factorization model which disassembles the high-dimensional appearance phenomena into components that can be independently modified and controlled for rendering.

We then focused on time-varying appearance of transparent objects.

Real-world transparent objects are seldom clean—over time their surfaces will gradually be covered by a variety of contaminants, which produce the weathered appearance that is essential for photorealism. We derived a physically-based analytic reflectance model for recreating the weathered appearance in real time, and developed single-image based methods to measure contaminant texture patterns from real samples.

The understanding of the weathered appearance of transparent surfaces was also used for removing image artifacts caused by dirty camera lenses. By incorporating priors on natural images, we developed two fully-automatic methods to remove the attenuation and scattering artifacts caused by dirty camera lenses. These image enhancement methods can be used for post-processing existing photographs and videos as well as for recovering clean images for automatic imaging systems such as outdoor security cameras.

Finally, we studied time-varying appearance of volumetric phenomena, such as smoke and liquid. For generating realistic animations of such phenomena, it is critical to obtain the time-varying volume densities, which requires either intensive modeling or extremely high speed cameras and projectors. By using structured light and exploring the sparsity of such natural phenomena, we developed an acquisition system to recover the time-varying volume densities, which is about 4 to 6 times more efficient than simple scanning. From the perspective of computer vision, our method provides a way to extend the applicable domain of structured light methods from 2D opaque surfaces to 3D volumes.

David Harmon

Robust, Efficient, and Accurate Contact Models

Advisor: Professor Eitan Grinspun

Abstract: Robust, efficient, and accurate collision response remains a difficult and challenging problem in simulation. Many methods exist that partially achieve these properties, but none yet fully attain all three. This thesis investigates existing methodologies with respect to these attributes, and proposes a novel algorithm for the simulation of deformable materials that demonstrate them all. This new method is analyzed and optimized, paving the way for future work in this greatly simplified but powerful manner of simulation.

Christian Murphy

Metamorphic Testing Techniques to Detect Defects in Applications without Test Oracles

Advisor: Professor Gail Kaiser

Abstract: Applications in the fields of scientific computing, simulation, optimization, machine learning, etc. are sometimes said to be "non-testable programs" because there is no reliable test oracle to indicate what the correct output should be for arbitrary input. In some cases, it may be impossible to know the program's correct output a priori; in other cases, the creation of an oracle may simply be too hard. These applications typically fall into a category of software that Weyuker describes as "Programs which were written in order to determine the answer in the first place. There would be no need to write such programs, if the correct answer were known." The absence of a test oracle clearly presents a challenge when it comes to detecting subtle errors, faults, defects or anomalies in software in these domains.

As these types of programs become more and more prevalent in various aspects of everyday life, the dependability of software in these domains takes on increasing importance. Machine learning and scientific computing software may be used for critical tasks such as helping doctors perform a medical diagnosis or enabling weather forecasters to more accurately predict the paths of hurricanes; hospitals may use simulation software to understand the impact of resource allocation on the time patients spend in the emergency room. Clearly, a software defect in any of these domains can cause great inconvenience or even physical harm if not detected in a timely manner.

Without a test oracle, it is impossible to know in general what the expected output should be for a given input, but it may be possible to predict how changes to the input should effect changes in the output, and thus identify expected relations among a set of inputs and among the set of their respective outputs. This approach, introduced by Chen et al., is known as "metamorphic testing". In metamorphic testing, if test case input x produces an output f(x), the function's so-called "metamorphic properties" can then be used to guide the creation of a transformation function t, which can then be applied to the input to produce t(x); this transformation then allows us to predict the expected output f(t(x)), based on the (already known) value of f(x). If the new output is as expected, it is not necessarily right, but any violation of the property indicates a defect. That is, though it may not be possible to know whether an output is correct, we can at least tell whether an output is incorrect.

This thesis investigates three hypotheses. First, I claim that an automated approach to metamorphic testing will advance the state of the art in detecting defects in programs without test oracles, particularly in the domains of machine learning, simulation, and optimization. To demonstrate this, I describe a tool for test automation, and present the results of new empirical studies comparing the effectiveness of metamorphic testing to that of other techniques for testing applications that do not have an oracle. Second, I suggest that conducting function-level metamorphic testing in the context of a running application will reveal defects not found by metamorphic testing using system-level properties alone, and introduce and evaluate a new testing technique called Metamorphic Runtime Checking. Third, I hypothesize that it is feasible to continue this type of testing in the deployment environment (i.e., after the software is released), with minimal impact on the user, and describe a generalized approach called In Vivo Testing.

Additionally, this thesis presents guidelines for identifying metamorphic properties, explains how metamorphic testing fits into the software development process, and discusses suggestions for both practitioners and researchers who need to test software without the help of a test oracle.

Andrew Rosenberg

Automatic Detection and Classification of Prosodic Events

Advisor: Professor Julia Hirschberg

Abstract: Prosody, or intonation, is a critically important component of spoken communication. The automatic extraction of prosodic information is necessary for machines to process speech with human levels of proficiency. In this thesis we describe work on the automatic detection and classification of prosodic events—specifically, pitch accents and prosodic phrase boundaries. We present novel techniques, feature representations and state of the art performance in each of these tasks. We also present three proof-of-concept applications—speech summarization, story segmentation and non-native speech assessment—showing that access to hypothesized prosodic event information can be used to improve the performance of downstream spoken language processing tasks. We believe the contributions of this thesis advance the understanding of prosodic events and the use of prosody in spoken language processing towards the goal of human-like processing of speech by machines.

Alexander Sherman

Guaranteeing Performance through Fairness in Peer-to-Peer File-Sharing and Streaming Systems

Advisor: Professor Jason Nieh

Abstract: Over the past decade Peer-to-Peer (P2P) file-sharing and streaming systems have evolved as a cheap and widely-used technology in distributing content to users. Guaranteeing a level of performance in P2P systems is, therefore, of utmost importance. However, P2P file-sharing and streaming applications suffer from a fundamental problem of unfairness, where many users have a tendency to free-ride by contributing little or no upload bandwidth while consuming much download bandwidth. By taking away an unfair share of resources, free-riders deteriorate the quality of service experience by other users, by causing slower download times and higher packet loss inside the file-sharing and streaming networks respectively. Previous attempts at addressing fair bandwidth allocation in P2P, such as BitTorrent-like systems, suffer from slow peer discovery, inaccurate predictions of neighboring peers' bandwidth allocations, under-utilization of bandwidth, and complex parameter tuning. We present FairTorrent, a new deficit-based distributed algorithm that accurately rewards peers in accordance with their contribution in a file-sharing P2P system. In a nutshell, a FairTorrent peer uploads the next data block to a peer to whom it owes the most data. FairTorrent is resilient to exploitation by free-riders and strategic peers, is simple to implement, requires no bandwidth over-allocation, no prediction of peers' rates, no centralized control, and no parameter tuning. We implemented FairTorrent in a BitTorrent client without modifications to the BitTorrent protocol, and evaluated its performance against other widely-used BitTorrent clients using various scenarios including live BitTorrent swarms. Our results show that FairTorrent provides up to two orders of magnitude better fairness, up to five times better download performance for contributing peers, and 60-100% better performance on average in live BitTorrent swarms. We show analytically that for a number of upload capacity distributions, in a n-node FairTorrent network no peer is ever owed more than 0(log(n)) data blocks with high probability.

Achieving fair bandwidth allocation in a P2P streaming scenario is even more difficult, as it comes with an additional constraint: each stream update must be received before its playback deadline. We introduce FairStream, a streaming P2P client, that leverages the FairTorrent algorithm for its peer reply policy, and implements additional optimizations for requesting stream updates. We evaluate FairStream on the PlanetLab and compare it against heuristics used by the popular P2P streaming systems, PPLive and GridMedia, as well as BitTorrent's tit-for-tat. We show that for peers that contribute upload bandwidth just above the stream rate, FairStream can reduce packet loss to just 0.04% down from 33-71% experienced by such peers in other systems. In addition, FairStream implements a new algorithm that allows a streaming server to distribute stream updates effectively to the users despite the presence of a large number of free-riders who may not forward these updates. We show that despite as many as 70% of free-riding peers, using its update distribution heuristics, FairStream is able to reduce the packet loss to just 1% for contributing peers as compared to up to 70% packet loss experienced in this scenario by the contributing peers in other systems.

Julia Stoyanovich

Search and Ranking in Semantically Rich Applications

Advisor: Professor Kenneth Ross

Abstract: This thesis proposes novel search and ranking approaches for semantically rich application domains.

The central role of Data Management in today's society may be compared to the role of Physics in early 20th Century when it entered its Golden Age. Data is the raw matter of the Universe of Information, and, in a process analogous to nuclear fusion, data is transformed progressively into information, and then into knowledge.

The advent of the World Wide Web as an information exchange platform and a social medium, both on an unprecedented scale, raises the user's expectations with respect to the availability and ease of access to relevant information. Web users build persistent on-line personas: they provide information about themselves in stored profiles, register their relationships with other users, and express their preferences with respect to information and products. As a result, rich semantic information about the user is readily available, or can be derived, and can be used to improve the user's online experience, making him more productive, more creative, and better entertained online. There is thus a need for context-aware data management mechanisms that support a user-centric data exploration experience, and do so efficiently on the large scale.

In a complementary trend, scientific domains, most notably the domain of life sciences, are experiencing unprecedented growth. The ever-increasing amount of data and knowledge requires the development of new semantically rich data management techniques that facilitate system-wide analysis and scientific collaboration. Literature search is a central task in scientific research. Controlled vocabularies and ontologies that exist in this domain present an opportunity for improving the quality of ranking.

The Web is a multifaceted medium that gives users access to a wide variety of datasets, and satisfies diverse information needs. Some Web users look for answers to specific questions, while others browse content and explore the richness of possibilities. The notion of relevance is intrinsically linked with preference and choice. Individual items and item collections are characterized in part by the semantic relationships that hold among values of their attributes. Exposing these semantic relationships helps the user gain a better understanding of the dataset, allowing him to make informed choices. This process is commonly known as data exploration, and has applications that range from analyzing the performance of the stock market, to identifying genetic disease susceptibility, to looking for a date.

In this thesis we propose novel search and ranking techniques that improve the user experience and facilitate information discovery in several semantically rich application domains. We show how the social context in social tagging sites can be used for user-centric information discovery. We also propose novel ontology-aware search and ranking techniques, and apply them to scientific literature search. Finally, we address data exploration in ranked structured datasets, and propose a rank-aware clustering algorithm that uses semantic relationships among item attributes to facilitate information discovery.

Andrew Wan

Learning, Cryptography and the Average Case

Advisors: Professors Tal Malkin and Rocco Servedio

Abstract: This thesis explores from an average-case perspective problems in computational learning theory and their connection to cryptography. We study the ways that learning problems change when relaxed to the average-case, and perhaps more importantly, how understanding this change enriches and deepens the connections between cryptography and learning. Through this understanding a variety of new results for both learning theory and cryptography are obtained:

1. We give an efficient algortihm which learns monotone DNF on the average.

The problem of learning DNF and even monotone DNF under the uniform distribution (here the distribution refers to the space of examples) is a notoriously difficult one in computational learning theory. We give the first algorithm that learns monotone DNF of arbitrary polynomial size in a reasonable average-case model of learning from random examples only. Our approach relies on the discovery and application of new Fourier properties of monotone functions which may be of independent interest.

2. Mansour's conjecture is true for random DNF.

In 1994, Y. Mansour conjectured that for every DNF formula f on n variables with t terms there exists a polynomial p with t^{0(log (1/epsilon))} non-zero coefficients such that the expected square error of p on f is at most epsilon. We make the first progress on this conjecture and show that it is true for several natural subclasses of DNF formulas including randomly chosen DNF formulas and read-k DNF formulas for constant k. Our result yields the first polynomial-time query algorithm for agnostically learning these subclasses of DNF formulas with respect to the uniform distribution on the Boolean hypercube (for any constant error parameter). Applying recent work on sandwiching polynomials, our results imply that a t^{-0(\log 1/\eps)}-biased distribution fools the above subclasses of DNF formulas. This gives pseudorandom generators for these subclasses with shorter seed length than all previous work.

3. We show that monotone polynomial-sized circuits are hard to learn assuming one-way functions exist. A wide range of positive and negative results have been established for learning different classes of Boolean functions from uniformly distributed random examples. However, polynomial-time algorithms have thus far been obtained almost exclusively for various classes of monotone functions, while the computational hardness results obtained to date have all been for various classes of general (nonmonotone) functions. Motivated by this disparity between known positive results (for monotone functions) and negative results (for nonmonotone functions), we establish strong computational limitations on the efficient learnability of various classes of monotone functions. Our main tool is a complexity-theoretic approach to hardness amplification via noise sensitivity of monotone functions that was pioneered by O'Donnell (JCSS '04).

4. Learning an overcomplete basis: analysis of lattice-based signatures with perturbations.

We propose a general technique for recovering parts of the secret key in lattice-based signature schemes that follow the Goldreich-Goldwasser-Halevi (GGH) and NTRUSign style of design with perturbations. Previously, Nguyen and Regev (Eurocrypt 2006) cryptanalyzed GGH-style signature schemes (including NTRUSign) without perturbations; their attack was by reduction to a learning task they called the hidden parallelepiped problem (HPP). The main problem left open in their work was to handle schemes that use perturbation techniques. We observe that in such schemes, recovery of the secret key may be modeled as the problem of learning an overcomplete basis, a generalization of the HPP in which the number of secret vectors exceeds the dimension. We propose an algorithm which solves random instances of this problem.

Department News & Awards

Two teams sent by Columbia University's Department of Computer Science have ranked among the top of competitors in the Greater New York Region of the 2009-2010 ACM International Collegiate Programming Contest. The teams ranked 2nd and 6th out of 51 teams.

Team Columbia 1
(ranked 2nd):

Jingyue Wu
(PhD, Computer Science)

Varun Jalan
(MS, Computer Science)

Zifeng Yuan
(PhD, Civil Engineering)

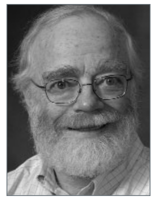
Team Columbia 2
(ranked 6th):

Chen Chen
(PhD, IEOR)

Huzaifa Neralwala
(MS, Computer Science)

Jiayang Jiang
(Junior, Mathematics)

Due to their performance, team Columbia 1 was also selected to be one of 100 teams (chosen from over 7,000 around the world) to advance to the world finals competition, held in Harbin, China from February 1-6. The teams were led by coach **John Zhang** (PhD student, Computer Science).



Professor **Steven Bellovin** is one of four new technical and scientific experts who were recently appointed to the United States Election Assistance Commission's Technical Guidelines Development Committee. The committee is charged under the Help America Vote Act with assisting the Electoral Assistance Commission in developing federal voluntary voting system guidelines that are used to test and certify voting systems.

Professor **Bellovin** was also recently named to the Computer Science and Telecommunications Board of the National Academies. According to the National Academies website, "The Computer Science and Telecommunications Board (CSTB) was established in 1986 to provide independent advice to the federal government on technical and public policy issues relating to computing and communications. It is composed of leaders in the information technology and complementary fields from industry and academia...CSTB is an operating unit within the National Research Council (NRC). The NRC is the principal working arm of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine—three honorific entities to which distinguished experts in their fields are elected by their peers."

PhD student **Ilias Diakonikolas** won Honorable Mention in the 2009 Nicholson Competition of the INFORMS society. The George Nicholson Student Paper Competition is held each year to honor outstanding papers in the field of operations research and the management sciences written by a student. Ilias Diakonikolas received an Honorable Mention Award in the 2009 Nicholson Competition for his paper "Small Approximate Pareto Sets for Biobjective Shortest Paths and Other Problems", coauthored with Professor Mihalis Yannakakis. The paper is published in the SIAM Journal on Computing.

Ashutosh Dutta, a PhD student working with Professor **Henning Schulzrinne**, received a 3rd best paper award at the IEEE International Conference on Internet Multimedia Systems Architecture and Application in Bangalore, India. The paper was titled "Self Organizing IP Multimedia Subsystem" and co-authored by Ashutosh Dutta (Telcordia Technologies, US), Christian Makaya (Ecole Polytechnique de Montreal, CA), Subir Das (Telcordia Technologies, US), Dana A Chee (Telcordia Technologies, US), Fuchun J Lin (Telcordia Technologies, US), Satoshi Komorita (KDDI R&D Laboratories Inc., JP), Tsunehiko Chiba (KDDI R&D Laboratories, Inc., JP), Hidetoshi Yokota (KDDI Labs, JP) and Henning Schulzrinne (Columbia University, US).

PhD student **Charles Han** won a Microsoft Research Fellowship—one of only ten worldwide—for his "great accomplishments and [Microsoft's] confidence in [Charles] as a future leader." Charles attended an award ceremony in March at Microsoft Research in Redmond, Washington, collocated with Microsoft Research's TechFest 2010.

PhD student **Steve Henderson** received the Best Paper Award at IEEE ISMAR 2009, held in Orlando, FL. IEEE ISMAR (International Symposium on Mixed and Augmented Reality) is the premier conference in its field. The paper, "Evaluating the

Department News & Awards (continued)

Benefits of Augmented Reality for Task Localization in Maintenance of an Armored Personnel Carrier Turret," was coauthored by Steve Henderson and Professor **Steve Feiner**.

Professor **Tony Jebara** delivered a keynote speech at the 21st International Conference on Tools with Artificial Intelligence (ICTAI) which was held in Newark, NJ.

Professor **John Kender** delivered the keynote address to this year's Family Weekend, on October 16, 2009, to the assembled first year CC and SEAS students and their visiting parents. He was the first SEAS professor to be invited to do so. The address is available online at: http://www.student affairs.columbia.edu/parents/communications.php

Professor
Angelos Keromytis

Professor
Sal Stolfo

Professor
Junfeng Yang

Professors **Angelos Keromytis**, **Sal Stolfo**, and **Junfeng Yang** will lead a consortium that includes Stanford University, George Mason University, and Symantec Corp. to develop novel software protection mechanisms in a 4-year, IARPA-funded effort. The project will develop a novel architecture that integrates static analysis, dynamic confinement, and code diversification techniques to enable the identification, mitigation and containment of a large class of software vulnerabilities. The system will permit the immediate deployment of new software and the protection of already deployed (legacy) software by transparently inserting extensive security instrumentation, while leveraging concurrent program analysis and runtime profiling data to gradually reduce the performance cost of the instrumentation by allowing its selective removal or refinement.

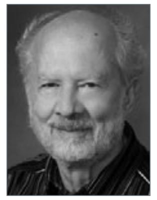
Henry and Gertrude Rothschild Professor of Computer Science **Kathleen McKeown** was selected as one of three recipients of a "Women of Vision Award" by the Anita Borg Institute of Women and Technology (ABI). The recipients will honored for their accomplishments and contributions as women in technology at ABI's fifth annual Women of Vision Awards Banquet at the Mission City Ballroom, Santa Clara, CA on May 12, 2010.

Professor **McKeown** also hosted a NACLO (North American Computational Linguistics Olympiad) site at Columbia in February. The NACLO is an Olympiad for high school students to expose them to computational linguistics and problem solving. More than 1100 students participated in the Olympiad across the country, including 58 students participating at Columbia.

T.C. Chang Professor of Computer Science and Computer Science Department Chair **Shree Nayar** has been awarded Carnegie Mellon University's 2009 Alumni Achievement Award, which recognizes an individual for exceptional accomplishments that have brought honor to the recipient and to Carnegie Mellon. He is being recognized for his "pioneering research contributions and teaching in the field of computer vision."

Professor **Rocco Servedio** was promoted to the rank of Associate Professor with tenure.

Professor **Sal Stolfo** was guest editor of a Special Issue of the IEEE Security and Privacy Magazine on Insider Threats (Nov/Dec 2009). Professor Stolfo also organized and chaired the National Cyber Defense Financial Service Workshop, which was held in October 2009 at the Financial Service Roundtable in Washington, DC. The official workshop report can be found at www.cs.columbia.edu/ncdi-fi-workshop and http://ncdi.nps.edu/.



Edwin Howard Armstrong Professor of Computer Science **Joseph F. Traub** has been appointed to the Division Committee on Engineering and Physical Sciences (DEPSCOM) of the National Academies in Washington, DC. The Committee provides advice and strategic insights to boards and standing committees within its purview. The DEPS portfolio ranges from disciplinary boards such as mathematics, physics, computer science, and astronomy to boards and standing committees serving each of the major military services as well as the intelligence community and the Department of Homeland Security.

Also, after 10 years of service Professor **Traub** has stepped down as Chair of the Computer Science and Telecommunications Board (CSTB). He served as founding chair from 1986 to 1992 and served again from 2005 to 2009.

Postdoc **Sean White** was named one of this year's 2009 Tech Award Laureates for his work addressing environmental issues. The Tech Awards 2009, a humanitarian program recognizing technological solutions aimed at worldwide challenges, selected 15 Laureates from a pool of 650 nominations representing 66 countries. Dr. White won for his work on the mobile, handheld, and augmented reality versions of the Electronic Field Guide. The 2009 Tech Awards Laureates represent regions as diverse as Nigeria, Brazil, Great Britain, the United States and Bangladesh. The Laureates and former Vice President Al Gore, this year's James C. Morgan Global Humanitarian Award recipient, were recognized at The Tech Awards Gala on November 19th at the San Jose McEnery Convention Center.

Alumni News

Knarig Arabshian (PhD '08) is a Member of Technical Staff at Alcatel-Lucent Bell Labs. Initially, she joined the team in Antwerp, Belgium. After working there for almost a year, she recently transferred to Murray Hill, NJ and resides once again in the Columbia University neighborhood. Her current research is focused on building a system that uses ontologies for context-aware tagging of resources on the Internet. She is also working on describing and composing services using ontologies in order to create personalized mashup applications.

German Creamer (PhD'06) writes, "After I left school in 2006, I joined the Risk, Information and Banking division of American Express as a manager, and then I was promoted to senior manager. I worked in projects related to data mining and direct marketing and enterprise-wide risk management. I recently co-authored with my former advisor Sal Stolfo the paper "A Link Mining Algorithm for Earnings Forecast and Trading" at Data Mining and Knowledge Discovery, and also co-authored with my former advisor Yoav Freund the paper "Automated Trading with Boosting and Expert Weighting" which is accepted for publication at the Quantitative Finance journal. I recently joined the Stevens Institute of Technology as an associate professor in quantitative finance and financial engineering.

Homin Lee (PhD'09) is doing postdoctoral research with Professor Adam Klivans at UT-Austin.

Adjunct Associate Professor **Alexander J. Pasik** (BA '82, MS '84, PhD '89) has joined IEEE as the Chief Information Officer and Staff Executive, Information Technology. For the past three years, Alex has been leading all information technology activities worldwide for the Solomon R. Guggenheim Foundation, with museums in New York, Venice, Bilbao, and Berlin. Previously, he was CEO and founder of Pasik Advisory Group, where his clients included Bain & Company and Citigroup. He also was a direct admit partner with Ernst & Young LLP, where he focused on enterprise architecture and new technology implementations for clients in a variety of industries. Prior to that, he served as Vice President and Director of IT Equity Research at Lazard Freres & Company, and was ranked in the top ten of U.S. software analysts in 1997. Alex's PhD thesis advisor was Professor Sal Stolfo.

Julia Stoyanovich (PhD '09) was awarded a highly selective Computing Innovation Fellowship. She has a postdoctoral position at University of Pennsylvania where she is working with Professor Susan Davidson.

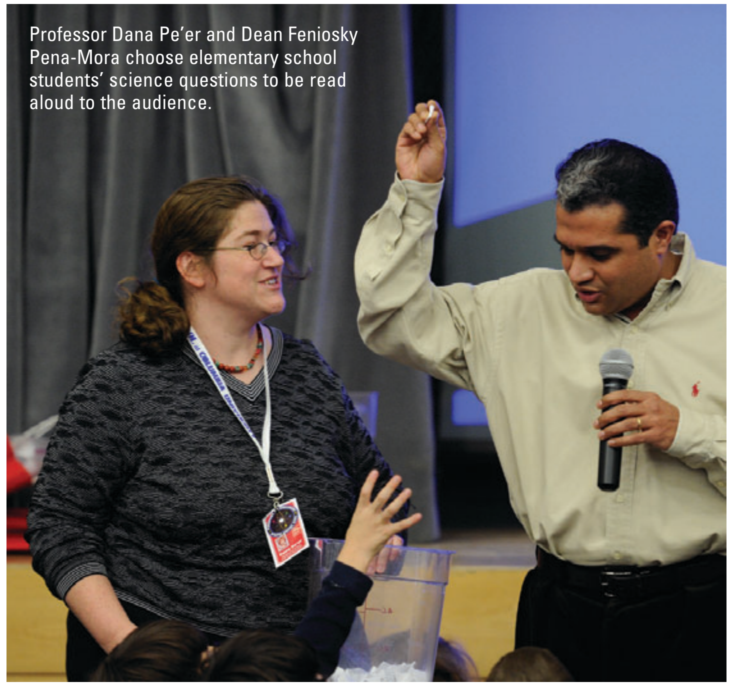
Michelle Zhou (PhD'99) was named an ACM Distinguished Scientist. Dr. Zhou is a research manager at IBM T.J. Watson Research Center, where she manages the department of intelligent multimedia interaction. ACM Distinguished Scientists are named in recognition of their individual contributions to both the practical and theoretical aspects of computing and information technology. The ACM Distinguished Member program, of which the ACM Distinguished Scientists program is a part, can recognize the top 10 percent of ACM worldwide membership based on professional experience as well as significant achievements in the computing field.

Computer Science Department Faculty Participate in **Elementary School Science Expo**

On Saturday February 6, more than 30 scientists from Columbia University and elsewhere led a Science Expo at The School at Columbia, a University-affiliated elementary school in Morningside Heights.

Each scientist manned an interactive museum-style exhibit and explained the great questions that motivate and inspire their work to hundreds of elementary school attendees and their families.

The Computer Science Department was well-represented at the Science Expo. Professor Dana Pe'er co-organized the event, and Professors Steven Feiner, Dana and Itsik Pe'er, and Rocco Servedio all led exhibits.



Professor Dana Pe'er and Dean Feniosky Pena-Mora choose elementary school students' science questions to be read aloud to the audience.

Professor Steven Feiner adjusts a visitor's headset as part of his interactive exhibit on virtual reality.

Professor Rocco Servedio explains the finer points of dynamic programming.

Professor Itsik Pe'er helps elementary school students decode the mysteries of DNA.