CS@CU

NEWSLETTER OF THE DEPARTMENT OF COMPUTER SCIENCE AT COLUMBIA UNIVERSITY VOL.12 NO.1 SPRING 2016

New Faces at CUCS

ALEXANDR ANDONI
ASSOCIATE PROFESSOR
OF COMPUTER SCIENCE

Advancing the algorithmic foundations of massive data

The impact of huge data sets is hard to understate, responsible for advancing almost every scientific and technological field, from machine learning and personalized medicine, to speech recognition and translation.

The flip side of the data revolution is that massive data has rendered many standard algorithms computationally expensive, necessitating a wholesale rethinking of even the most basic computational methods. To take one example: Nearest neighbors search, the classic method since the 70s for finding similarity among objects, is straightforward for small data sets: compare every object to every other one and compute their similarity for each pair. But as the number of objects increases into the billions, computing time grows quadratically, making the task prohibitively expensive, at least in terms of traditional expectations.

Alexandr Andoni, a theoretical computer scientist focused on developing algorithmic foundations for massive data, sees the need to reframe the issue: "The question today is not 'what can we solve in polynomial time?' but 'what is possible in time proportional to data size, or even less?' With all this data, what is the best we can do given the resources? Fortunately, vast improvements are possible in both theory and practice once we settle for approximate answers."

Rather than searching an entire data set for the single most nearest neighbor, a search would go much faster if objects were pre-grouped according to some shared attribute, making it easy to zero in on just the small subset of objects most likely to contain the most similar neighbor. The new challenge then becomes: what attribute shall we use to make such a pre-grouping maximally efficient. The speed-up thus gained reverberates across a wide range of computational methods since nearest neighbors search is ubiquitous and serves as a primitive in higher-level algorithms, particularly in machine learning.

More generally, in the same spirit of relying on (approximate) attributes to speed operations, Andoni has developed a theory of sketching that represents complex objects by smaller,

simpler "sketches" that capture the main structure and essential properties of the original objects yet use less (sublinear) space and time to compute. For many tasks, such as estimating similarity of a pair of objects, a sketch may work just as well as a fully realized object. While relaxing strict formulations is happening generally throughout the community in most part by necessity, Andoni is carrying the idea further and is in the forefront of those inventing new primitives and new data structures that explicitly incorporate the concept of sketches.

In early work applying a sketch primitive (Locality Sensitive Hashing) to nearest neighbor search, Andoni in 2006 with Piotr Indyk was able, for the most basic Euclidean distances, to improve over a seminal 1998 algorithm widely used for classification. The *Communications of the ACM* later (2008, vol. 51) hailed the new primitive as a breakthrough technology that allowed researchers to revisit decades-old problems and solve them faster. Few expected more progress to be possible. Yet when Andoni again revisited the problem (in papers published in 2014 and in 2015), he unexpectedly made more headway, this time by using data-dependent, rather than random, hash functions, a novel idea not previously conceptualized as something useful for improving algorithms.

More seems to be possible by continually re-examining and applying fresh perspectives to classic, well-studied problems. It's one reason, Andoni is returning to academia after four years at Microsoft Research Silicon Valley, which closed in 2014. One point in Columbia's favor was the Data Science Institute (DSI) where exceptional researchers from across the university and diverse disciplines study different aspects of data.

Says Kathy McKeown, director of the DSI: "We're thrilled to have Alex join the Data Science Institute. His research is directly relevant to big data applications where the ability to provide computationally efficient solutions depends on the development of new algorithms."

Andoni also looks for inspiration from students. "You feel the new energy. Students are excited, and that excitement and enthusiasm is invigorating. It leads you to think about even things you've already checked off, to believe there might be new ways of doing things. You want to try again."

SUMAN JANA
ASSISTANT PROFESSOR
OF COMPUTER SCIENCE

Protecting security and privacy in an age of perceptual computing

"I like to break things. To open a thing and see its inner workings and really understand how it works... and then break it in some clever way that maybe no one else thought of and in the worst possible way that doesn't seem possible, but is."

It's a hacker mindset, and **Suman Jana** is a hacker—of a sort. Though he admits enjoying the thrill of destruction and subversion, Jana wants to make systems more secure, and companies have hired him to find security flaws in their systems so those flaws can be fixed, not exploited. This semester he joins the Computer Science department to more broadly research issues related to security and privacy. It's a wide-open field.

Two years ago, for his thesis he looked hard at the security risks inherent in perceptual computing, where devices equipped with cameras, microphones, and sensors are able to perceive the world around them so they can operate and interact more intelligently: lights that dim when a person leaves the room, games that react to a player's throwing motion, doors that unlock when recognizing the owner.

It all comes at a cost, of course, especially in terms of privacy and security.

"Features don't come for free; they require incredible amounts of data. And that brings risks. The same data that tells the thermostat no one is home might also be telling a would-be burglar," says Jana.

What data is being collecting isn't always known, even by the device manufacturers who, pursuing features, default to collecting as much data as they can. This data is handed off to gaming, health, home monitoring, and other apps: not all are trusted; all are possible hacking targets.

There is no opting out. The inexorable trend is toward more perception in devices and more data collection, with privacy and security secondary considerations. For this reason, Jana sees the need for built-in privacy protections. A paper he co-authored, *A Scanner Darkly*, shows how privacy protection might work in an age of perceptual computing.

"Can we disguise some data? Should a camera for detecting hand gestures also read labels on prescription medicines inadvertently left in the camera's view? Would an app work just as well if it detected approximate contours of the hand? If so, we can pass on lower-resolution data so the prescription label isn't readable."

Jana's opinion is that users should decide what data apps are able to see. His DARKLY platform—named after a dystopian novel by Philip K. Dick—inserts a privacy protection layer that intercepts the data apps receive and displays it in a console so users can see and, if they want, limit how much data is passed on to the app. The platform, which integrates with the popular computer vision library OpenCV, is designed to make it easy for companies to implement and requires no changes to apps. The DARKLY paper, called revolutionary, won the 2014 PET Award for Outstanding Research in Privacy Enhancing Technologies.

Making it easy to build in safety and privacy mechanisms is critical. Manufacturers have little incentive to construct privacy protections; in any case, determining what data is sensitive is not easy. A single data point by itself—a random security photo of a passerby, for instance—might seem harmless, but combined with another data point or aggregated over time—similar photos over several weeks—reveals patterns and personal behaviors. The challenge to preventing security leaks is first finding them amidst a deluge of data; a single image of a prescription label or credit card might be hidden within entire sequences of images.

Perceptual computing is rife with other such vulnerabilities made possible by devices that see, hear, and sense what goes on in their immediate environment; but the landscape of vulnerabilities is even larger than it would appear since perceptual computing, while creating new vulnerabilities, inherits all the old ones associated with any software, namely buggy code. While Jana gets a bigger space to explore, for the rest of us, it spells potential privacy disaster.

Preventing such an outcome will come from enlisting help from other technology experts. "For finding images with hidden personal information, we need classifiers. Machine learners over the years have learned how to train classifiers to recognize spam, recommend movies, and target ads. Why not train classifiers to find security risks? I'm fortunate to be working within Columbia's Data Science Institute alongside machine learners who can build such classifiers." For guarding against buggy code, Jana imagines adapting program analysis, an existing technology for automatically finding software bugs, so it specifically searches out those bugs that concern security and privacy.

Technology alone, however, isn't the answer. Companies are unlikely to fix privacy problems unless pressured by the public, and Jana sees his role encompassing the policy arena, where he will work to propose and enact workable regulations and legislation to protect data and security.

At least for perceptual computing, Jana says there an opening to do something about privacy risks. "The field is still relatively new, and we have the chance to build in security from the beginning and make life better so people can trust these devices and use them."

EUGENE WU
ASSISTANT PROFESSOR
OF COMPUTER SCIENCE

Fast, accurate enough for the human in the loop: Visualizing and interacting with big data sets

For exploring complex data sets, nothing matches the power of interactive visualizations that let people directly manipulate data and arrange it in new ways. Unfortunately, that level of interactivity is not yet possible for massive data sets.

"Computing power has grown, data sets have grown, what hasn't kept pace is the ability to visualize and interact with all this data in a way that's easy and intuitive for people to understand," says **Eugene Wu**, who recently received his PhD from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), where he was a member of the database group.

Speed is one important component for visualizing data, but there are others, such as the ease with which interactive visualizations can be created and the ability to help understand what the results actually say. For his PhD thesis, Wu tackled the latter problem by developing a visualization tool that automatically generates explanations for anomalies in a user's visualization. This is important because while visualizations are very good at showing what's happening in the data, they are not good at explaining why. A visualization might show that company expenses shot up 400% in a single month, and an analyst would naturally want to understand what types of expenditures are responsible. However, the monthly statistic is often computed from thousands or millions of input data points, and identifying a simple description of the exact subset causing the spike (e.g., California shops overspent their budgets) requires laborious, error-prone effort.

Now starting at Columbia, Wu is broadening the scope of his research and is among the first looking at the challenging problems in the overlap between databases and how people want to interact with and visualize the data in those databases. Visualization systems currently being built must take an all-or-nothing approach. "You either get performance for small data sets using a small set of fixed interactions, or you get full expressiveness with SQL and queries but you have to wait and give up interactivity."

Part of the problem is that the database and the visualization communities have traditionally been separate, with the database side focusing on efficient query processing and accuracy, and the visualization community focusing on usability and interactions. Says Wu, "If you look at visualizations from a database perspective, a lot of it looks like database operations. In both cases, you're computing sums, you're computing common aggregates. We can remove many of the perceived differences between databases and visualization systems." Wu wants to bridge the two sides to operate more closely together so both consider first the expectations and requirements of the human in the loop.

For instance, what does database accuracy mean when a human analyst can't differentiate 3.4 from 3.45 in a scatterplot? A slight relaxation of accuracy requirements—unnoticeable to users—would conserve resources while speeding up query operations. In understanding the boundary between what a human can perceive and what amounts to wasted computations, Wu hopes to develop models of human perception that are both faithful to studies in the Human Computer Interaction and Psychology literatures, and applicable to database and visualization system performance.

On the visualization side, less attention has been paid to the programming languages (like JavaScript) used to construct the visualizations; consequently, visualizations are hard to write, to debug, and even harder to scale. A similar situation once prevailed in the database world, where application developers wrote complex and brittle code to fetch data from their databases; but the invention of SQL, a high-level, declarative language, made it easier for developers to express relationships within the data without having to worry about the underlying data representations, paving the way towards today's ubiquitous use of data.

For Wu, the natural progression is to extend the declarative approach to interactive visualizations. With colleagues at Berkeley and University of Washington, Wu is designing a declarative visualization language to provide a set of logical operations and mappings that would free programmers from implementation details so they can logically state what they want while letting the database figure out the best way to do it.

A declarative language for visualization would have additional positive benefits. "Once you have a high-level language capable of expressing analyses, all of these analysis tools such as the explanatory analysis from my thesis is in a sense baked into whatever you build; it comes for free. There will be less need for individuals to write their own ad hoc analysis programs."

As interactions become portable and sharable, they can be copied and pasted from one interactive visualization to another for someone else to modify. And it becomes easier to build tools, which fits with Wu's focus in making data accessible and understandable to all users.

"When a diverse group of people look at the same data, the questions you get are more interesting than if just other computer scientists or business people are asking questions." One of the attractions for Wu in coming to Columbia is the chance to work within the Data Science Institute and collaborate with researchers from across the university, all sharing ideas on new ways to investigate data. "Columbia has a huge range of leaders in nearly every discipline from Journalism, to Bioinformatics to Government studies. Our use of data is ultimately driven by the applications built on top, and I'm excited about working on research that can help improve and benefit from the depth and breath of research at the university."

Linda Crane

PAUL BLAER
LECTURER IN DISCIPLINE

Columbia this fall promoted **Paul Blaer** from adjunct professor to Lecturer in Discipline, a full-time faculty position that makes teaching Blaer's primary focus, something he's wanted for a long time.

Hiring Blaer full time is not exactly a stab in the dark for Columbia where Blaer is a well-known quantity. Since he was 3, he has been floating around campus. His father is physics professor Allan Blaer, who did both undergraduate and graduate work at Columbia and who—after teaching stints at Princeton and Swarthmore—returned to Columbia where his son would likewise attend as both undergraduate and graduate student.

Paul Blaer did his PhD research in the area of mobile robotics and 3D vision, working in Peter Allen's lab. It was there, while a grad student leading recitations, that he got his first taste of teaching. He knew immediately that teaching was what he wanted to do. For him, it was the fun stuff, a chance to engage with students, to think on his feet to get them work through problems themselves. For three years before graduating, he was a preceptor, running classes and seeing results from the front of a class.

With teaching in mind, Blaer originally planned to seek a position at a small four-year college, but the combined draw of Columbia and New York City proved strong, and Blaer, knowing Columbia as well as he did, figured someday a teaching opportunity would open up. Until it did, there were other ways on campus for him to contribute.

In Allen's lab Blaer had been doing systems work—his skills for controlling his own computing environments scaled up for a lab of 50 or more—which led to a full-time position at Computing Research Facilities (CRF); precepting led to part-time adjuncting. For seven years now, Blaer has been teaching introductory computer science classes part-time while working at CRF full time to help faculty design and build backend systems for all types of research projects.

With his new position, the mix gets recalibrated: teaching becomes full-time and CRF part-time.

"I'm thrilled to be working full time with students here at Columbia. It's the best of both worlds: a large university environment with highly motivated students, yet like a college professor I have this direct interaction with the students, which is the favorite part of my job."

Blaer knows the classes, the students and faculty, the projects, and how the computer systems are set up; in a department dependent on systems, that's better than knowing where the bodies are buried. He's involved also in the administrative aspects that touch on teaching; he is Director of Undergraduate Studies for BS Programs and is active in the Science Honors Program for area high-school science and math students.

Deep institutional and systems knowledge is all well and good, but a lecturer first and foremost has to be able to teach. Blaer has that angle covered especially well. As someone who genuinely cares about teaching, he pays attention to what resonates with students and what doesn't, and strives to keep his lectures engaging, using humor and real-life stories from his own research to keep students interested. That he succeeds is clear from student comments on the Columbia Underground Listing on Teacher Abilities (CULPA) site, where Blaer has earned a silver nugget for his teaching and approachability.

"We're thrilled to have Paul join the faculty full-time as a lecturer. The department has rock-solid confidence in his classroom skills because we have the strongest possible kind of evidence—actual results over several years," says Rocco Servedio, chair of the Computer Science department.

ANSAF SALLEB-AOUISSI
LECTURER IN DISCIPLINE

The increasingly data-centric approach in all aspects of science and technology means students need to learn what algorithms and methods can stand up to the immense scale of today's data sets. Teaching computer science from the perspective of large data sets is the job of **Ansaf Salleb-Aouissi**. A data scientist from before the term was commonly understood, Salleb-Aouissi has worked with all types of data on projects ranging from geology and geographic information systems early in her career, to social sciences and urban design, and more recently to medical informatics and to education.

"The common denominator is data. The context may be different and the goals may be different, but at the end of the day, data is data and you try to leverage that data somehow to learn something new," says Salleb-Aouissi.

An associate research scientist at Columbia's Center for Computational Learning Systems (CCLS) since 2006, she has worked on both fundamental research into new machine learning and data mining algorithms and methods as well as real-world applications of those methods.

Many of her projects are predictive in nature, forecasting when power-grid failures are likely to occur in one case, and in another predicting which expectant mothers are most, or least, likely to deliver preterm. In this last example, Salleb-Aouissi, with support from the National Science Foundation Smart and Connected Health program, used advanced machine-learning methods to vastly expand the number of risk factors to be considered, including socio-economic, psychological and behavioral factors.

Prediction is also at the heart of her most recent (and current favorite) project: a web browser optimized for self-learning. "We want to create a personalized self-learning experience by sifting through huge number of search results to identify and return those customized for student's learning preferences—whether they be videos, books, blogs—and that also fit within the student's short or long time constraints. The challenge here, as it was with the preterm study, is making all these different and heterogeneous resources work together in a system. It's an ambitious project and I am very excited to work on it. More so because it is a link between my research and my teaching."

Though research forms the bulk of her recent work, teaching has also been a component. Post-PhD, she worked as an adjunct professor at the University of Orléans and discovered how much she enjoyed interacting with students. She would have gladly accepted the position of assistant professor except for her plans to eventually move to the US. Instead she took the offer of a Postdoctoral Fellowship at the prestigious research lab INRIA (French National Institute of Computer Science and Control) at Rennes, France. There she did more fundamental investigation of new algorithms, particularly new methods for quantitative association rules, but also for frequent patterns matching, ranking, characterization, and action recommendation.

While still at INRIA, she applied to the CCLS for an open position. Though that position filled quickly, David Waltz, then director of the CCLS, took note of her INRIA fellowship and her growing publications list and contacted her when a different position came up. She and Waltz later collaborated on a number of papers and projects. "Dave smoothed my transition to the CCLS and helped make it an enriching experience where I could grow and learn. I will always be grateful to him."

Once settled in at the CCLS, she was able to get back into teaching, adjuncting in the Computer Science department, teaching courses in data science, discrete math, and artificial intelligence. As a lecturer, teaching will now be her primary focus, but she will continue doing research, which will now serve a double purpose. "I like to deliver my lecture in an engaging and interactive way, my own way and to keep the material fresh and alive so students actively absorb it rather than just be passive recipients. My own research may serve to give students a peek into what you can do with computer science, and I hope that can motivate them and spark their interest so they learn now so they can do later."

Allison Bishop Wins **NSF CAREER Award**

Allison Bishop has been awarded a five-year $500,000 National Science Foundation (NSF) CAREER award to develop tools for designing and proving the security of new cryptographic systems.

Allison Bishop, an assistant professor within Columbia's Computer Science Department and a member of the Data Science Institute, has been awarded a five-year $500,000 National Science Foundation (NSF) CAREER award to develop tools for designing and proving the security of new cryptographic systems. The CAREER award is the NSF's most prestigious honor designed to support junior faculty who exemplify the role of teacher-scholars through their outstanding research and excellent teaching.

With the award, Bishop will build on her current research into provably secure cryptographic systems that can accommodate various levels of access to data, thus allowing different people to access different data within the same data source. The need for fine-grained control over data access has never been greater as vast amounts of sensitive data have to be simultaneously shared and protected, such as when a hospital needs to see almost all of a patient's data but an insurance company needs to see only what procedures have been done.

Achieving more nuanced cryptographic capabilities means also enhancing the mathematical foundations that support such capabilities so that security at each access level can be enforced in a provable way. For this, Bishop is looking to integrate recent advances in lattice cryptography with her progress in designing security reductions.

Bishop will also use the award to provide an entry point and training ground for emerging young scientists of all ages, giving advanced graduate classes a more integrated view of cryptographic system design principles, while opening valuable research opportunities to students at both the undergraduate and graduate levels. The educational outreach aspect extends to students of elementary-school age, for whom Bishop is producing a book that uses a fairy-tale setting to introduce mathematical reasoning. Motivating others will help ensure faster progress towards a flexible and more unified theory of cryptography to meet the mounting challenges of huge data sets, cloud computing, and other emerging data systems.

Linda Crane

Three Columbia Engineering Professors **Win Sloan Fellowships**

Matei Ciocarlie

Roxana Geambasu

Daniel Hsu

Three Columbia Engineering professors—**Matei Ciocarlie** (Mechanical Engineering), **Roxana Geambasu** (Computer Science), and **Daniel Hsu** (Computer Science)—have won 2016 Sloan Research Fellowships. They are among 126 outstanding young scientists and scholars announced by the Alfred P. Sloan Foundation.

Awarded annually since 1955, the Sloan Fellowships honor early-career scientists and scholars whose achievements and potential identify them as rising stars, the next generation of scientific leaders. The 2016 fellows, who receive $50,000 to further their research, have been drawn from 52 colleges and universities in the United States and Canada, and represent a wide range of research interests.

Matei Ciocarlie's research is focused on developing versatile manipulation and mobility in robotics, in particular on building dexterity into robotic hands, and he sees robotic manipulation in unstructured environments as a critical research area. "We aim to discover how artificial mechanisms can interact with the world as skillfully as biological organisms," he notes. So far, robotic applications that have had significant impact (especially in industrial domains) have done it by being fast, precise, and tireless. In order to advance to less constrained domains, robots need to become more versatile and learn to handle variability, or be more intelligent in their environment interaction. "True dexterity in interacting with the world will play a role in the more general problem of developing cognitively advanced computers and machines," Ciocarlie adds. His Robotic Manipulation and Mobility Lab is working on a range of applications, from versatile automation in manufacturing and logistics to mobile manipulation in unstructured environments to assistive and rehabilitation robotics in healthcare. He is a member of the Data Science Institute and has won numerous prestigious honors, including the 2013 IEEE Robotics and Automation Society Early Career Award, a 2015 Young Investigator Program grant from the Office of Naval Research, a 2015 NASA Early Stage Innovations grant, and a 2016 CAREER Award from the National Science Foundation.

Computer Science Professor Roxana Geambasu is working to ensure data security and privacy in an era of cloud computing and ubiquitous mobile devices—technologies upon which billions of users rely to access and host sensitive data and which have become easy targets for theft, espionage, hacking, and legal attacks. Our mobile devices are packed with confidential information under operating systems that never securely erase data. And at the other end, cloud services not only accumulate endless logs of user activity, such as searches, site visits, and locations, but also keep them for extended periods of time, mine them for business value, and at times share them with others—all without the user's knowledge or control. Geambasu, a member of the Data Science Institute, is working to identify the security and privacy risks inherent in current mobile and web technology and designs, and constructing systems to address those problems. Her research spans broad areas of systems research, including cloud and mobile computing, operating systems, and databases, all with a focus on security and privacy. She integrates cryptography, distributed systems, database principles, and operating systems techniques and works collaboratively in developing cross-field ideas in order to solve today's data privacy issues.

A computer science professor at Columbia Engineering and a member of the Data Science Institute, Daniel Hsu develops machine learning algorithms that have been used in automated language translation, personalized medicine, and privacy transparency systems. His work making computers smarter was recently recognized in IEEE's *Intelligent Systems* magazine. Hsu specializes in a branch of machine learning called interactive learning, which turns an algorithm loose on a small set of hand-labeled data. When the algorithm encounters a term it doesn't recognize, it requests a label, massively speeding up the training process. As a graduate student in the late 2000s, Hsu helped develop an active learning method that was later applied to electrocardiograms, reducing the amount of training data needed by 90 percent. His work on Hidden Markov Models has been applied in genomics to understand the role of gene regulation in disease, and how the chromatin packaging a cell's DNA may be implicated. More recently, he helped develop a tool to bring greater transparency to how personal data is used on the Web.

The Sloan Fellowships are awarded in eight scientific and technical fields—chemistry, computer science, economics, mathematics, computational and evolutionary molecular biology, neuroscience, ocean sciences, and physics. Candidates are nominated by their fellow scientists and winning fellows are selected by an independent panel of senior scholars.

Holly Evarts and Kim Martineau

Steve Bellovin Named **First Technology Scholar** by the Privacy and Civil Liberties Oversight Board

Computer Science Professor **Steven Bellovin** has been appointed the first Technology Scholar by the Privacy and Civil Liberties Oversight Board (PCLOB). A nationally recognized expert in technology and network security, Bellovin has examined technology and its privacy implications throughout his career.

"I'm delighted to be joining PCLOB," says Bellovin. "Modern intelligence agencies rely heavily on technology; many of their collection and analysis systems are based on software. My role will be to help the Board members understand these mechanisms and their implications."

Bellovin has taught computer science at Columbia since 2005. During more than 20 years at Bell Labs and AT&T Labs Research, he focused on network security firewalls, protocol failures, routing security, and cryptographic protocols. He is a member of the National Academy of Engineering and the Computer Science and Telecommunications Board of the National Academies. He has served on the Science and Technology Advisory Committee of the U.S. Department of Homeland Security, the Technical Guidelines Development Committee of the U.S. Election Assistance Commission, and as Chief Technologist of the Federal Trade Commission. He also has authored numerous publications and has received awards and national recognition for his work. He holds a BA from Columbia University and an MS and PhD in Computer Science from the University of North Carolina at Chapel Hill.

In announcing the appointment, PCLOB Chairman David Medine said, "I am pleased that Professor Bellovin will be joining our team as our first Technology Scholar. His vast knowledge and significant expertise in both the private and public sectors will be of great benefit to our agency's mission to ensure that the federal government's efforts to prevent terrorism are balanced with the need to protect privacy and civil liberties."

The PCLOB is an independent agency within the executive branch established by the Implementing Recommendations of the 9/11 Commission Act of 2007. The bipartisan, five-member Board is appointed by the President and confirmed by the Senate. The PCLOB's mission is to ensure that the federal government's efforts to prevent terrorism are balanced with the need to protect privacy and civil liberties.

Vishal Misra Named **IEEE Fellow**

For contributions to "network traffic modeling, congestion control and Internet economics," **Vishal Misra** has been named a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the highest grade of IEEE membership and limited every year to one-tenth of one-percent of the total voting membership.

"Throughout my career I have attempted to solve real world problems via mathematical modeling and analysis," says Misra. "While I am deeply honored by this recognition by the IEEE, the credit for it goes to all my great collaborators who have ensured that our work has had an impact."

On the faculty of the Computer Science department at Columbia University, Misra in his research emphasizes the use of mathematical modeling to examine complex network systems, particularly the Internet. It's an approach that has been highly productive from the start. His PhD thesis work on modeling Internet congestion, done in collaboration with colleagues, opened up entirely new directions in TCP analysis and led to better control mechanisms, helping achieve high throughput, low latency, and low packet loss on Internet links. Software that grew out of Misra's PhD thesis is now being deployed in all cable modems worldwide.

In a 2008 paper, he and colleagues used mathematical modeling to examine the pricing policies and profit motives of Internet service providers, in the process identifying at an early stage the economic incentives that would later give rise to paid peering; Misra was thus one of the first in academic circles to warn that network neutrality issues are not resolvable without first understanding Internet economics.

Recently as network neutrality has become a political issue, particularly in the US and India, Misra has actively participated in the public debate, contributing articles and interviews to leading media outlets. Earlier this year, he appeared before the Indian Parliament to present his views on zero rating, a policy contrary to network neutrality. Misra's opinions and expertise are sought not only for his deep technical research, but also for his real-world experience building Internet-based businesses. While still a graduate student, he co-founded the sport website Cricinfo (acquired by ESPN in 2007); more recently he founded the data center storage startup Infinio.

Misra's elevation to IEEE fellow is an important achievement in a career that has previously earned him a National Science foundation CAREER Award, a Department of Energy CAREER Award, and Google and IBM Faculty Awards.

Linda Crane

Julia Hirschberg and David Blei Elected **2015 ACM Fellows**

Julia Hirschberg

David Blei

Two professors in the Computer Science department at Columbia University have been elected 2015 Association for Computing Machinery (ACM) Fellows: **Julia Hirschberg** for "contributions to spoken language processing," and **David Blei**, for "contributions to the theory and practice of probabilistic topic modeling and Bayesian machine learning." The ACM fellowship grade recognizes the top 1% of ACM members for their outstanding accomplishments in computing and information technology or outstanding service to ACM and the larger computing community. This year, 42 have been named ACM Fellows.

Julia Hirschberg is the Percy K. and Vida L.W. Hudson Professor of Computer Science and Chair of the Computer Science Department. She is also a member of the Data Science Institute. Her main area of research is computational linguistics, with a focus on the relationship between intonation and discourse. Her current projects include deceptive speech; spoken dialogue systems; entrainment in dialogue; speech synthesis; speech search in low-resource languages; and hedging behaviors.

"I'm deeply honored to be joining this wonderful group of computer scientists," says Hirschberg. "The ACM has done a wonderful job of supporting and promoting computer science for many years."

Upon receiving her PhD in Computer and Information Science from the University of Pennsylvania, Hirschberg went to work at AT&T Bell Laboratories, where in the 1980s and 1990s she pioneered techniques in text analysis for prosody assignment in text-to-speech synthesis, developing corpus-based statistical models that incorporate syntactic and discourse information, models that are in general use today. She joined Columbia University faculty in 2002 as a Professor in the Department of Computer Science and has served as department chair since 2012.

As of November 2015, her publications have been cited 14,161 times, and she has an *h*-index of 60.

Hirschberg serves on numerous technical boards and editorial committees, including the IEEE Speech and Language Processing Technical Committee and the board of CRA-W. Previously she served as editor-in-chief of *Computational Linguistics* and co-editor-in-chief of *Speech Communication* and was on the Executive Board of the Association for Computational Linguistics (ACL); on the Executive Board of the North American ACL; on the CRA Board of Directors; on the AAAI Council; on the Permanent Council of International Conference on Spoken Language Processing (ICSLP); and on the board of the International Speech Communication Association (ISCA). She is also noted for her leadership in promoting diversity, both at AT&T and Columbia, and broadening participation in computing.

Among many honors, she is a fellow of the Association for Computational Linguistics (2011), of the International Speech Communication Association (2008), of the Association for the Advancement of Artificial Intelligence (1994); and she is a recipient of the IEEE James L. Flanagan Speech and Audio Processing Award (2011) and the ISCA Medal for Scientific Achievement (2011). In 2007, she received an Honorary Doctorate from the Royal Institute of Technology, Stockholm, and in 2014 was elected to the American Philosophical Society.

David Blei is a Professor of Computer Science and Statistics and a member of the Data Science Institute. He is a leading researcher in the field of probabilistic statistical machine learning and topic models, having co-authored (with Michael I. Jordan and Andrew Y. Ng) the seminal paper on latent Dirichlet allocation (LDA), the standard algorithm for discovering the abstract "topics" that occur in a collection of documents. LDA has become an important statistical tool and is used to capture interpretable patterns in a range of applications, including document summarization, indexing, genomics, and image database analysis.

In addition to continuing work on topic models, Blei develops models of social networks, music and audio, images and computer vision, and neuroscience and brain activity. Recent work with students has resulted in efficient algorithms to fit a wide class of statistical models to massive data sets, enlarging the scale of data that can be analyzed using sophisticated methods.

"I am deeply honored to have been elected an ACM fellow," says Blei. "The ACM is a wonderful organization—for many years it has nurtured the fantastic intellectual and community spirit of computer science."

Blei's research has earned him a Sloan Fellowship (2010), an Office of Naval Research Young Investigator Award (2011), the NSF Presidential Early Career Award for Scientists and Engineers (2011), the Blavatnik Faculty Award (2013), and the ACM-Infosys Foundation Award (2013). He is the author and co-author of over 80 research papers.

Before coming to Columbia in 2014, Blei was an Associate Professor of Computer Science at Princeton University. He received his PhD in Computer Science from UC Berkeley and his BSc in Computer Science and Mathematics from Brown University.

Linda Crane

Henning Schulzrinne Named Recipient of **2016 IEEE Internet Award**

Henning Schulzrinne, the Julian Clarence Levi Professor of Mathematical Methods and Computer Science at The Fu Foundation School of Engineering at Columbia University, has been named the recipient of the 2016 IEEE Internet Award for exceptional contributions to the advancement of Internet technology.

Schulzrinne was recognized "for formative contributions to the design and standardization of Internet multimedia protocols and applications." Schulzrinne is particularly known for his contributions in developing the Session Initiation Protocol (SIP) and Real-Time Transport Protocol (RTP), the key protocols that enable Voice-over-IP (VoIP) and other multimedia applications. Each is now an Internet standard and together they have had an immense impact on telecommunications, both by greatly reducing consumer costs and by providing a flexible alternative to the traditional and expensive public-switched telephone network.

"This award also recognizes the work by my students and visitors in the Columbia IRT lab as well as all the other colleagues who contributed to making Internet-based multimedia possible," says Schulzrinne, in referring to the Internet Real-Time (IRT) Lab, which he directs and which conducts research in the areas of Internet and multimedia services.

The Internet award follows on the heel of two other honors recently accorded Schulzrinne. In January, he was named an ACM Fellow, and in December 2014 he received an Outstanding Service Award by the Internet Technical Committee (ITC), of which he was the founding chair. In 2013, Schulzrinne was inducted into the Internet Hall of Fame. Other notable awards include the New York City Mayor's Award for Excellence in Science and Technology and the VON Pioneer Award.

Schulzrinne whose research interests include applied network engineering, wireless networks, security, quality of service, and performance evaluation, continues to work on VoIP and other multimedia applications and is currently investigating an overall architecture for the Internet of Things and making it easier to diagnose network problems. He is also active in designing technology solutions to limit phone spam ("robocalls") and recently testified on this topic before the Senate Special Committee on Aging.

In addition to his research, Schulzrinne is active in public policy and in serving the broader technology community. From 2012 until 2014, he was the Chief Technology Officer for the Federal Communications Committee where he guided the FCC's work on technology and engineering issues and played a major role in the FCC's decision to require mobile carriers to support customers' abilities to contact 911 using text messages. He continues to serve as a technical advisor to the FCC.

Schulzrinne is a past member of the Board of Governors of the IEEE Communications Society and a current vice chair of ACM SIGCOMM. He has served on the editorial board of several key publications, chaired important conferences, and published more than 250 journal and conference papers and more than 86 Internet Requests for Comment.

Linda Crane

Jonathan Gross Retires After 47 Years of Teaching and Research at Columbia

Jonathan Gross retired last semester, following a highly active career that allowed him to indulge his lifelong love of mathematics while doing pioneering work in graph theory, three-dimensional topology, shape modeling, and sociological modeling.

Professor Gross's main specialty is topological graph theory, a math subdiscipline straddling combinatorics and geometry and marked by a strong visual component. In several of his 17 books and in over 100 papers and journal articles, Gross expanded topological graph theory by initiating new programs of investigation and by developing new methods for them, often collaborating with Thomas W. Tucker. Together Gross and Tucker authored the influential and comprehensive T*opological Graph Theory*, which at its release in 1987 represented the state-of-the-art in graph theory. Their objective in writing that book was to create a single source that would provide someone new to topological graph theory with sufficient background to move as quickly as possible into frontier research. It remains a standard reference today.

Gross invented the voltage graph construction in 1973, which is the basis for a concise algebraic specification of infinite families of large graphs and also of placements of such graphs on increasingly complicated surfaces. Gross's joint work with Tucker on its generalization, published in 1977, includes some of the most frequently cited publications in topological graph theory. The name voltage graph plays on the fact that one of the key properties that sometimes

occurs in the specification of placements in surfaces is an algebraic generalization of the Kirchhoff voltage law, which is a property of electrical circuits well known to electrical engineers and physicists. Another paper by Gross and Tucker explains how the voltage graph construction unifies dozens of special cases that occur in the solution of the Heawood map-coloring problem.

Topological graph theory has connections to many other areas of mathematics, including combinatorial and probabilistic models, as well as to knot theory. Since 2009, Gross has been working with Jianer Chen, one of his former Columbia PhD students, to apply topological graph theory to the computer graphics area called shape modeling. Another area that Gross tackled and examined for several years is behavioral and cultural rule systems, for which he developed information-theoretic models and measurement techniques. Working with the eminent British anthropologist Dame Mary Douglas, Gross demonstrated how such high-powered tools can be harnessed to better understand human social behavior. In his book, *Measuring Culture*, Gross and his co-author Steve Rayner describe how to measure information content in societal patterns, making it possible to obtain objective comparisons of different target populations.

For his research, Gross has earned multiple honors and awards: an Alfred P. Sloan Fellowship, an IBM Postdoctoral Fellowship, and numerous research grants from the Office of Naval Research, the National Science Foundation, the Russell Sage Foundation, and, most recently, from the Simons Foundation.

Gross began his formal mathematics education as an undergraduate at MIT, graduating in 1964. From MIT, he went to Dartmouth College where his PhD thesis on three-dimensional topology (1968) solved a published problem of Fields Medalist John Milnor. After graduate school, he joined the Mathematics Department at Princeton University, working with Ralph Fox, renowned for his work on knot theory and three-dimensional topology.

Though primarily a mathematician, Gross had an early interest in computers, and it was in computer science that he felt that his teaching would have greater impact. He has believed since his high school days that computing was for everybody, and his earliest books are concerned with computer programming. It was to set up a computer science curriculum for arts and science students that he was invited in 1969 to join the Statistics Department at Columbia. His first class in introductory computer programming at Columbia had eight students. Within a few years, 300 students in that same course filled the seats in the large lecture room in Havemeyer. The university expanded the computer science contingent that he headed within Statistics one by one, to five faculty members.

In the late 1960s and the 1970s, computer science was also taught by a small nucleus of professors of Electrical Engineering. In 1978-79, while Gross was Acting Chair of Statistics, Dean Peter Likins of SEAS committed funds from a substantial gift to SEAS to found a separate Computer Science Department, which both contingents agreed to join. Merging the computer science course offerings from Statistics and from Electrical Engineering was among the first initiatives that Gross orchestrated for the new department. He strongly encouraged faculty to balance their teaching assignments between undergraduate and graduate levels. His role in starting Columbia's computer science department was fundamental; as the department grew over the years—it now numbers 44 professors and 5 lecturers—Gross was the organizer of department-wide efforts to keep the academic curriculum at the educational forefront. Over the years, he became the keeper of institutional memory.

"Not only did we have no cell-phones or personal computers when I was young, most families did not have a television before 1950. We would start being nice to the rich kid around Thursday, in the hope that he would invite us to watch television at his house over the weekend."

Mathematician, researcher, author, and computer scientist, Gross was also an instructor to thousands of Columbia students. He taught discrete mathematics, graph theory, and combinatorial theory, lecturing with humor and with what he called "enhancement," short historic anecdotes from science and mathematics as well as from his own mathematical career and personal history. "Enhancements" were as integral to his courses as his meticulously put-together notes, often giving students insight into a different time and place.

He proved popular with students, who variously described him as devoted to his work, brilliant, idiosyncratic, and highly quotable.

"When I say a baby-level proof, that's just how mathematicians talk. I don't actually know any babies who can do algebraic topology."

"Negativebplusormi-nusthesquarerootofbsquared-minusfouracovertwoa. You have to say it very quickly, or you'll get it wrong."

"I have no idea what liquid soap will make your dishes sparkle, but I recommend liquid Joy for making high-quality knotted soap bubbles with interesting mathematical properties."

– From a collection of quotes compiled by students

For his excellent teaching, Gross received two SEAS awards; in 1994 he received as well the career Great Teacher Award from the Society of Columbia Graduates.

In late career and retirement, Gross continues his research work with his co-authors around the world. Each year he produces numerous journal papers in topological graph theory, and he continues to travel to national and international mathematics meeting to give talks about his research and to chair sessions in his specialty. One math friend has joked, "Jonathan, you are in danger of flunking retirement." To this, Gross responds that math is too much fun to stop and that he intends to flunk retirement for years to come.

His conclusion of active service at Columbia was marked in December with a dinner amidst remembrances by colleagues and family. Among those who shared their personal stories of Professor Gross, it was perhaps his daughter Rena who most closely articulated how much mathematics infused her father's life when she recounted how, as a child and misbehaving, her father would threaten "Stop, or I'll map you into the complex plane."

Professor Gross plans to be back for the parties.

Inaugural Morton B. Friedman Prize
Honors Robotics Innovator

Jonathan Weisz

Jonathan Weisz, a computer science PhD candidate at Columbia Engineering, has been named the recipient of the inaugural Morton B. Friedman Memorial Prize for Excellence.

Named after the beloved professor and senior vice dean who was an integral part of Columbia Engineering for nearly 60 years, the prize honors undergraduate and graduate students who exemplify "Mort's" legacy of academic excellence, visionary leadership, and outstanding promise for the future.

Weisz's work at the Columbia University Robotics Group, led by Computer Science Professor **Peter K. Allen**, advances real-time grasp planning through brain-computer interfaces. He has developed code for a range of robotics platforms spanning research and industry and published several papers at peer-reviewed conferences. His research with Allen has included measures of grasp stability under uncertainty, "human-in-the-loop" grasping, and data-driven hand design optimization. He also helped manage integrating the various components of the Robotics lab's grasping platform, arm trajectory planning, vision, grasp planning, and tactile sensing.

"The work our lab is doing with brain-computer interfaces and assistive robotics is exploring how far we can push practical, affordable technologies to help people with motor impairments regain some autonomy," said Weisz.

Weisz participated in Phase 1 of DARPA's Autonomous Robotic Manipulation (ARM) Challenge to create a manipulator capable of high-level tasks and adapting to real-world environments with little supervision, as well as the DARPA Robotics Challenge to develop innovative ground robots for use in disaster response operations. Previously, as a student and researcher at Johns Hopkins and the University of Southern California, he contributed to augmented reality projects to combat phantom limb pain and small devices to measure motor impairment of cerebral palsy and osteoarthritis patients.

Professor Friedman, who founded the Division of Mathematical Methods, the precursor to the applied mathematics component of the Department of Applied Physics and Applied Mathematics, chaired the Department of Civil Engineering and Engineering Mechanics for 14 years. In his role as associate dean, vice dean, and senior vice dean, he was in the vanguard of engineering education and helped shaped the School for many decades. He died last year at age 86.

"I'm very humbled by the link to someone who contributed so much," Weisz said. "I'm hopeful that the community of researchers that Dr. Friedman helped build will continue to have an impact."

Jesse Adams

The Morton B. Friedman Memorial Prize for Excellence is awarded periodically to an undergraduate or graduate student who best exhibits Professor Friedman's characteristics of academic excellence, leadership, and outstanding promise.

PhD Student Wins **Google Fellowship**

Riley Spahn

Riley Spahn, a computer science (CS) PhD student working with CS professors **Roxana Geambasu** and **Gail Kaiser**, was recently awarded a North American Google PhD Fellowship for his work on privacy issues. He is one of 15 students chosen from a highly competitive group who represent the next generation of researchers working to solve some of the most interesting challenges in computer science.

"I'm very happy that Google will be supporting my research," says Spahn, who will pursue research on operating and distributed systems with a focus on security, privacy, and data management. "How we manage and control data is a very important aspect of modern life and I'm excited to build tools that allow programmers to manage data in more secure ways and add transparency to how web services put our data to use."

Google created the PhD Fellowship program in 2009 to recognize and support outstanding graduate students doing exceptional work in computer science and related disciplines.

Columbia Engineering's Computer Vision Laboratory Develops Cambits, a Modular Imaging System that Can Transform Into Many Different Cameras

Makoto Odamaki, visiting scientist from Ricoh Corporation, and Computer Science Professor Shree Nayar examine Cambits.

Computer Science Professor **Shree Nayar** and Makoto Odamaki, a visiting scientist from Ricoh Corporation, have developed Cambits, a modular imaging system that enables the user to create a wide range of computational cameras. Cambits comprises a set of colorful plastic blocks of five different types—sensors, light sources, actuators, lenses, and optical attachments. The blocks can easily be assembled to make a variety of cameras with different functionalities such as high dynamic range imaging, panoramic imaging, refocusing, light field imaging, depth imaging using stereo, kaleidoscopic imaging and even microscopy.

"We wanted to redefine what we mean by a camera," says Nayar, who is the T.C. Chang Professor of Computer Science at Columbia Engineering and a pioneer in the field of computational imaging. "Traditional cameras are really like black boxes that take one type of image. We wanted to rethink the instrument, to come up with a hardware and software system that is modular, reconfigurable, and able to capture all kinds of images. We see Cambits as a wonderful way to unleash the creativity in all of us."

Cambit blocks, whose exteriors were 3D-printed, are easy and quick to configure. They are attached through magnets: no screws, no cables. When two blocks are attached, they are electrically connected by spring-loaded pins. The pins carry the power (from a host computer, tablet, or smartphone), data, and control signals.

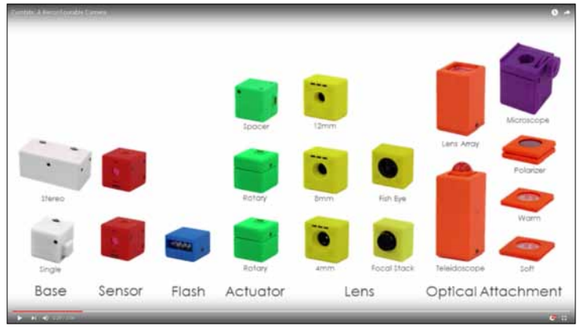
Each block has an ID and when a set of blocks are put together, the host computer recognizes the current configuration and provides a menu of options for what the user might want to do. Cambits is scalable: new blocks can be added to the existing set.

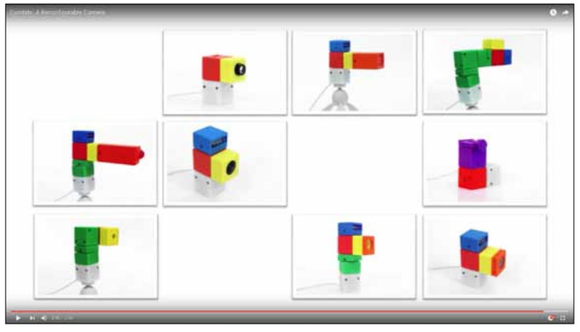
A key aspect of the Cambits design is a circuit board designed by Odamaki that sits inside each block. The board includes a microcontroller, an upstream interface, and a downstream interface. Through the circuit, each block can provide power downstream and receive data upstream. Control signals are conveyed both up and downstream.

"Using our novel architecture, we were able to configure a wide range of cameras," adds Odamaki, who spent two years working with Nayar on the proof-of-concept project. The suite of computational photography algorithms used by Cambits was implemented by a group of MS project students at Columbia Engineering. Odamaki and Nayar are hoping to partner with a manufacturer to bring their concept to the public.

"There are so many exciting advances in computational photography these days," Nayar adds. "We hope this reconfigurable system will open the door to new avenues of creativity, bringing new dimensions to an art form we all enjoy."

Holly Evarts





Cambits, A Reconfigurable Camera: Cambits comprises a set of colorful plastic blocks of five different types—sensors, light sources, actuators, lenses, and optical attachments. The blocks can easily be assembled to make a variety of cameras with different functionalities.

The Future of DNA Sequencing is Already in the Classroom

Halfway through their Ubiquitous Genomics class, 20 students were handed a MinION device, a mobile DNA sequencer the size of two matchboxes laid end to end. This $1000 device, now fully available after being introduced in an early access program, is expected to play an important role in advancing the goal of real-time, on-site DNA sequencing, vastly increasing the applications for DNA sequencing and, just as far-reaching, expanding the number of people who can do DNA sequencing. For their professor, Yaniv Erlich, the device has a more immediate purpose: a teaching tool that gives students direct experience with handling and sequencing DNA samples for themselves. Plus he was curious. What happens when you give smart, ambitious students a new device not yet fully explored?

The parasites were a surprise. In sequencing a food sample premeasured to contain 80% beef and 20% tomato, the students identified the DNA of three parasites (*babesia bigemina, wuchereria bancrofti, onchocerca ochengi*) and duly noted it as part of their assignment. Identifying parasites in food hadn't been the original intent, but when you give students a brand new tool not yet in general use, it's never clear how they are going to use it or what they will find. That's part of the fun, and the learning, too, and it shows the promise of onsite, immediate DNA sequencing.

But it was not all smooth sailing. While students found the accidental parasites, some also misidentified the beef—purchased from a local New York City grocery store—as bighorn sheep. Not a huge leap (both animals are in the same family), but it does give pause to the idea that real-time DNA sequencing will soon be in use at airports to screen passengers.

Classroom encounters with DNA sequencing

Sequencing DNA from food samples was the first of two hackathons in the class Ubiquitous Genomics, offered for the first time at Columbia and developed by Dr. Yaniv Erlich, an assistant professor of computer science at Columbia who is also faculty member of the New York Genome Center.

The class teaches the basics of DNA sequencing with an eye on future sequencing technologies that promise to make DNA identification possible in real time at almost any location.

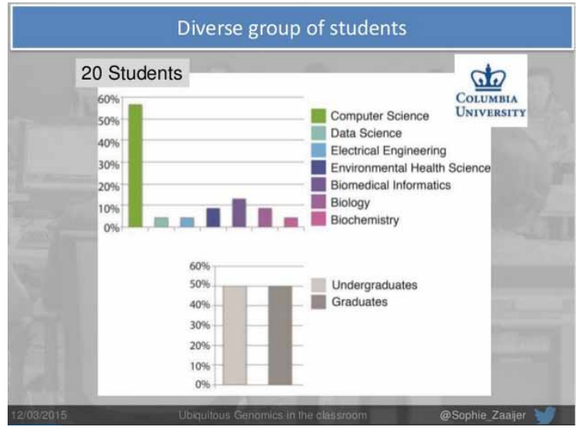
Taught in conjunction with Sophie Zaaijer, a postdoc in Erlich's New York Genome Center lab, the class combines aspects of computer science, biology, electrical engineering, algorithms, and data science, particularly the special challenges of acquiring, storing, and analyzing huge amounts of genomic data. (The first reading assignment was *Big Data: Astronomical or Genomical?* by ZD Stephens and others.)

The class, however, has a major DIY twist. Rather than sending out DNA samples to a lab equipped with $1M sequencing machines, Erlich would have students learn DNA sequencing by actually doing it themselves.

What makes this scenario even imaginable let alone possible is a new, portable DNA sequencing device called a MinION. Inexpensive (approximately $1000), portable, and capable of sequencing DNA in almost real time, the MinION will vastly broaden the applications of DNA sequencing and who can accomplish it.

The MinION uses a sequencing method different from traditional (or sequential) DNA sequencing, which works by first breaking up the DNA into tiny snippets before painstakingly reassembling them, mapping them against a template DNA—a process that can take days and requires a high level of expertise.

Instead, MinION relies on nanopore sequencing, where a single-stranded DNA molecule passes through a small biological pore, or nanopore, embedded in an elec-



Slightly more than half the students were computer science majors.

The MinION is four inches long, weighs 4 ounces, and gets power from a computer's USB port.

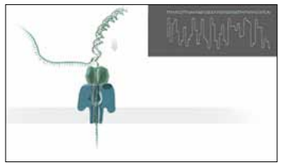


Image credit: Oxford Nanopore Technologies

trical field. As the DNA molecule transits through the nanopore, the individual nucleotides (A, T, G, C) that construct a string of the DNA disrupt the ion current in characteristic ways, creating a profile (called a squiggle) that can be analyzed by software to "decode" the nucleotide sequence, almost in real time.

Erlich was able to procure for his class five MinIONs because the device's manufacturer, Oxford Nanopore Technologies, is interested in exploring the potential applications of the MinION in education. (The class has generated interest among the community growing up around the MinION and was covered by a *GenomeWeb article*.)

Two hackathons count for half the grade

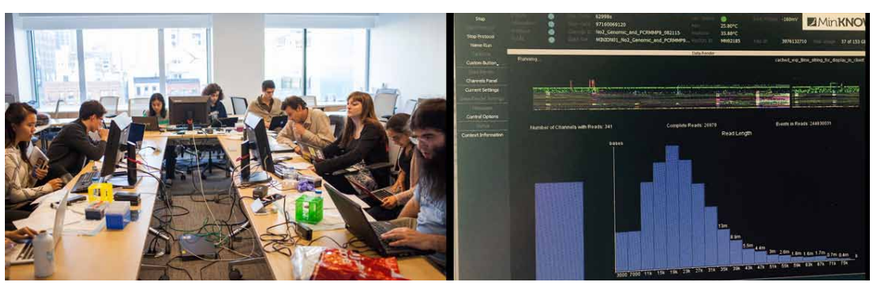
Half the grade would be determined by two hackathons, where the 20 graduate and undergraduate students, working in small groups, would be given the five MinIONs along with five PCs running MinION software. The first hackathon, "Snack to Sequence," required student teams to identify ingredients of a food sample prepared by Zaaijer. In the second, "CSI Columbia," students were given human DNA and asked to identify the specific individual who donated it. The first went much smoother than the second.

Before each hackathon, DNA samples were first prepared to create a DNA library for feeding into the MinION, a step that was done by Zaaijer. Though generating DNA libraries for MinIONs is much simpler than for other sequencers, it is time-consuming, requires a lab setting (and is therefore not mobile yet), and takes some finesse and experience.

With the libraries prepared, the students take over. Using a pipette, they dispense a solution containing the prepared DNA into the MinION's flow cell (which contains 512 channels containing nanopores). Care must be taken to not introduce air bubbles that render the pores inaccessible. Pipetting is tricky, and generally one person on each team learned how to do it and performed the task each time.

As the solution seeps through the flow cell, individual modules transit the nanopores, and software on the PC powering the MinION starts detecting the ion current disruptions. This raw data (in HDF5 format) gets uploaded to the cloud where software analyzes the recorded events to identify the individual bases. Minutes later, students begin seeing preliminary sequencing data on their screens. (All reads—along with new code written—were posted to the class github site.)

Not all 512 channels contain a nanopore that produces reads, but those that do produce individual files for each sequenced read. It's a lot of data in a very short time, both the promise of the MinION and the beginning of the difficulty for the students.



In a classroom at the New York Genome Center, students observe MinION data during second hackathon. Screenshot shows stats on number and length of reads.

Right away, students were faced with the question of how to transfer thousands of individual files from the lab-supplied PCs to their own (mostly Mac) computers where they could carry out their analysis. The sizes of the files precluded using cloud-based products such as Dropbox whose free accounts don't support synchronizing data at such large scale. The file-transfer issue, after some grappling, was finally solved by placing the data in a BitTorrent Sync folder that was then synched to students' computers (maxing out the hard drive in at least one case).

With the sequenced data downloaded, the students head out. Their task is now to compare their reads with existing DNA sequences found online to identify the sample DNA. This they do using existing alignment tools, many free, that compare two or more reads and produce a similarity score.

For the snack hackathon, students all used NBCI BLAST, a tool that makes it easy to run stand-alone searches for similar sequences and to discover, for instance, whether a given read aligns more closely with a template read from a tomato or from a zucchini. The concept is simple, but the difficulty level can ratchet up quickly depending on what two sequences are being compared. Discriminating between two species is one thing; differentiating between two humans who share many of the same traits is something else entirely.

Difficulty level increases in second hackathon

Of the two hackathons, CSI Columbia proved to be much more open-ended. Here the aim was to test whether MinION sequencing could be used to identify a single person. Normally short tandem reads (STRs) are used to identify individuals (the FBI typically uses 13 different STRs for identification purposes), not the long reads returned by nanopore sequencing. As yet, no scientific framework exists on how to identify an individual using the reads generated from the MinION nanopore sequencer. While there are existing alignment tools for comparing two or more human DNA sequences, almost all were developed for traditional sequencing methods.

Choosing an alignment tool took time. With many different ones, it was hard to know where to begin. Even downloading the tools took time, a step that often had to be repeated when students discovered their first tool choice didn't work well.

File formats were another issue and consumed a significant amount of time for the teams. Different tools accept and output different file formats. Many were incompatible; only some were standard.

For CSI Columbia, the difficulty level ratcheted up much more than even Erlich and Zaaijer had imagined. (In fact, CSI Columbia had initially been slated to occur first, ahead of the snack hackathon. However, preparing the DNA libraries for CSI Columbia took longer than planned, necessitating a switch in the order of hackathons.)

Students were not originally given any clues as to the identities of the individuals whose DNA was being sequenced; they were told only to search several online genetic databases for a close match. With students having to spend considerable time finding the right tool and overcoming file incompatibilities, halfway through the assignment Erlich narrowed the scope, naming himself, Craig Venter, James Watson, or someone in the 1000 Genomes Project as the possible suspects. This extra information changed the scope considerably: rather than finding a single individual in a sea of others, the task became to look closely at a few individuals, and rule out others. Even then, only one of the five groups made the correct identification.

The main issue had to do with the number of reads students actually had to work with. Nanopore sequencing is less accurate and has more errors (deletions, insertions, and substitutions) and more noise than traditional sequencing. After filtering out those reads not meeting quality requirements for nanopore sequencing, students were left with a subset of reads covering the genome to around 1%. Such low coverage poses a challenge since much information about ancestry or traits is derived from tiny changes in the DNA (SNPs). Even so, students were able to learn some aspects of an individual's ancestry and traits (including susceptibility to diseases).

(Erlich wants to offer the class again and is considering adding an intermediate, "where-you-are" report so students can help one another over encountered roadblocks.)

Fortunately for the students, the grade depended more on methodology and designing a workable sequencing pipeline than coming up with a correct identification. In this regard, the students excelled, even with the severe computational challenges of constructing an integrated pipeline out of several distinct steps (acquisition, storage, distribution, and analysis), each with its own particular file incompatibilities and data storage problems. Without a clear route already mapped out by others, students responded by writing their own code to plug up the holes and seamlessly transition data from one step to another.

The fundamental structure was sound; it was the data that was lacking. But even then, students demonstrated they were able to properly interpret the data they had. If they couldn't identify the exact donor, they still were able to provide a list of traits that in the real world would help narrow the number of suspects.

Zaaijer points out also that students were dealing with a technology that is not yet mature. "Mobile sequencing is just now getting off the ground, and the error-rate in the reads is still relatively high compared to traditional DNA sequencing—though many scientific groups are working on improving this. It was good for the students to experience that not everything is an iPhone where you open the box and it works. Technology evolves by hard work of many people who see a future (and applications) for new types of devices and machines. The hackathons were a good learning experience. Even though there are obstacles to overcome, the students also saw the opportunities the technology has."

Students not only demonstrated they absorbed the basics of DNA sequencing but added ideas and strategies of their own. One team had taken a throw-processing-power-at-the-problem approach, setting up a dedicated server for the sole purpose of downloading the entire genomes of Watson and Venter—enormous files weighing in at 100 gigs for Watson, 80 gigs for Venter. It ran for over 24 hours before the team called a halt.

Interestingly the one group that did correctly identify its suspect actually had the fewest reads but compensated by using a statistical approach that assigned probabilities to different templates, thus narrowing choices to the most likely candidate. It was an impressive and highly workable solution that Erlich sees as the subject of a possible scientific paper.

Final project

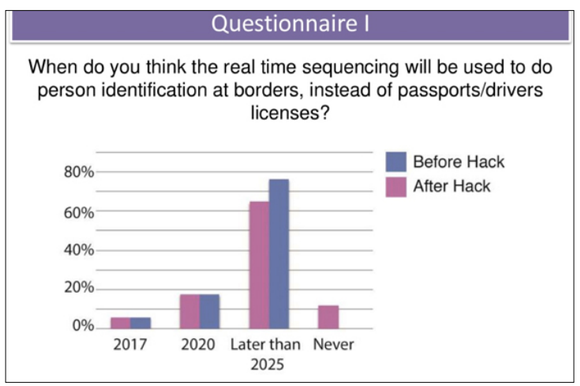
The final project, good for 25% of the grade, had students work in pairs to describe a new use for the MinION. Each group had different applications, from wastewater management, to safe person identification at borders, to sequencing by zero gravity. Especially innovative was the idea for at-home sequencing to trace potential transplant rejection; another proposed using the sequencer when traveling to find edible food and clean water resources.

How soon before these applications or any others start appearing in the real world? Once before and once after the hackathons, students were asked to estimate when mobile DNA sequencing might replace passport checks at national borders. Their answers were more conservative at the second asking, but not by much. Only one or two students revised their answer. Students clearly see the potential for mobile DNA sequencing, even with first-hand knowledge of the work and dedication still needed to optimize the technology.

Though there were hiccups, the problems had more to do with finding the proper tools and overcoming incompatible file formats. Erlich and Zaaijer had been pushing from the beginning to see how far the students could go; that some original assumptions didn't work out was only to be expected. However, the main goal was clearly achieved: students new to DNA sequencing were able —with a little training—to successfully set up a sequencing pipeline and imagine new uses for the MinION. That a sophisticated process once relegated to specialized labs played out relatively smoothly in the classroom points to the huge possibilities of mobile, onsite DNA sequencing.

Says Erlich, "The future is here: we can place DNA sequencers in the hands of our students. No more theoretical explanation of how sequencers work, no more just data wrangling. We can let them feel the internal, promote critical thinking, and a sense of ownership. DNA is everywhere. In your food, on your clothes, everything you touch. By having these sequencers, we can let students get a glimpse for this rich data layer around them."

Linda Crane



Researchers Develop Algorithm to **3D Print Vibrational Sounds**

3D metallophone cups automatically created by computers.

In creating what looks to be a simple children's musical instrument—a xylophone with keys in the shape of zoo animals—computer scientists at Columbia Engineering, Harvard, and MIT have demonstrated that sound can be controlled by 3D-printing shapes. They designed an optimization algorithm and used computational methods and digital fabrication to control acoustic properties—both sound and vibration—by altering the shape of 2D and 3D objects. Their work—"Computational Design of Metallophone Contact Sounds"—will be presented at SIGGRAPH Asia on November 4 in Kobe, Japan.

"Our discovery could lead to a wealth of possibilities that go well beyond musical instruments," says **Changxi Zheng**, assistant professor of computer science at Columbia Engineering, who led the research team. "Our algorithm could lead to ways to build less noisy computer fans, bridges that don't amplify vibrations under stress, and advance the construction of micro-electro-mechanical resonators whose vibration modes are of great importance."

Zheng, who works in the area of dynamic, physics-based computational sound for immersive environments, wanted to see if he could use computation and digital fabrication to actively control the acoustical property, or vibration, of an object. Simulation of contact sounds has long interested the computer graphics community, as has computational fabrication, and, he explains, "We hoped to bridge these two disciplines and explore how much control one can garner over the vibrational frequency spectra of complex geometrics."

Zheng's team decided to focus on simplifying the slow, complicated, manual process of designing idiophones, musical instruments that produce sounds through vibrations in the instrument itself, not through strings or reeds. Because the surface vibration and resulting sounds depend on the idiophone's shape in a complex way, designing the shapes to obtain desired sound characteristics is not straightforward, and their forms have been limited to well-understood designs such as bars that are tuned by careful drilling of dimples on the underside of the instrument.

To demonstrate their new technique, the team settled on building a "zoolophone," a metallophone with playful animal shapes (a metallophone is an idiophone made of tuned metal bars that can be struck to make sound, such as a glockenspiel). Their algorithm optimized and 3D-printed the instrument's keys in the shape of colorful lions, turtles, elephants, giraffes, and more, modeling the geometry to achieve the desired pitch and amplitude of each part.

"Our zoolophone's keys are automatically tuned to play notes on a scale with overtones and frequency of a professionally produced xylophone," says Zheng, whose team spent nearly two years on developing new computational methods while borrowing concepts from computer graphics, acoustic modeling, mechanical engineering, and 3D printing. "By automatically optimizing the shape of 2D and 3D objects through deformation and perforation, we were able to produce such professional sounds that our technique will enable even novices to design metallophones with unique sound and appearance."

Though a fun toy, the zoolophone represents fundamental research into understanding the complex relationships between an object's geometry and its material properties, and the vibrations and sounds it produces when struck. While previous algorithms attempted to optimize either amplitude (loudness) or frequency, the zoolophone required optimizing both simultaneously to fully control its acoustic properties. Creating realistic musical sounds required more work to add in overtones, secondary frequencies higher than the main one that contribute to the timbre associated with notes played on a professionally produced instrument.

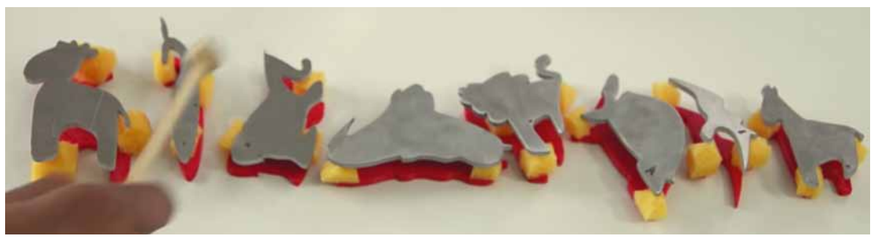
Looking for the most optimal shape that produces the desired sound when struck proved to be the core computational difficulty: the search space for optimizing both amplitude and frequency is immense. To increase the chances of finding the most optimal shape, Zheng and his colleagues developed a new, fast stochastic optimization method, which they called Latin Complement Sampling (LCS). They input shape and user-specified frequency and amplitude spectra (for instance, users can specify which shapes produce which note) and, from that information, optimized the shape of the objects through deformation and perforation to produce the wanted sounds. LCS outperformed all other alternative optimizations and can be used in a variety of other problems.

"Acoustic design of objects today remains slow and expensive," Zheng notes. "We would like to explore computational design algorithms to improve the process for better controlling an object's acoustic properties, whether to achieve desired sound spectra or to reduce undesired noise. This project underscores our first step toward this exciting direction in helping us design objects in a new way."

Zheng, whose previous work in computer graphics includes synthesizing realistic sounds that are automatically synchronized to simulated motions, has already been contacted by researchers interested in applying his approach to micro-electro-mechanical systems (MEMS), in which vibrations filter RF signals.

The work at Columbia Engineering was supported in part by the National Science Foundation (NSF) and Intel, at Harvard and MIT by NSF, Air Force Research Laboratory, and DARPA

Holly Evarts



To demonstrate their optimization algorithm, the researchers built a "zoolophone," a metallophone with playful animal shapes.

In U.S. Senate Testimony, Henning Schulzrinne Offers **Technology Solutions to Unwanted Calls**

Robocalls are proliferating and becoming increasingly sophisticated and deceptive, purporting to be from banks or government agencies to trick and scare people into revealing personal information or transferring money. Recent advances in technology have reduced the cost of calling to close to nothing and made it easier to "spoof," or misrepresent, the originating number or caller ID. The famous Do Not Call list, while effective against unwanted calls from legitimate businesses, is no deterrent to criminals intent on fraud. Seniors are especially vulnerable, and for this reason, the Senate Special Committee on Aging held hearings in June 2015 on possible new legislation to prevent unwanted calls. Among those testifying was Henning Schulzrinne who provided the biggest takeaway of the day: technology offers solutions.

More than 10 years after the Do Not Call list was instituted, more robocall complaints than ever are being received by the Federal Trade Commission (FTC) and the Federal Communications Commission (FCC).

Technological advances are partly to blame. As the telephone infrastructure is changing from traditional copper wires to Voice-over-IP (VoIP) technology, what was once expensive and difficult—international calling, auto-dialing, falsifying caller ID information—has become cheap and easy, making it possible for almost anyone with a laptop and an Internet connection to flood phones with millions of robocalls and to do so from any location in the world.

The nature of the calls themselves has changed also. Before the list, most robocalls were legitimate telemarketers looking to make a sale. Against those calls, the Do Not Call list has been largely effective, leaving the field wide open to illegitimate operators who, like bank robbers walking past the meter on the way into the bank, ignore the Do Not Call list to commit the bigger crime of fraud, either conning victims into divulging personal information or of selling services or products that never materialize.

WHAT YOU CAN DO AGAINST ROBOCALLING

1. **Hang up immediately. Do not press buttons or engage the caller.**
2. Sign up for Nomorobo or other services that blacklist numbers of known robocallers. (Nomorobo is available only in the US and only from certain carriers.) Or sign up for services such as GoogleVoice's free feature that prompts callers to say their names before you pick up.
3. File a complaint with the FTC. Complaints help define patterns of fraud and abuse, sometimes leading to investigations that result in fines.

To increase their odds of success and because VoIP makes it easy, robocallers often impersonate a legitimate bank or government agency. It's called spoofing, and it is quasi-legal. The Caller ID Act of 2009 does make spoofing a crime but only when it is used to harm or defraud someone, something possible to prove only after the fact. No one seems too concerned, and companies openly sell spoofing software. There is even a free iPhone app for spoofing. An app is strictly small scale and for targeting specific individuals; for spoofing at industrial-scale, robocallers are likely to turn to open-source phone switch software when inserting fake phone numbers into millions of calls.

And they usually get away with it. Experiments done by system staff at Columbia University showed that even large carriers do not reject implausible phone numbers such as 311-555-2368.

The ability of robocallers to associate their numbers with any other number or caller ID name gives rise to a whole slew of semi-plausible scams: the IRS demanding payment for overdue taxes, the Social Security Administration requesting an account number to make a deposit, an extradition threat from local police if a debt is not immediately repaid. There are many others, like the one that promises a "free" medical alert system. Most people today know enough to be wary of such calls, but the robocallers' simple business model—flood phones with millions of cheap calls to flush out the few naïve victims that make the business model work—is robust against a low success rate. Even a 95% or 99% suppression rate would not sufficiently discourage robocallers if it leaves the most likely victims unprotected.

Because senior citizens are especially vulnerable to such scams, the Senate Special Committee on Aging in June held hearings on possible legislative solutions. Chaired by Susan Collins (R-Maine), the committee called four witnesses—a small business owner who logged 62 robocalls within a month, an FTC representative who testified about her agency's difficulty in dealing with the problem, and a Missouri Deputy Attorney General whose office last year fielded 57,000 complaints, 52,000 of which concerned unwanted calls.

Testifying about the technology aspects was Henning Schulzrinne, who developed the key protocols that enable VoIP and who continues to work on VoIP protocols as a professor of computer science at Columbia University. He is also knowledgeable about the policy issues, having served as the Chief Technologist at the FCC from 2012 to 2014. While currently consulting for the agency, it was in his private role as a technology expert that he addressed the committee.

After summarizing eight categories of scams, Schulzrinne described the technology solutions, which fall into roughly three categories: filtering, caller ID and name authentication, and gateway blocking. Each, summarized below, has its strong points and limitations.

Filtering

Filtering, either through a third-party service or a downloaded app, works by checking each incoming call against a white list of trustworthy phone numbers or a

Testing showed that clearly fictitious numbers were transmitted even though it would be easy for phone carriers to identify and block them.

black list of nonacceptable ones compiled in one of several ways: from FTC and FCC customer complaints, crowd-sourced by consumers, or collected through honeypots. (Honeypots are stealth servers programmed to act like normal phones—with numbers not assigned to any individual or company—for the express purpose of capturing the phone numbers of robocallers.) Built-in safeguards can ensure emergency alert calls get through as do calls placed from medical facilities; unknown phone numbers can be verified by making callers prove that they are human rather than robotic.

Filtering today has several drawbacks. It puts the onus on individuals, and it protects only those who know about filtering and are willing to do the setup, generally the most sophisticated people who are unlikely to fall for a scam in any case. By protecting the people who least need it, filtering today leaves the most vulnerable even more exposed.

Extending filtering to others is not currently easy. Filtering works on many landlines, and it is usually available only through large cable companies like Time Warner or Comcast that support external filtering services such as Nomorobo.

And filters are easily avoided by robocallers' use of spoofing.

Caller ID and name authentication

Spoofing is perhaps the most nefarious aspect of the scamming schemes; almost anyone is likely to pick up when seeing the phone number of the local police department or the IRS. Spoofing has other bad uses as well since a caller ID is often used to verify one's identity when gaining access to voicemail or when calling a bank, utility, or airline.

Preventing spoofing is necessary both to make filtering effective and to stop robocallers from impersonating others, and Schulzrinne offered possible ways to do it. One is to authenticate the originating number to ensure the caller is authorized to use the caller ID contained in the call setup message. Authentication would require phone carriers to insert links to new cryptographic certificates so any carrier along the way could validate the signature and detect spoofed caller IDs. These calls could then be labeled in some way or, if the customer prefers, rejected.

However, it's not clear how much the phone carriers will do voluntarily. For years, carriers have resisted appeals to block robocalls, claiming that federal law prohibits them as common carriers from doing so. The FCC pulled the rug out from this excuse in a June 18 vote that explicitly states that phone companies are legally allowed to provide filtering to those customers who request it. (The FCC does not currently, however, obligate phone companies to provide filtering.)

Using his deep knowledge of the protocols, Schulzrinne offered an alternative approach to preventing spoofing, one that does not rely on carriers. The VoIP protocols (specifically the Session Initiation Protocol, or SIP) allow for changing the mechanics how caller ID information is generated, and thus make it difficult to do spoofing in the first place.

Currently ID information is collected from many different databases and is often not validated, making it easy for fraudulent callers to insert any information they like, especially for numbers that have not been assigned to a carrier. Because SIP allows the calling carrier to insert name information directly into the call signaling request, it's possible to avoid looking up the information in databases and making it easier to track who generated the information. Longer term, carriers may also indicate that they have validated the information by cross-checking them against service address records or credit card billing information, for example.

Blocking at the VoIP gateway ("do not originate")

Perhaps Schulzrinne's most innovative proposal is a do-not-originate list that would cut off robocalls closer to the source: at the VoIP gateways that connect VoIP calls to the traditional phone system. While VoIP robocalls can be placed from anywhere in the world, all such calls pass through such gateways to enter the traditional circuit-switched phone lines used by most large US companies and large carriers. (Companies generally contract with a carrier that operates a VoIP gateway on their behalf to handle the transition for all incoming and outgoing calls.)

VoIP gateways currently do not check whether the originating number is valid or not. However, it would be easy to program them to reject originating phone numbers of companies that did not contract for their services or numbers known to be out of service. Any calls from numbers on a list to not originate—a reverse do-not-call list—would be rejected by the gateway and thus blocked from entering the phone system. Alternatively, the gateway could replace the fake caller ID information with a fraud indicator, such as the (made-up) area code 666. Consumer-chosen call filtering technologies can then reject those calls if the carrier prefers not to. While companies would have to list themselves on do-not-originate lists, those companies most likely to be impersonated would have incentive to do so.

The do-not-originate approach has the advantage that it can be implemented quickly and easily, without any changes in telephony protocols. Nor does it require cooperation of other phone carriers. It is no substitute for authentication, but it should prevent many of the most harmful calls from reaching consumers.

Breaking the business model

Each of the three methods—filtering, authentication, VoIP gateway blocking—does its part to add to the difficulty and expense of robocalling, but each addresses only a subpart of the problem. The do-not-originate list addresses spoofing of high-profile numbers of government agencies and banks but not other legitimate-sounding numbers robocallers invent ("Card Svcs," "Medcare"). Authentication stops robocallers from impersonating legitimate businesses and government agencies (and makes fraudulent calls less likely to pay off) but does nothing to prevent robocalls themselves. Filtering can stop robocalls but currently protects the relatively few individuals who use it and is easily circumvented by spoofing.

But used in combination with one another, the three methods complement one another to undermine the economics of robocalling. Once authentication is in place to prevent spoofing and people can trust that phone numbers are legitimate, white lists of acceptable numbers—government agencies, banks, doctors—can be compiled and safely and widely distributed to protect even the most vulnerable. And without spoofing to disguise their calls, robocallers quickly get identified and blacklisted (and in the best case, shut down by law enforcement).

It's the combination of methods, working in conjunction with the VoIP technology and the supporting protocols, that stands the best chance of approaching the 100% suppression rate needed to put an end to robocalling. Since it was technology that allowed robocalling in the first place, it's only fitting that technology be part of the solution.

Linda Crane

The full transcript of Schulzrinne's testimony is at aging.senate.gov/imo/media/doc/Schulzrinne_6_10_15.pdf

Author Interview: **Steven Bellovin on *Thinking Security***

In the face of relentless security attacks, is it possible to keep systems, data, and networks protected? Yes, says respected security expert Steven Bellovin, but it requires more than a static checklist of standard security measures. It requires looking ahead of current technology to anticipate vulnerabilities and understand how and why they exist; only then is it possible to identify the most effective defense mechanism and guard against new attacks. To help security specialists and other IT professionals foster a security mindset, Bellovin in his latest book, Thinking Security, ***describes fundamental security principles that are true no matter the computing environment or how much technology changes. It's a pragmatic approach that presents security as a systems issue while considering cost, the value of the assets being protected, the actual threat, and employees' need to be productive.***

Why did you write Thinking Security*?*

Dissatisfaction with how security is practiced in the real world. Security today tends to rely on checklists based on yesterday's technology and yesterday's threats. Checklists can't cover every situation, and they can't anticipate new types of attacks.

After years of seeing misleading and simplistic security recommendations in the mainstream press, I started thinking about underlying principles and what security advice is always going to be the same no matter what happens in technology; it's the way I try to get my students to think about security. All those things together went into the book.

What is an example of misleading security advice?

That a strong password will protect you. The rules on picking strong passwords go back to a paper in 1979, so this is not new technology and there are ways to bypass strong passwords. Keystroke loggers and phishing attacks, for example, don't care how strong your password is.

The underlying vulnerability here is the reuse that occurs when you're sending something to one site that can be stolen from that site and reused against you. In RSA SecurID—generally considered very secure—a cryptographic secret is embedded in the token but a server somewhere has a copy of that secret. Anyone hacking into that server can impersonate the file of tokens kept there. Some will say, *Lock down the server*. If you can lock down the server, why can't you lock down your password file? Why is that server more secure than a password file? It's not.

A one-time password if done right is secure, not because it is hard to guess, but because it can't be reused.

Your most valuable password is your email account password, because that's used for all the password resets. Anyone having your email password can potentially learn any password emailed to you, no matter how strong the password.

Passwords are ubiquitous. Can they be used safely?

I use and recommend password managers, though there are some bad designs out there. The book discusses the characteristics that make for a good password manager.

***Thinking Security* is written for network and security administrators, but some security advice applies to everybody.**

- **Use a password manager to securely store a different credential for every site and avoid reuse of keys.**

  Look for password managers that encrypt URLs and that add "salt" (a random string of data) to each password to add an extra layer of protection.

  Though web access to a password collection is convenient, it is also more dangerous, especially when using potentially insecure machines.

  One nice feature is the ability to copy a password to the clipboard for easy pasting into web forms; however, check that the clipboard gets automatically cleared.

  The more integrated the manager is with a browser, the more risk there is that malware can abuse it to steal your credentials.

- **If your bank offers an online access to your account, use it.** By regularly logging in, you'll detect fraudulent activity more quickly.

- **Use a credit card rather than a debit card when making purchases, especially when you don't completely trust a site.** US law limits cardholders to $50 liability in the case of unauthorized card use. (For debit cards, which are covered under a different law, you're liable for up to $50 if you report within two days; after two days, you're liable for up to $500. After 60 days, you're liable for the entire stolen amount.)

  Debit cards have the added risk of being a direct line into your bank account.

If checklists are too static to be useful, how should people go about ensuring their systems are secure?

It starts with two fundamental questions: what are you protecting, and against whom are you protecting it.

If you don't answer those questions, you're doing security just to do security, forgetting that the purpose of security is not to increase security, but to prevent loss.

Defenses have to be matched to the likely attacks. If you're protecting a single database accessed only by employees, a firewall will probably suffice. However, if it's 17 databases all tied together and made to function as a single resource while also needing to be accessible by those inside and outside the company, a firewall is not going to work.

Nor will a firewall protect you from an attack launched from inside. Even employees might work against a firewall if it prevents them from getting their work done.

People might be surprised to see you say firewalls may not provide needed security. You and William Cheswick wrote the first book on firewalls (Firewalls and Internet Security: Repelling the Wily Hacker).

That book was written in 1994. Networks and systems are more complicated and interconnected these days. From a security perspective, complexity is fatal.

Firewalls work well when there is a clear distinction between the inside and outside of what you're protecting. Today it's not always clear-cut. Companies often make their databases and parts of their network accessible to outside contractors, vendors, or auditors. In such cases, a firewall is not appropriate, but not

having a firewall creates vulnerabilities that can be exploited.

Which is what happened in the breach at Target. Attackers obtained the network credentials used by Target's HVAC vendor, which had external access to Target's network. Once inside, hackers were able to move freely over Target's network, which from all accounts was rather loosely structured with little segmentation. Internal firewalls should have been used to cordon off sensitive parts of the network, like the payment system, which is how the attackers were ultimately able to steal credit card information.

Legacy systems are a problem; an internal network might have started off simple but just then grew, with security put in this one spot here and another spot over there, but never with any overarching vision of how it should be done; before too long it's too late to do a coordinated plan.

The Target breach was one of many big ones in the last couple of years. Federal agencies were attacked multiple times, Home Depot, Anthem Health, even Chase.

Attackers seem to be always one step ahead...

...mostly, but not completely. We don't hear about the attacks that get repulsed.

With all the risks, would you recommend people not use online banking?

No. And I'll tell you why. As a matter of practice, banks don't hold customers liable for money hacked from their bank accounts because the next bank down the street won't. It's the competitive landscape.

Merchants and businesses, however, are generally liable. A customer is not. There are a lot of things I will do if I'm not liable.

Many things come down to economics, and security is one of them.

That is one of the main themes in the book.

Yes, security costs money. Companies have to spend resources, understand the need, and they have to be willing to accept inconvenience in order to protect themselves. If it takes two signatures to fully protect something, do the extra bit of work to get two signatures.

But companies are under other pressures. Online vendors need to make their sites easy to use for customers.

Amazon for instance generally does not make you go through the extra step of inputting the 3- or 4-digit card security code that other sites require because Amazon has made one-click ordering a business priority. Less secure verification will incur some loss, but Amazon is willing to eat those losses, figuring net profit is greater than the loss if it's easy for people to buy.

Insecurity is not a state of sin; it's part of running a business and business can be risky.

Are you optimistic people can secure sites and data?

Yes and no. The biggest cause of security problems is buggy code. This is not a new thought of mine. It was true 20 years ago, and though code is better written today, programs are bigger and more complex. It's hard to imagine what a defense against buggy code would look like.

Any system must also be periodically re-evaluated for vulnerabilities, something that rarely happens.

Research into new security measures is ongoing. When I came to Columbia ten years ago, sandboxing was known to have good properties, but it was not then in general use. Today it's a mainstream part of all operating systems.

Digital rights management has also been more successful at protecting proprietary content than I thought it would be. It works because most of the content that people were pirating can now be bought at reasonable prices online. From a technological perspective, digital rights management doesn't seem to be something that should work, but it works from an economic perspective. Not perfectly, of course, but good enough.

I'm morally certain that right now someone in Silicon Valley or Tel Aviv or Hyderabad or Beijing or Accra or somewhere is devising something that 10 years from now, we'll find indispensable, but will have as profound an effect on security as today's smartphones have had on communications and society. We just don't know what it is yet.

Right Time, Right Place: A Collaborative Approach for More Accurate Context-awareness in Mobile Apps and Ads

The right information delivered at the right time can make apps and ads more appealing and relevant to customers: a traffic app that auto-updates for the work or home commute as appropriate; a restaurant that offers lunch coupons for people who work in the area but dinner coupons for people who live nearby. This level of customization requires taking into account a user's immediate context, something that is not easy to do. It requires both location data and a temporal framework that gives meaning to each location, identifying it as home, work, commute, or another place frequented by a user. But location data is often surprisingly sparse for any one user. To overcome sparsity and construct reliable weekly routines for individual users, researchers integrated global temporal patterns inferred from the entire data set with user-specific spatiotemporal data. The resulting method is entirely data-driven—requiring no labeling—and flexible to accommodate variations in a user's weekly schedule.

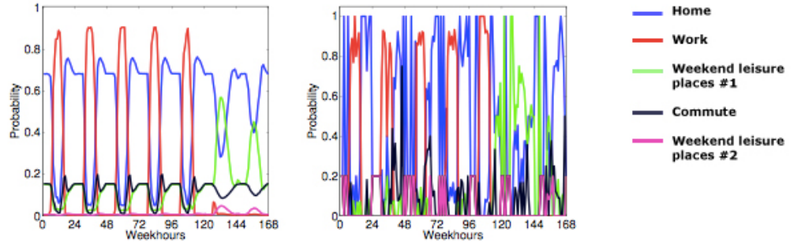
The location data needed for context-aware ads and apps is surprisingly sparse for any single user. For privacy and to conserve energy, most smart phone apps log users' locations only when the app is active. The result is that location data sets collected from apps comprise many users but few observations per user. This sparsity makes it difficult to know how a particular GPS position is relevant to a user, whether it represents a work place, home, a point along the morning or evening commute, or some other frequently visited destination. It's this contextual information that allows companies to customize their apps and ads for their customers' immediate or near-future locations. Local businesses especially benefit when they can accurately predict who is or will soon be in their area.

(It's not just companies that want context-aware apps and ads; there is evidence users do, too. Cisco found that half of customers surveyed would use coupons sent from a nearby store.)

Location data for any single user may be too sparse to understand when a user transitions between places, but the collection of data across all users represents much more information that can help illuminate broader patterns. To exploit this collective information, four researchers—Berk Kapicioglu, David S. Rosenberg, Robert Schapire, and **Tony Jebara**—developed a data-driven method that learns people's important places based on global temporal patterns inferred from the entire data set. They described this method in the paper *Collaborative Place Models* presented in July 2015 at the International Joint Conference on Artificial Intelligence.

Collaborative place models differ from previous methods that first label locations according to time of day and day of week. By assuming, for example, a 9-to-5 workday Monday through Friday, methods that rely on labeling might average positions between 8AM and 6PM and call that home while averaging positions between 9 and 5 and calling that work. It's an intuitive approach but it lacks flexibility—not everyone has the same schedule—and it ignores the commute, which can be a significant amount of time for some people and a missed opportunity for those businesses located along the commute.

Rather than imposing a static temporal framework, collaborative place models learn the quantitative relationship between week-hours by inferring similarities across all users, relying on Bayesian estimation techniques to do so. With a global temporal framework thus set, the relevance of the sparser latitude-longitude GPS coordinates from individual users can then be determined from how they fit into the global temporal pattern. In this way, the model re-constructs a particular user's home-work-commuting schedule even though a user might have been observed only at Thursday 3PM and Monday 1PM.



The strong, well-defined pattern on the left results from combining global weekly patterns with spatiotemporal data of an individual user arbitrarily chosen from the dense dataset. The right distribution (for the same user) represents a previous baseline model that did not infer global patterns and so as not able to correctly identify important places.

To prove the concept, the researchers tested the model using two real-world data sets, a sparse one collected from a mobile ad exchange, and a dense data set from a cellular carrier. In both cases, the only inputs were user IDs, latitudes, longitudes, and time stamps. (Data was anonymized by removing all personal information.)

With data aggregated across all users, a strong, global temporal pattern emerged fairly quickly, one that contained within it several temporal clusters correlated with work, morning and evening commutes, leisure times after work, and sleeping at night. With the global pattern thus established, the individual spatiotemporal patterns of individual users became apparent even with few data points associated with each user.

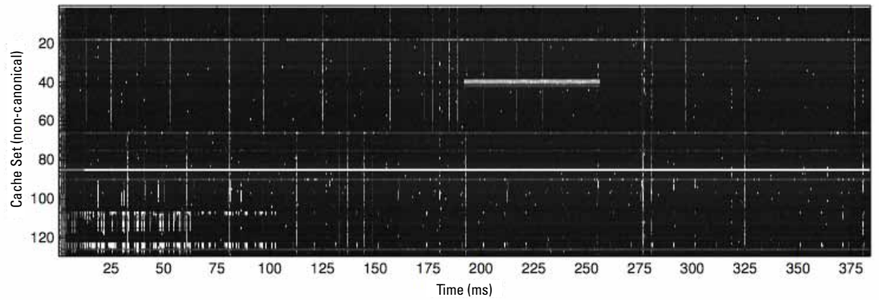
The spatial extent of place types associated with temporal clusters were determined by replacing multiple observations logged during the same hour with their geometric median (computed using Weiszfeld's algorithm and by clustering nearby points using a Gaussian mixture model that is a subcomponent of the collaborative place model). This contrasts with the use of averaging in other place models to handle redundant observations and the noise that occurs from GPS errors and from having multiple cell towers covering the same location; by not averaging, the collaborative place model avoids the strange results sometimes caused by deviations in the regular routine, such as a late work evening or a night or weekend away from home.

Flexibility was built into the model by allowing users to have varying numbers of places or week hours. This flexibility turned out to be key; an early, simpler prototype that constrained users to have the same week-hour distribution performed worse than a baseline model.

In the end, data by itself was enough to reliably assess a user's spatiotemporal schedule. Without the need to label or average location places, the collaborative approach of combining global patterns with sparse user location data reduced the median distance error by 8% from a simpler non-collaborative baseline model.

Linda Crane

Side-channel Attacks in Web Browsers:
Practical, Low-cost, and Highly Scalable



A memorygram of L3 cache activity: Vertical line segments indicate multiple adjacent cache sets are active during the same time period. Since consecutive cache sets (within the same page frame) correspond to consecutive addresses in physical memory, it may indicate the execution of a function call spanning more than 64 bytes of assembler instructions. The white horizontal line indicates a variable constantly being accessed during measurements, and probably belongs to the measurement code or to the underlying JavaScript runtime.

In a paper presented at the ACM Conference on Computer and Communications Security, four computer scientists from Columbia University—Yossi Oren, Vasileios P. Kemerlis, Simha Sethumadhavan, and Angelos D. Keromytis—demonstrate that it's possible to spy on activities of a computer user from a web browser, even in some cases determining what website(s) a user is visiting. This type of attack, dubbed spy-in-the-sandbox, works by observing activity in the CPU cache on Intel microprocessors. It affects close to 80% of PCs, and it represents an escalation and scaling up of what's possible with side-channel attacks, requiring no special software or close proximity to the victim. Fortunately the fix is easy and Web browser vendors, alerted to the problem, are updating their code bases to prevent such attacks. One other upside: the spy-in-the-sandbox attack may serve as a primitive for secure communications.

In a side-channel attack, an attacker is able to glean crucial information by analyzing physical emissions (power, radiation, heat, vibrations) produced during an otherwise secure computation. Side-channel attacks are not new; Cold-War examples abound, from aiming a laser beam at a window to pick up vibrations from conversations inside, or installing microphones in typewriters to identify letters being typed. On computers, side-channel-attacks often work by inferring information from how much time or battery power is required to process an input or execute an operation. Given precise side-channel measurements, an attacker can work backward to reconstruct the input.

Side channel attacks can be particularly insidious because they circumvent security mechanisms. Traditionally they are directed against targeted individuals and assume proximity and special software installed on the victim's computer. However, those assumptions may have to be rethought after four computer scientists from Columbia University (Yossef Oren, Vasileios P. Kemerlis, Simha Sethumadhavan, and Angelos D. Keromytis) demonstrated for the first time that it is possible to launch a side channel attack from within a web browser. The method is detailed in their paper *The Spy in the Sandbox—Practical Cache Attacks in JavaScript and Their Implications*, which was presented October 12, 2015 at the ACM Conference on Computer and Communications Security.

The attack, dubbed spy-in-the-sandbox by the researchers, does not steal passwords or extract encryption keys. Instead, it shows that the privacy of computer users can be compromised from code running inside the highly restricted (sandboxed) environment of a web browser. The researchers were able to tell for instance whether a user was sitting at the computer and hitting keys or moving the mouse; more worrisome from a privacy perspective, the researchers could determine with 80% accuracy whether the victim was visiting certain websites.

More may be possible. As Yossef Oren, a postdoctoral researcher who worked on the project (now an Assistant Professor at the Department of Information Systems Engineering in Ben-Gurion University) puts it, "Attacks always become worse."

In one sense at least, spy-in-the-sandbox attacks are more dangerous than other side-channel attacks because they can scale up to attack 1,000, 10,000, or even a million users at once. Nor are only a few users vulnerable; the attack works against users running an HTML5-capable browser on a PC with an Intel CPU based on the Sandy Bridge, Ivy Bridge, Haswell, or Broadwell micro-architectures, which account for approximately 80% of PCs sold after 2011.

How it was done

Neither proximity or special software is required; the one assumption is that the victim can be lured to a website controlled by the attacker and leaves open the browser window.

What's running in that open browser window is JavaScript code capable of viewing and recording the flow of data in and out of the computer's cache, specifically the L3, or last-level, cache. (A cache is extra-fast memory close to the CPU to hold data currently in use; caching data saves the time it would take to fetch data from regular memory.)

That an attacker can launch a side-channel attack from a web browser is somewhat surprising. Websites running on a computer operate within a tightly contained environment (the sandbox) that restricts what the website's JavaScript can do.

However, the sandbox does not prevent JavaScript running in an open browser window from observing activity in the L3 cache, where websites interact with other processes running on the computer, even those processes protected by higher-level security mechanisms like virtual memory, privilege rings, hypervisors, and sandboxing.

The attack is possible because memory location information leaks out by timing cache events. If a needed element is not in the cache (a cache miss event), for instance, it takes longer to retrieve the data element. This allows the researchers to know what data is currently being used by the computer. To add a new data element to the cache, the CPU will need to evict data elements to make room. The data element is evicted not only from the L3 cache but from lower-level caches as well. To check whether data residing at a certain physical address are present in the L3 cache as well, the CPU calculates which part of the cache (cache set) is responsible for the address, then only checks the certain lines within the cache that correspond to this set, allowing the researchers to associate cache lines with physical memory.

In timing events, researchers were able to infer which instruction sets are active and which are not, and what areas in memory are active when data is being fetched. "It's remarkable that such a wealth of information about the system is available to an unprivileged webpage," says Oren.

"While previous studies have been able to see some of the same behavior, they relied on specially written software that had to be installed on the victim's machine. What's remarkable here is that we see some of the same information using only a browser," says Vasileios Kemerlis, a PhD student who worked on the project (now an Assistant Professor in the Computer Science Department at Brown University).

By selecting a group of cache sets and repeatedly measuring their access latencies over time, the researchers were able to construct a very detailed picture, or memorygram, of the real-time activity of the cache.

Such a detailed picture is possible only because many web browsers recently upgraded the precision of their timers, making it possible to time events with microsecond precision. If memorygrams were fuzzier and less detailed, it would not be possible to capture such small events as a cache miss. (Different browsers implement this new feature with different precisions.) High-resolution timers have recently been added to browsers as a way to give developers, especially game developers, sufficient fine-grained detail to know what processes might be slowing performance. Of course, the more information developers have, the more information an attacker can access also.

Different processes have different memorygrams, and the same is true for different websites; their memorygrams will differ depending on the data the site is using, how the site is structured, how many images it contains and the size of those images. These various parts of the website end up in different locations in memory, and need to be called and cached, giving each website its own distinctive signature.

The researchers visited 10 sites and recorded multiple memorygrams in each case to build a classifier that could, with 80% accuracy, determine if a website open on a victim's machine matched one of the 10 pre-selected sites. (The same website viewed on different browsers will exhibit slight differences; it's this noise that prevents 100% accuracy when matching memorygrams.)

Future work

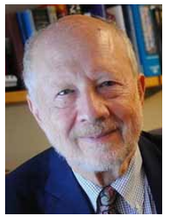
As pernicious as is the side-channel attack, especially considering how practical, scalable, and low-cost it is, avoiding it is surprisingly easy: run only a single web browser window at a time. An across-the-board fix to prevent the attacks is easy also; have browsers return to using less precise timers (or alert users of the high-precision timers that there exist possible security vulnerabilities).

And in this story of data privacy at least there is a happy ending. In March 2015, the researchers shared their research with all major browser vendors; by September 2015, Apple, Google, and Mozilla had released updated versions of their browsers to close the identified security hole.

The researchers are not yet done examining the potential of web-based side-channel attacks. They will continue looking at the problem (on old versions of browsers) to test the attack at larger scale. They are also considering a more interesting question; can memorygrams be used for good purposes?

A pre-set memorygrams might be placed in memory to be viewed by a trusted party to convey information. One memorygram might represent a 1 bit, another a 0 bit. The process of communicating in this fashion would be slow, but it would be extremely difficult for an attacker to even figure out that explicit communication is occurring between two parties. Memorygrams might thus serve as a primitive in securely conveying information, and what was once a threat to security may serve to enhance it.

Linda Crane



Joseph F. Traub

"Let me tell you how I got hooked on computing. For my thesis I worked for six months starting from a mathematical model of the helium atom and writing a program to compute the energy and other parameters of the atom. I took the cards from the IBM 650 and loaded them on the printer. The printer started spewing out approximations to the ground state energy of helium. I was using a variational principle which means I was converging down to the ground state energy of the helium matter. Watching, after the six months of work, the numbers rolling off the printer, and seeing that the initial numbers approximated the experimentally measured ground state energy of the helium atom good to four places. That was the moment."

Joseph F. Traub, a pioneering computer scientist and founder of the Computer Science department at Columbia University, died Monday, August 24, 2015 in Santa Fe, NM. He was 83. Most recently the Edwin Howard Armstrong Professor of Computer Science, Traub was an early pioneer in computer science years before such a discipline existed, and he would do a lot to shape the field.

Traub was most known for his work on optimal algorithms and computational complexity applied to continuous scientific problems. In collaboration with Henryk Wozniakowski, he created the field of information-based complexity, where the goal is to understand the cost of solving problems when information is partial, contaminated, or priced. Applications for information-based complexity are diverse and include differential and integral equations, continuous optimization, path integrals, high-dimensional integration and approximation, and low-discrepancy sequences.

Understanding the role of information about a problem was a unifying theme of Traub's contributions to a number of diverse areas of computing. Often collaborating with others, he created significant new algorithms, including the Jenkins-Traub algorithm for polynomial zeros, the Kung-Traub algorithm for comparing the expansion of an algebraic function, and the Shaw-Traub algorithm to increase computational speed. He authored or edited ten monographs and some 120 papers in computer science, mathematics, physics, computational finance, and quantum computing.

Apart from his scientific research, he had a major role in building and leading organizations that promoted computer science. In 1971, at the age of 38, he was appointed chair of the computer science department at Carnegie Mellon University (CMU), overseeing its expansion from fewer than 10 professors to 50, and making it one of the strongest computer science departments in the country. Based on his achievements at CMU, Columbia University in 1979 extended an offer to Traub to found the University's Computer Science department. He accepted the offer and chose to locate Computer Science within the Engineering School, which at the time offered a single computer, only three tenured faculty members teaching computer science, and a huge demand for computer classes.

After securing a $600,000 gift from IBM (which later provided another $4 million), he was able to add faculty and attract top students. Within a year the department was awarding bachelor's and master's degrees as well as PhDs. He would chair the department until 1989.

In 1982 he oversaw the construction of the Computer Science Building, working closely with architects to come up with a final design that would later win awards.

Traub liked building things from scratch. In 1985 while still chair of the Computer Science department, he became the founding editor-in-chief of the *Journal of Complexity* (a position he held at the time of his death). In 1986, he founded the Computer Science and Technology Board (CSTB) of the National Research Council, serving as its chair from 1986 until 1992 and again in 2005 and 2009.

His awards and honors are many and include election to the National Academy of Engineering in 1985, the 1991 Emanuel R. Piore Gold Medal from IEEE, and the 1992 Distinguished Service Award from the Computer Research Association (CRA). He is a Fellow of the Association for Computing Machinery (ACM), the American Association for the Advancement of Science (AAAS), the Society for Industrial and Applied Mathematics (SIAM), and the New York Academy of Sciences (NYAS). He was selected by the Accademia Nazionale dei Lincei in Rome to present the 1993 Lezione Lincee, a cycle of six lectures. Traub received the 1999 Mayor's Award for Excellence in Science and Technology, an award presented by Mayor Rudy Giuliani.

In 2012, his 80th birthday was commemorated by a symposium at Columbia's Davis Auditorium to celebrate his research and contributions to computer science.

Traub's "contributions to Columbia's Computer Science Department have been instrumental in establishing the strong foundation of excellence of our Computer Science department today, enabling our ongoing frontier leadership in this field," said Dean Mary C. Boyce. "Joe will be sorely missed by all of us at Columbia and by the computer science community across the globe."

A life of science and discovery

Traub always described himself as lucky: Lucky in his early life that his parents were able to flee Nazi Germany in 1939 and settle in New York City; that he had a knack for math and problem-solving just when those skills were needed; that a fellow student's prescient suggestion led him to visit IBM's Watson Laboratories where he first encountered computers. And lucky to be among the first to enter a new, unexplored field when he had the ambition to make new discoveries and a hunger to do something significant. In an interview recalling his life, he once said "I'm almost moved to tears but who could have expected such a wonderful life and such a wonderful career."

That he returned to New York City to found Columbia's computer science department is entirely appropriate. He attended both Bronx High School of Science and City College of New York (earning degrees in math and physics) before entering Columbia University in 1954 intent on a PhD in theoretical physics. That plan changed when he discovered computers, not at Columbia—which had no computers—but at the IBM Watson lab then located in Casa Hispanica, just off campus at 612 W. 116th Street. He was hired there as a fellow, gaining the perk of unlimited computer time.

In 1959 he earned his PhD under the Committee of Applied Mathematics at Columbia. After his first choice to work on a chess problem was rejected, he proposed instead a quantum problem that involved six months of programing to calculate the ground energy state of a helium atom, correct to four decimal points.

After graduating Columbia, Traub went to work at Bell Labs then in its "golden 60s" when researchers were given wide latitude to choose projects and conduct pure research. It was there that a colleague one day walked into his office with a problem. Could Traub find the zero of a function that involved an integral? Mulling over the problem led to two observations: one, it was expensive to compute the function; and two, there were lots of ways of solving it. His thinking about how to select the best, most optimal algorithm culminated in his 1964 monograph ***Iterative Methods for the Solution of Equations***. It was the start of his career with many publications to come.

His luck extended to his personal life. He was married to Pamela McCorduck, a noted author who also taught science writing at Columbia. He enjoyed skiing, tennis, hiking, travel, and good food.

He regularly spent his summers in Santa Fe, where he was an External Professor at the Santa Fe Institute and played a variety of roles over the years, often organizing workshops to bring together those working in science and math. It was in Santa Fe where he died Monday morning, unexpectedly and quickly, after having made plans to travel to Germany, Poland, and CMU. He is survived by his wife Pamela and two daughters, Claudia Traub-Cooper and Hillary Spector.

Joseph Traub was an important and valued member of the Computer Science department he founded. He will be missed by faculty, staff, and students.

In Memoriam
David S. Johnson

David S. Johnson, a leading expert in the area of computational complexity and the design and analysis of algorithms, died Tuesday, March 8, 2016. Since 2014, Johnson was a visiting professor at Columbia University.

The winner of the 2010 Knuth Prize for his contributions to theoretical and experimental analysis of algorithms, Johnson helped lay the foundation for algorithms used to address optimization problems, in which a best solution is sought among a large set of possible solutions to a problem. His papers on the experimental analysis of approximation algorithms were influential in establishing rigorous standards for algorithms that find an approximately optimal rather than exactly optimal solution. Such approximation algorithms play an important role within computer science both in theory and in practice.

Johnson researched and contributed to a range of foundational topics in both mathematics and computer science, including combinatorial optimization, network design, routing and scheduling, facility location, bin packing, graph coloring, and the Traveling Salesman Problem.

It is however his pioneering work on NP-completeness for which he is best known. He was one of the first to investigate NP-completeness, which deals with problems that are believed to be unsolvable within a reasonable amount of time in the worst case. His book, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, co-authored with Michael Garey and written in 1979, has been called a classic for its rigorous treatment of NP completeness and for its clear, concise exposition. The book is one of the most cited references in all of computer science, with over 55,000 citations. Johnson continued to write on NP-completeness throughout his career, maintaining a column on the subject from 1982 until 1992 in the *Journal of Algorithms*.

Born December 9, 1945, Johnson attended Amherst as an undergraduate studying mathematics and went on to MIT where he earned a PhD in mathematics in 1973 for his thesis Near-Optimal Bin Packing Algorithms. The same year, he started his long and productive career at Bell Labs (and later AT&T Research) that would last until 2014. During this time, he published continuously, including several books and well over 100 papers and articles, many of which concern the best ways to cope with computational intractability and his developing interest in the interplay between theoretical and experimental analysis in computer science.

Johnson was an active member and leader in the theoretical computer science community, founding the Symposium on Discrete Algorithms (SODA), a conference that has become a top theory venue; for 25 years he served as SODA's committee chair. He created also the DIMACS (Center for Discrete Mathematics and Theoretical Computer Science) Implementation Challenges. His work within the community was unflagging. He served on the ACM Council as Member-at-Large (1996-2004), chaired ACM SIGACT (1987-1991), edited the Journal of the Association for Computing Machinery (1983-1987), and he served as associate editor of ACM Transactions on Algorithms (TALG) since its founding in 2004.

In 2014, Johnson joined Columbia's computer science faculty as a visiting professor, teaching CS students and interacting with faculty.

"We will miss David very much. He was a wonderful colleague and mentor for students," said Julia Hirschberg, chair of Columbia's Computer Science department.

In addition to the Knuth Prize awarded to him in 2010, Johnson is a 1995 Fellow of the Association for Computing Machinery and was just this year elected to the National Academy of Engineering. Johnson has an Erdös number of 2.

These awards do not do justice to his many contributions to the field of computer science, both written and in private consultation with colleagues and students. David Johnson will be missed for his expertise and for the modest and unassuming way in which he set about to better understand and communicate to others the foundational topics in computer science.

Li-Yang Tan

Prof. **Rocco Servedio**

Professor **Rocco Servedio** of Computer Science, and his former student **Li-Yang Tan** (PhD '14) are recipients of a four-year, $1.2M National Science Foundation (NSF) Award for their proposal to use random projections to prove lower bounds on Boolean circuits. The award will allow Servedio and Tan, now on the faculty of Toyota Technological Institute at Chicago, to continue work they started last year in their paper "An average-case depth hierarchy theorem for Boolean circuits." Named Best Paper at the FOCS 2015 conference, it resolved a conjecture that had been open for close to 30 years.

Professor **Tony Jebara** of Computer Science, with Lamont-Doherty Earth Observatory scientists Joaquim Goes, Ryan Abernethey, and Helga Gomes, win the University's Research Initiatives in Science & Engineering (RISE) competition, for their project titled "Inferring Spatial Heterogeneity in Marine Phytoplankton Using Fluid Dynamics and Bayesian Machine Learning Techniques." From 53 teams that entered this year's competition, only six were chosen to receive funding for each project for up to two years.

Professor **Junfeng Yang** of Columbia University and Yinzhi Cao of Lehigh University are recipients of a four-year, $1.2M National Science Foundation (NSF) grant for their proposal to develop a new approach towards making systems forget data, or the concept they called "machine unlearning." The success of their approach was demonstrated in their paper "Towards Making Systems Forget with Machine Unlearning" that appeared earlier in the 2015 IEEE Symposium on Security and Privacy.

Professor **Luca Carloni** of Computer Science is a guest editor of a special issue of Proceedings of the IEEE that focused on the evolution of Electronic Design Automation (EDA) and its future developments.

Kathryn Angeles, Student Affairs Officer in the Computer Science department, was awarded the inaugural American College Personnel Association (ACPA) CASHE Fellowship.

The article titled "Spatial Computing," coauthored by Professor **Steven Feiner** of Computer Science, was the cover story of the January 2016 issue of Communications of the ACM (CACM), where past accomplishments, short-term opportunities and long-term research challenges of spatial computing are discussed.

A paper co-authored by Professor **Changxi Zheng** of Computer Science received the "Hot Paper Award" at ACM's Hotwireless 2015. The paper, titled "3D Printing Your Wireless Coverage," proposed WiPrint, a new computational approach to control wireless coverage by mounting signal reflectors in carefully optimized shapes on wireless routers.

Professor **Steven M. Bellovin**, coauthor of "Keys Under Doormats: Mandating Insecurity by Requiring Government Access to All Data and Communications," received the M³AAWG 2015 J.D. Falk Award. The paper explains potential issues raised from the government's request for a system that would allow it to access any secured file.

Professor **Steven Nowick** of Computer Science received a $420,000 National Science Foundation (NSF) award titled "An Asynchronous Network-on-Chip Methodology for Cost-Effective and Fault-Tolerant Heterogeneous SoC (System-on-Chip) Architectures" to explore and significantly advance plug-and-play systems for industrial applications. This grant will fund several significant new research directions in the area of asynchronous on-chip networks and systems.

Professor **Martha Kim** was named the recipient of the Edward and Carole Kim Award for Faculty Involvement. This award honors a faculty member demonstrating teaching excellence and a special, personal commitment to students. Nominations are made by undergraduate and graduate students.

Professor **Shree Nayar** received the Distinguished Faculty Teaching Award along with James Hone (Mechanical Engineering). This award is given on behalf of students and alumni for excellence in teaching, including dedication to undergraduate students. Selection is based on student evaluations and recommendations of a selection committee made up of three students and two alumni.

Aaron Bernstein

Prof. **Clifford Stein**

A paper by CS PhD student **Aaron Bernstein** and Professor **Clifford Stein** won the Best Paper Award at the International Colloquium on Automata, Languages and Programming (ICALP 2016), the main European conference in Theoretical Computer Science. The paper is titled "Fully Dynamic Matching in Bipartite Graphs."

Jessica Ouyang

Prof **Kathleen McKeown**

A paper by CS PhD student **Jessica Ouyang** and Professor **Kathleen McKeown** received the Notable Data Set Award at the 2015 Conference on Empirical Methods in Natural Language Processing. The paper is titled "Modeling reportable events as turning points in narrative."

Daniel Sims

A paper co-authored by CS Professor **Shree Nayar** and Columbia Engineering researcher **Daniel Sims** won the Best Paper Award at the International Conference on Computational Photography. The paper is titled "Towards Self-Powered Cameras."

Professor **Shih-Fu Chang** of Electrical Engineering and of Computer Science was awarded an honorary doctorate by the University of Amsterdam "in recognition of his pioneering contribution to our understanding of the digital universe, particularly in the areas of imagery, language, and sound."

Yunsung Kim

Alison Y. Chang

Robert Ying

Three undergraduate computer science majors have been recognized by the Computing Research Association (CRA) for showing outstanding research potential in an area of computing research, including **Yunsung Kim** (SEAS'16) for his work on information privacy and anonymity in big data, **Alison Y. Chang** (CC'16) for her research in code switching, web scraping, and text-to-speech data selection, and **Robert Ying** (MS'16) for his work on assistive robotics and brain-computer interfaces.

A paper coauthored by CS PhD student **Georgios Kontaxis** received the Best Student Paper Award at the 2015 workshop on Web 2.0 Security and Privacy. The paper is titled "Tracking Protection in Firefox For Privacy and Performance."

Lucas Kowalczyk, a first-year PhD student in computer science, has been awarded a National Science Foundation (NSF) Graduate Research Fellowship, which recognizes and supports outstanding graduate students in science, technology, engineering, and mathematics disciplines.

Henrique Teles Maia, currently completing dual degrees in computer science and mechanical engineering, has recently been awarded a National Science Foundation (NSF) Graduate Research Fellowship and a GEM Fellowship awarded by The National Consortium for Graduate Degrees for Minorities in Engineering and Science (GEM).

CUCS student **Danfei Xu** (currently a PhD student at Stanford University) won Computing Research Association (CRA) Outstanding Undergraduate Researcher Award, for his research in sensory perception of robotic systems; in particular, tactile sensing, visual perception, and sensor fusion.

Adrian Tang

John Demme

Prof. **Simha Sethumadhavan**

Prof. **Salvatore Stolfo**

A paper by CS PhD student **Adrian Tang**, Associate Research Scientist **John Demme**, Professor **Simha Sethumadhavan**, and Professor **Salvatore Stolfo** received a Best Poster Award at the Hot Chips conference. The paper is titled "Anti-Virus in Silicon."

CS Postdoctoral Researcher **Alec Jacobson** received the 2015 SGP software award for leading the development of the widely used geometry processing library, libigl.

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 3593