CS@CU

NEWSLETTER OF THE DEPARTMENT OF COMPUTER SCIENCE AT COLUMBIA UNIVERSITY VOL.11 NO.1 SPRING 2015

CUCS **Welcomes Two New Faculty**

YANIV ERLICH

ASSISTANT PROFESSOR OF COMPUTER SCIENCE

Dissecting the Complex Relationships of Genes, Health, and Privacy

Health and predisposition to disease strongly depend on the genetic material carried deep within the cells of every individual. Increasingly this genetic information, unique to every individual, is becoming public. For **Yaniv Erlich**, it's two parts of the complex relationship between genes and health, one increasingly understood through quantitative analysis and computer algorithms.

"Computational methods are necessary at every step in examining the genome," says Erlich, assistant professor of computer science. "The strings of nucleotides (A, T, G, and C) that form each person's unique genome are billions of letters long. It's not even possible to look at these long sequences without a computer. As more genetic data is collected, concepts from machine learning, statistics, and signal processing are needed to detect the subtle variations and statistical patterns that reveal traits and predisposition to disease."

Erlich is well-positioned for untangling the complex genetic underpinnings of human biology. An initial love of math and a friend's chance remark that biology entailed a lot of mathematics piqued his interest, and he went on to study both biology and genetics, two sciences increasingly awash in data and in need of algorithms for inferring information from that data. He approached both subjects from a computational perspective.

At MIT's Whitehead Institute of Biomedical Research, where he headed a research lab from 2010 through 2014, he had the chance to develop and then apply new computational tools to genetics.

Results were impressive. One method sequenced tens of thousands of samples at a time. Another harnessed signal processing and statistical learning to extract genetic information from short tandem repeats, or STRs, a fast-mutating fragment of DNA so small it had been mostly neglected by the research community. Both methods contributed new information about how genes operate at the molecular level.

But Erlich was also interested in how genes affect health and traits. Is the relationship linear where mutations predictably sum to a trait, or is the relationship nonlinear, with mutations interacting unpredictably with one another? The answer required a population-scale analysis, one too large to be constructed using traditional data collection methods. Here, Erlich and his Whitehead colleagues came up with an innovative solution. Using genealogical data already uploaded to social media sites, they created a genealogical tree of 13 million individuals dating back to the 15th century. Such a deep genealogy reveals clusters of genetic variation, with some tied to longevity and some to rare disorders. (And offering as a side bonus an intriguing view into human migration.) Looking at how mutations ripple through populations, researchers can measure the frequency of a certain trait, and thus evaluate the genetic contribution. With a larger tree, even more will become possible.

The promise is great but it all means nothing if there isn't genetic material to work with. And that requires trust.

People today are more wary about revealing personal data, and they are right to be so. Erlich himself is one of the first to flag the ethical complexities involving genetics and privacy. A paper he spearheaded, released in January 2013, caused a stir by showing how easy it is to take apparently anonymized genetic information donated by research participants and cross-reference it to online data sources. Using only Internet searches with no actual DNA, Erlich and his research team were able to correlate the donated DNA to a surname in 13% of the U.S. population, a result that surprised even Erlich.

"Our study highlighted current gaps in genetic privacy as we enter to the brave new era of ubiquitous genetic information," explains Erlich. "However, we must remember that sharing genetic information is crucial to understanding the hereditary basis of devastating disorders. We were pleased to see that our work has helped facilitate discussions and procedures to better share genetic information in ways that respect participants' preferences."

Erlich hopes that ensuring better safeguarding of genetic information will encourage more people to contribute their DNA, speeding the day when personalized medicine becomes a reality for all. For those most at risk of rare genetic diseases, the bigger danger may be in not contributing genetic information.

BSc, Tel-Aviv University, 2006;
PhD, Watson School of Biological Sciences, Cold Spring Harbor Laboratory, 2010

By Linda Crane

DAVID BLEI
PROFESSOR OF COMPUTER SCIENCE AND STATISTICS

Embracing the Science of Teaching Topics to Machines

Consider the challenge of the modern-day researcher: Potentially millions of pages of information dating back hundreds of years are available to be read from a computer screen. How does a simple Internet search deliver appropriate findings?

The answer is topic modeling, a mathematical approach to uncovering the hidden themes in collections of documents. **David Blei**, professor of computer science and statistics, led the groundbreaking research that resulted in the development of the Latent Dirichlet Allocation (LDA) model, a topic model capable of exploiting hidden topics and semantic themes among countless lines in billions of documents via language processing. By co-developing the LDA, Blei effectively influenced the research and development of topic models and machine learning.

"We were fortunate that LDA has become influential for analyzing all kinds of digital information, including text, images, user data and recommendation systems, social networks, and survey data," says Blei. "Even in population genetics, LDA-like ideas were developed independently and are used to uncover patterns in collections of individuals' gene sequences."

Blei, whose challenges and rewards both come from solving large-scale machine learning problems, appreciates the interconnectivity of disciplines within his field.

"Columbia has an extremely strong faculty in machine learning and statistics, and across other departments, such as in medicine and neuroscience. This makes it an ideal university for my group to pursue our interdisciplinary research agenda," he says. "I like working on interesting applied problems about data—in the sciences, social sciences, and humanities—and letting real-world problems drive the methodological advances that we develop."

Blei's interest in data science and machine learning was reinforced by the work of Herbert Robbins, a former professor of mathematical statistics at Columbia. "Robbins developed an algorithm that allows the machine learning community to scale up algorithms to massive data," says Blei, who predicts that one of the next breakthroughs in modern machine learning and statistics will involve observational data, or that which is observed but not collected as part of a carefully controlled experiment.

"Previously, machine learning has focused on prediction; observational data has been difficult to understand and exploring that data was a fuzzily defined activity," Blei explains. "Now, massive collections of observational data are everywhere—in government, industry, natural science, and social science—and practitioners need to be able to quickly explore, understand, visualize, and summarize them. But to do so, we need new statistical and machine learning tools that can help reveal what such data sets say about the real world."

Prior to joining Columbia Engineering in July of 2014, Blei was an associate professor in the Department of Computer Science at Princeton University. He is a faculty affiliate of Columbia's Institute for Data Sciences and Engineering. Blei is a recipient of the 2013 ACM-Infosys Foundation Award in Computing Sciences and is a winner of an NSF CAREER Award; an Alfred P. Sloan Fellowship, and the NSF Presidential Early Career Award for Scientists and Engineers. His accolades also include the Office of Naval Research Young Investigator Award and the New York Academy of Sciences Blavatnik Award for Young Scientists.

BSc, Brown University, 1997;
PhD, University of California Berkeley, 2004

By Dave Meyers

Columbia Engineering Team **Finds Thousands of Secret Keys in Android Apps**

Professor of Computer Science **Jason Nieh**

PhD Candidate **Nicolas Viennot**

In a paper presented—and awarded the prestigious Ken Sevcik Outstanding Student Paper Award—at the ACM SIGMETRICS conference on June 18, **Jason Nieh**, professor of computer science at Columbia Engineering, and PhD candidate **Nicolas Viennot** reported that they have discovered a crucial security problem in Google Play, the official Android app store.

"Google Play has more than one million apps and over 50 billion app downloads, but no one reviews what gets put into Google Play—anyone can get a $25 account and upload whatever they want. Very little is known about what's there at an aggregate level," says Nieh, who is also a member of the University's Institute for Data Sciences and Engineering's Cybersecurity Center. "Given the huge popularity of Google Play and the potential risks to millions of users, we thought it was important to take a close look at Google Play content."

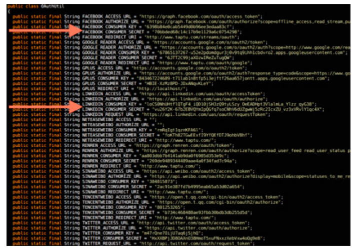
Nieh and Viennot's paper is the first to make a large-scale measurement of the huge Google Play marketplace. To do this, they developed PlayDrone, a tool that uses various hacking techniques to circumvent Google security to successfully download Google Play apps and recover their sources. PlayDrone scales by simply adding more servers and is fast enough to crawl Google Play on a daily basis, downloading more than 1.1 million Android apps and decompiling over 880,000 free applications.

Nieh and Viennot discovered all kinds of new information about the content in Google Play, including a critical security problem: developers often store their secret keys in their apps software, similar to usernames/passwords info, and these can be then used by anyone to maliciously steal user data or resources from service providers such as Amazon and Facebook. These vulnerabilities can affect users even if they are not actively running the Android apps. Nieh notes that even "Top Developers," designated by the Google Play team as the best developers on Google Play, included these vulnerabilities in their apps.

"We've been working closely with Google, Amazon, Facebook, and other service providers to identify and notify customers at risk, and make the Google Play store a safer place," says Viennot. "Google is now using our techniques to proactively scan apps for these problems to prevent this from happening again in the future."

In fact, Nieh adds, developers are already receiving notifications from Google to fix their apps and remove the secret keys.

Nieh and Viennot expect PlayDrone to lay a foundation for new kinds of analysis of Android apps. "Big data is increasingly important and Android apps are just one form of interesting data," Nieh observes. "Our work makes it possible to analyze Android apps at large scale in new ways, and we expect that PlayDrone will be a useful tool to better understand Android apps and improve the quality of application content in Google Play."



The researchers used PlayDrone to recover app sources and, in the process, uncovered crucial security flaws.

Other findings of the research include:

- Showing that roughly a quarter of all Google Play free apps are clones: these apps are duplicative of other apps already in Google Play.
- Identifying a performance problem resulting in very slow app purchases in Google Play: this has since been fixed.
- A list of the top 10 most highly rated apps and the top 10 worst rated apps in Google Play that included surprises such as an app that, while the worst rated, still had more than a million downloads: it purports to be a scale that measures the weight of an object placed on the touchscreen of an Android device, but instead displays a random number for the weight.

Good news for the hundreds of thousands of developers who upload content to Google Play and even more so for the millions of users who download the content!

By Holly Evarts

Chasing Real Security

An interview with **Allison Bishop Lewko**

Assistant Professor
Allison Bishop Lewko

Encrypting software so it performs a function without the code being intelligible was long thought by many to be impossible. But in 2013, researchers—unexpectedly and in a surprising way—demonstrated how to obfuscate software in a mathematically rigorous way using multilinear maps. This breakthrough, combined with almost equally heady progress in fully homomorphic encryption, is upending the field of cryptography and giving researchers for the first time an idea of what real, provable security looks like. The implications are immense, but two large problems remain: creating efficiency and providing a clear mathematical foundation. For these reasons, five researchers, including many of those responsible for the recent advances, have formed the Center for Encrypted Functionality with funding provided by NSF. ***Allison Bishop Lewko*** *is part of this center. In this interview she talks about the rapidly evolving security landscape, the center's mandate, and her own role.*

What happened to make secure software suddenly seem possible?

The short answer is that in 2013, six researchers—Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova , Amit Sahai, Brent Waters—published a paper showing for the first time that it is possible to obfuscate software in a mathematically rigorous way using a new type of mathematical structure called *multilinear maps*. This means I can take a software program and put it through the mathematical structure provided by these multilinear maps and encrypt software to the machine level while preserving its functionality. Anyone can take this obfuscated software and run it without a secret key and without being able to reverse-engineer the code. To see what's happening in the code requires solving the hard problem of breaking these multilinear maps.

This process of encrypting software so it still executes, which we call indistinguishability obfuscation, was something we didn't believe was out there even two or three years ago. Now there is this first plausible candidate indistinguishabilty obfuscator using multilinear maps.

And it's rigorous. Today there are programs that purport to obfuscate software. They add dead code, change the variables, things you imagine you would get if you gave a sixth-grader the assignment to make your software still work but hard to read. It is very unprincipled and people can break it easily. Indistinguishability obfuscation is something else entirely.

How does obfuscation differ from encryption?

Both obfuscation and encryption protect software, but the nature of the security guarantee is different. Encryption tells you your data or software is safe because it is unintelligible to anyone who doesn't have the secret key to decrypt it. The software is protected but at a cost of not being usable. The software is either encrypted and protected, or it is decrypted and readable.

Obfuscation is a guarantee to protect software, but there is this added requirement that the encrypted software still has to work, that it still performs the function it was written to do. So I can give you code and you can run it without a secret key but you can't reverse-engineer the code to learn how it works.

The bar is much higher for obfuscation because obfuscated software must remain a functional, executable program. Protecting and using at the same time is what is hard. It's easy to protect software if you never have to use it again.

These concepts of encryption and decryption, however, do live very close together and they interact. Encryption is the tool used to accomplish obfuscation, and obfuscation can be used to build an encryption.

How so? Do you use the same methods to encrypt data to obfuscate code?

Not exactly, though the general concept of substituting something familiar with something seemingly random exists in both.

In the well-known RSA encryption method, each character of the output text is replaced by a number that gets multiplied by itself a number of times (the public key) to get a random-looking number that only the secret, private key can unlock and make sense of.

When you obfuscate software, each bit of the original program gets associated with some number of slots. For every slot there are two matrices, and every input bit determines what matrix to use.

The result is a huge chain of matrices. To compute the program on an input bit stream requires selecting the corresponding matrices and multiplying them.

Is having two matrices for every slot a way to add randomness?

No. The choice of what matrix to use is deterministic based on the input. But we do add randomness, which is a very important part of any cryptography scheme since it's randomness that makes it hard to make out what's going on. For this current candidate obfuscation scheme, we add randomness to the matrices themselves (by multiplying them by random matrices), and to representations we create of the matrices—like these multilinear maps.

But we do it very carefully. Among all these matrices, there are lots of relationships that have to be maintained. Adding randomness has to be done in a precise way so it's enough to hide things but not so much to change software behavior.

There is this tension. The more functionality to preserve, the more there is to hide, and the harder it is to balance this tension. It's not an obvious thing to figure out and it's one reason why this concept of obfuscation didn't exist until recently.

Amit Sahai, director of the Center, compares obfuscated software to a jigsaw puzzle, where the puzzle pieces have to fit together the right way before it's possible to learn anything about the software. The puzzle pieces are the matrix encoded in the multilinear map, and the added randomness makes it that much harder to understand how these pieces fit together.

What does obfuscated software look like?

You would not want to see that. A small line becomes a page of random characters. The original code becomes bit by bit multiplication matrices. And this is one source of inefficiency.

Actually for this first candidate it's not the original software itself being put through the obfuscation, but the software transformed into a circuit (a bunch of AND, OR, NOT gates), and this is another source of inefficiency.

Indistinguishability obfuscation is said to provide a richer structure for cryptographic methods. What does that mean?

All cryptography rests on trying to build up structures where certain things are fast to compute, and certain things are hard. The more you can build a richer structure of things that are fast, the more applications there are that can take advantage of this encryption. What is needed is a huge mathematical structure with a rich variety of data problems, where some are easy, some are hard.

In RSA the hard problem is factoring large numbers. One way—multiplying two prime numbers—is easy; that's the encryption. Going back—undoing the factorization—is hard unless you have the key. But it's hard in the sense that no one has yet figured out how to factorize large numbers quickly. That nobody knows how to factor numbers isn't proven. We're always vulnerable to having someone come out of nowhere with a new method.

RSA is what we've had since the 80s. It works reasonably well, assuming no one builds a quantum computer.

But RSA and other similar public key encryption schemes don't work for more complicated policies where you want some people to see some data and some other people to see other data. RSA has one hard problem and allows for one key that reads everything. There's no nuance. If you know the factors, you know the factors and you can read everything.

A dozen years ago (in 2001) Dan Boneh working with other people discovered how to build crypto systems using bilinear maps on elliptic curves (where the set of elements used to satisfy an equation consists of points on a curve rather than integers). Elliptic curves give you a little more structure than a single function; some things are easy to compute but there are things that are hard to compute. The same hard problem has different levels of structure.

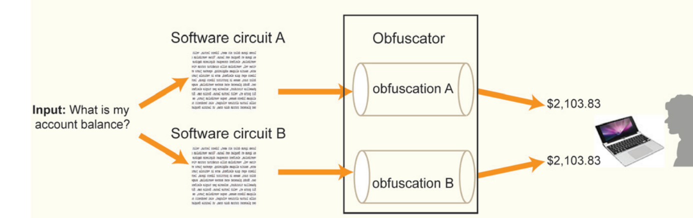
Think of a group having exponents. Instead of telling you the process for getting an exponent out, I can give you one exponent, which to some extent represents partial information, and each piece of partial information lets you do one thing but not another. The same hard problem can be used to generate encryption keys with different levels of functionality, and it all exist within the same system.

This is the beginning of identity-based security where you give different people particular levels of access to encrypted data. Maybe my hospital sees a great deal of my data, but my insurance company sees only the procedures that were done.

The structure of bilinear maps allows for more kinds of keys with more levels of functionality than you would get with a simple one-way function. It's not much of a stretch to intuitively understand that more is possible with trilinear maps or with multilinear maps.

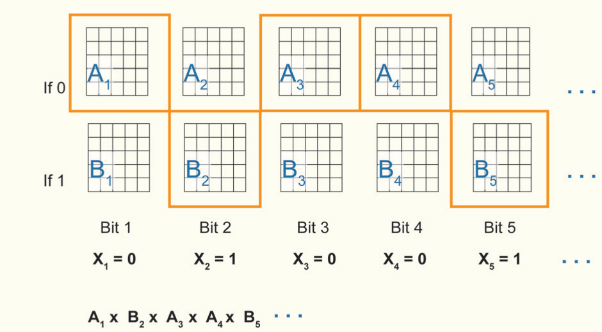
For a dozen years it was an open problem on how to extend bilinear maps to multilinear maps in the realm of algebraic geometry, but elliptic curves didn't seem to have a more computationally efficient structure on which to build. During this time, Craig Gentry (now at IBM) was working on fully homomorphic encryption, which allows someone to perform computations on encrypted data (the results will also be encrypted). Gentry was using lattices, a type of mathematical structure from number theory, an entirely different mathematical world than algebraic geometry.

Indistinguishability obfuscation



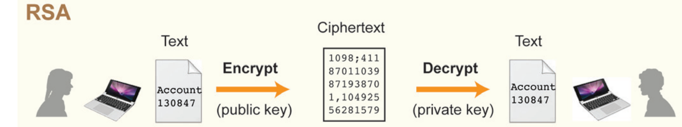
Indistinguishability obfuscation is achieved when a software circuit is obfuscated in such a way that two different obfuscations are indistinguishable.

Bits into matrices



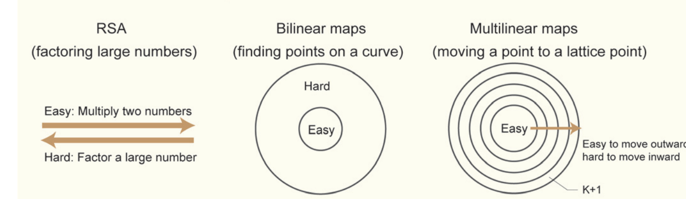
Each bit is represented by one of two matrices, with 0 and 1 determining which matrix to use. The selected matrices are then multiplied together. Multiplying two matrices not meant to be multiplied prevents the software from being correctly reassembled.

RSA



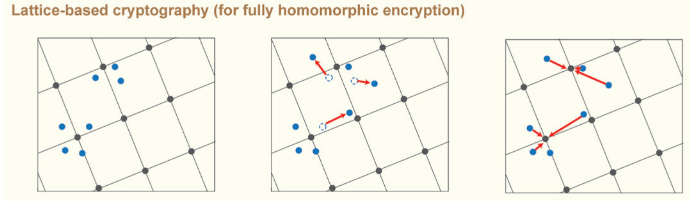
RSA is a single function where going one way is easy and one way is hard. It requires generating both a private and public key at the same time.

Structures for hard problems



Richer structures for hard problems expand the applications for cryptography.

Lattice-based cryptography (for fully homomorphic encryption)



In lattice-based cryptography, encrypted data can be envisioned as points at some random distance from a point on a lattice (shown here in two dimensions, though lattices can have hundreds). On decryption, each blue point is attracted to the nearest lattice point. Homomorphic operations amplify the random displacements. If the noise exceeds a threshold, some points gravitate to the wrong lattice point, leading to the wrong decryption.

Lattices provide a rich family of structures, and lattice schemes hide things by adding noise. As operations go on, noise gradually builds up and that's what makes homomorphic encryption hard.

You might think immediately that these lattices schemes serve as a multilinear map but there isn't a sharp divide between what's easy and hard; there is a falloff. With fully homomorphic encryption having this gradual quality, it wasn't at all clear how to use it in a way to build something where multiplying three things is easy but multiplying four things is hard.

The solution came in 2013 with the publication of *Candidate Multilinear Maps from Ideal Lattices*, which describes this candidate I've been so far describing. The authors were able to make a sharp threshold between easy and hard problems within lattice problems. They did it by taking something like a fully homomorphic scheme and giving out a partial secret key; one that lets you do this one thing at level 10 but not at level 11. These partial keys, which are really weak secret keys, add more nuance, but it's extremely tricky. If you give out a weak key, you may make things easy that you don't want to be easy. You want it to break a little but not too badly. This is another source of inefficiency.

How soon before obfuscated software becomes a reality?

A long time. Right now, indistinguishability obfuscation is a proof of concept. Before it can have practical application, it should have two features: it should be fast, and we should have a good understanding of why it is secure, more than "this uses math and it looks like we don't know how to break it."

Indistinguishability obfuscation currently doesn't meet either criteria. It is extremely inefficient to what you want to run in real life—running an instance of a very simple obfuscated function would take seconds or minutes instead of microseconds —and we don't have a good answer for why we believe it's hard to break.

Is that the purpose of the Center for Encryption Functionalities?

Yes, the center's main thrusts will be to increase efficiency and provide a provable mathematical foundation. Many of the same people responsible for the big advances in indistinguishability obfuscation and fully homomorphic encryption are working at the center, and building on previous work.

We'll be examining every aspect of the current recipe to increase efficiency. There are many avenues to explore. Maybe there is a more efficient way of implementing multilinear maps that doesn't involve matrices. Maybe only certain parts of a program have to be encrypted. We're still early in the process.

The center also has a social agenda, to both educate others on the rapid developments in cryptography and bring together those working on the problem.

What is your role?

My role is to really narrow down our understanding of what we're assuming when we say indistinguishability obfuscation is secure. If someone today breaks RSA, we know they've factored big numbers, but what would it take to break indistinguishability obfuscation?

Indistinguishability obfuscation is still this very complicated, amorphous thing of a system with all these different possible avenues of attack. It's difficult to comprehensively study the candidate scheme in its current form and understand what makes it secure and to know exactly when we are vulnerable and when we are not.

I did my PhD in topics to expand the set of things that we can provably reduce to nice computational assumptions, which is very much what is at the heart of where this new obfuscation stuff is going. We need to take this huge, complicated scheme down to a simple computational problem to make it easy for people to study. Saying "factor two numbers" is a lot easier than saying break this process that takes a circuit to another circuit. We have to have this tantalizing carrot of a challenge to gain interest.

And on this front we have made great progress. We have a somewhat long, highly technical proof that takes an abstract obfuscation for all circuits, and reduces it to a sufficiently simple problem that others can study and try to break.

That's the first piece. What we don't yet have is the second piece, which is instantiating the simple problem as a hard problem in multilinear maps.

Have there been any attempts to "break" indistinguishability obfuscation?

Yes, in the past few months some attacks did beat back on some problems contained in our proof that we thought might be hard, but these attacks did not translate into breaks on the candidate obfuscation scheme. While we wished the problems had been hard, the attacks help us narrow and refine the class of problems we should be focusing on, and they are a reminder that this line between what is hard and what is not hard is really very much still in flux.

The attackers didn't just break our problem, they broke a landscape of problems for current multilinear map candidates, making them fragile to use.

Were you surprised that the problem was broken?

Yes, very surprised, though not super-surprised because this whole thing is very new. There are always breaks you can't imagine.

Does it shake your faith in indistinguishability obfuscation?

I have no faith to be shaken. I simply want to know the answer.

We learn something from this process either way. If the current multilinear map candidates turn out to be irretrievably broken—and we don't know that they are—we still learn something amazing. We learn about these mathematical structures. We learn about the structure of lattice-based cryptography, and the landscape of easy and hard problems.

Every time something is broken, it means someone came up with an algorithm for something that we didn't previously know. Bilinear maps were first used to break things; then Dan Boneh said if it's a fast algorithm that can break things, let's re-appropriate it and use it to encrypt things. Every new algorithm is a potential new tool.

The fact that a particular candidate can be broken doesn't make it implausible that other good candidates exist. We've also made progress on working with different kinds of computational models. In June, we will present *Indistinguishability Obfuscation for Turing Machines with Unbounded Memory* at STOC, which replaces circuits with a Turing machine to conceptually show that we can do obfuscation for varying input sizes.

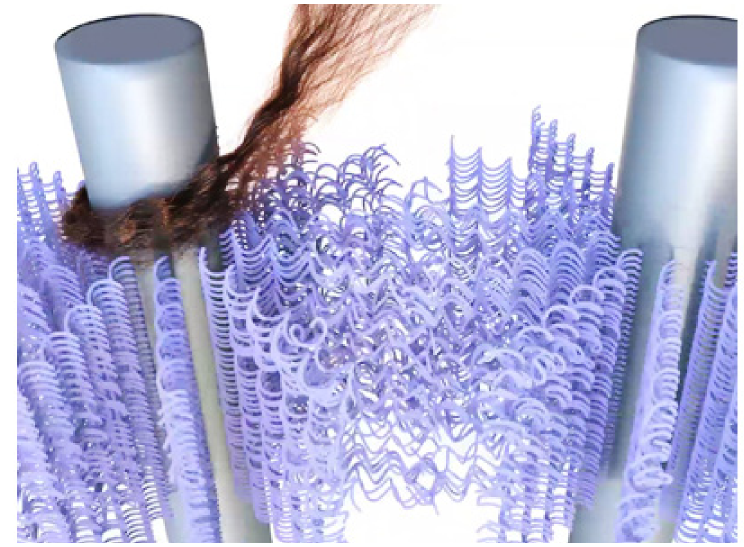
Things are still new; what we've done this year at the center is motivating work in this field. People will be trying to build the best multilinear maps for the next 20 years regardless of what happens with these particular candidates.

The ongoing cryptanalysis is an important feedback loop that points to targets that attackers will go after to try to break the scheme, and it will help drive the direction of cryptography. It's really an exciting time to be a researcher in this field.

About The Center for Encrypted Functionalities (CEF)

The Center's primary mission is to transform program obfuscation from an art to a rigorous mathematical discipline. It supports direct research, organizes retreats and workshops to bring together researchers, and engages in outreach and educational efforts.

How **Things Coil**



Discrete elastic rods, used for hair simulation here, are also being used to predict the coiling of undersea communication cables.

IMAGE CREDIT: DANNY M. KAUFMAN, RASMUS TAMSTORF, BREANNAN SMITH, JEAN-MARIE AUBRY, AND EITAN GRINSPUN

When one sends an email from Boston to Beijing, it travels through submarine optical cables that someone had to install at some point. The positioning of these cables can generate intriguing coiling patterns that can also cause problems if, for instance, they are tangled or kinked. The deployment of a rodlike structure onto a moving substrate is commonly found in a variety of engineering applications, from the fabrication of nanotube serpentines to the laying of submarine cables and pipelines, and engineers for years have been interested in predicting the mechanics of filamentary structures and the coiling process.

A team led by **Eitan Grinspun**, associate professor of computer science at Columbia Engineering, and Pedro Reis, associate professor of mechanical engineering and civil and environmental engineering at MIT, has been collaborating on a project that, in exploring these issues, bridges engineering mechanics (Reis's group) and computer graphics (Grinspun's group). The researchers combined precision model experiments with computer simulations and examined the mechanics of coiling, discovering in particular that the natural curvature of the rod dramatically affects the coiling process. Their study is published in the September 29 Early Online edition of *Proceedings of the National Academy of Sciences* (PNAS).

"This was a fun, fruitful collaboration," says Grinspun. "We did something totally new and different: we took a computer algorithm that we had designed for Hollywood, and, by teaming up with Reis's group, discovered that this same algorithm served as a predictive tool for engineering mechanics of thin filaments, rods, and pipes. It's exciting to think that this computer model can serve both creative and engineering enterprises."

Grinspun's simulation technology, Discrete Elastic Rods, was originally developed to animate hair and fur in film and graphics applications, licensed and used in Photoshop for realistic paintbrushes, and by Weta Digital for use in films such as The Hobbit and Planet of the Apes series. Reis, who is an experimental mechanician at MIT, was studying how buckling of thin elastic structures can be turned on its head: buckling is normally feared by engineers as a potential failure of a design, but what if it could be used as a functional component of a design? The two researchers decided to investigate how cables are deployed, both at the nanoscale, in stretchable electronics, and the macroscale, such as the deployment of internet communication cables on the ocean floor.

"This has been a wonderful example of two seemingly unrelated fields coming together to address a practical problem to introducing powerful and novel computational tools that were not previously available in our engineering community," Reis notes.

The collaboration between Grinspun and Reis began when Reis invited Grinspun to visit his lab at MIT. "We wondered if our seemingly distant worlds could be bridged by a common vision," says Grinspun. "We both wanted to understand how physical objects move by looking at how their geometry, or shape, affects their motion. Cables, being long and slender, were ideal candidates for study. But could the technology we built at Columbia Engineering for visually striking film and special effects be sufficiently accurate to agree with Reis's hard and precise experimental data?"

With support from the National Science Foundation, Reis and Grinspun recruited doctoral students Khalid Jawed (MIT) and Fang Da (Columbia Engineering) to study cable deployment in detail. In their PNAS article, the researchers describe how seemingly benign decisions, such as the diameter of a spool, or the speed at which a cable is deployed, can dramatically affect the way that the cable lies on the ground. They created a map of the different patterns that can arise, from a wiggling meandering mode to steady coiling and on to alternating loops, as the spool diameter or deployment speed are varied. The researchers also identified factors that have relatively little impact on the deployment, among them the height from which a cable is dropped.

"These findings have practical impacts on our everyday lives," Reis adds. "Take, for instance, an email that travels along a transoceanic communication cable. By better understanding the variables that impact the deployment of such cables, we can better balance considerations such as expense (the length of the cable deployed, the amount of time to deploy the cable), signal quality (tangled cables can be more prone to interference), and the resilience of the connection (taut cables are more prone to damage due to external factors, such as seismic activity."

"Translating computer tools from computers and validating them against precision model experiments has provided a novel tool for engineering mechanics to tackle the design and analysis of other rodlike structures, which are common in nature and technology," Reis continues.

"As we move to the next stage, we would like to pursue engineering problems that combine the mechanics of slender filaments with additional ingredients, such as drag, contact, and friction," adds Grinspun. "We are looking, for example, at locomotion of bacteria, tying of shoelaces, and hair blowing in the wind."

This work is funded by a National Science Foundation MoM-IDR Collaborative grant under CMMI (1129894).

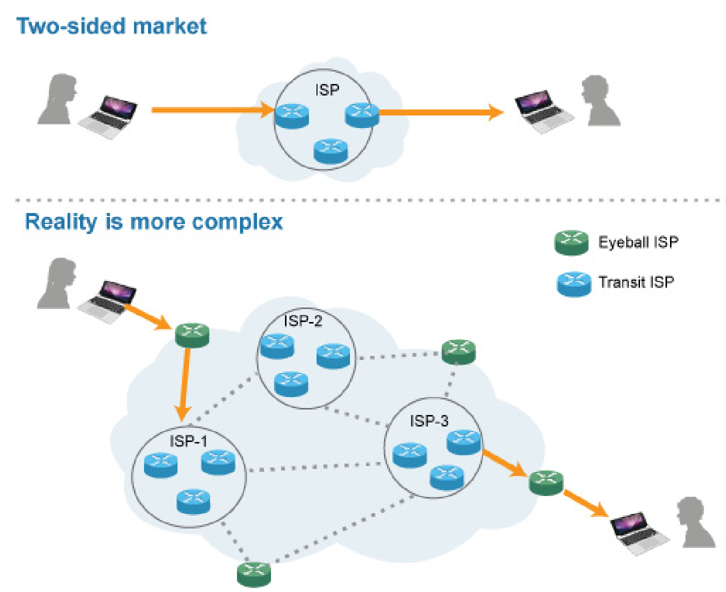
By Holly Evarts

Net neutrality is all good and fine; the real problem is elsewhere.

UPDATE: Since this article was posted in November 2014, the FCC voted (2/26/15) to enforce net neutrality rules that ban paid prioritization.

Associate Professor
Vishal Misra

As the debate over net neutrality rages, two papers co-authored by Vishal Misra have been recently reprinted in a special IEEE *issue on net neutrality. Both papers examine the economics of the Internet. The first (from 2008) looks at how profit-motivated decisions by the ISPs work against customers' preferences for fast Internet speeds at low prices; it also predicted the rise of paid peering, which plays a central role in today's net neutrality debate. The second (from 2012) shows how the introduction of a public-option ISP works to align the interests of both customers and ISPs, in the process achieving a free, open Internet without regulation.*



The Internet is operated by thousands of interconnected ISPs; each requires the cooperation of other ISPs to provide Internet services.

Understanding the economics of the Internet is crucial to understanding how best to maintain a free and open Internet—the stated goal of net neutrality.

The connection between Internet economics and net neutrality is not well understood, partly because net neutrality is considered more of a legal and regulatory issue, and partly because the Internet, a vast maze of inter-connected networks, is hugely complex. Economists themselves when analyzing Internet economics fail to take into account the complexity and instead rely on abstractions that simplify traffic flow, routing decisions, and the impact of congestion, and as a consequence fail to account for how the profit-driven decisions negatively affect Internet service for customers.

To bring a more networking angle to the economic analysis, Vishal Misra with his colleague Dan Rubenstein and their student Richard T. B. Ma (now Assistant Professor with the School of Computing, National University of Singapore) along with collaborators Dah-Ming Chiu and John Lui in Hong Kong set out in 2008 to realistically model the way Internet service providers (ISPs) interact with one another, looking particularly at the pricing policies used by ISPs to compensate one another for transit services. Their findings are described in *On Cooperative Settlement Between Content, Transit and Eyeball Internet Service Providers*, which used sophisticated mathematical modeling to accurately depict the complex interactions among different ISPs. Originally published in 2008, the paper obtained new insights into ISPs' routing decisions and the settlement (payment) among different ISPs, showing the inevitability of paid peering and the incentives supporting it.

This prediction made in 2008 is now coming true and is central to today's heated debate on net neutrality. This paper (one of two co-authored by Misra) was republished by the IEEE in a special net-neutrality issue that appeared in June 2014.

From symmetry to asymmetry

One main goal of the paper was to more realistically model the interactions among ISPs. Economists have typically analyzed and modeled the Internet as a two-sided market, with consumers on one side and content providers on the other, and a single ISP sitting in the middle serving them both. The reality even in 2008 was far more complex, with packets traveling through several ISPs before reaching the intended destination.

With mathematical models providing a more realistic picture of the Internet, the authors use game-theory economics, specifically Shapley values, to understand the optimal arrangement under which each ISP is incentivized to contribute to the overall good. This methodology showed how in the early Internet, the flow of traffic (mainly emails and files) was roughly symmetrical. A packet originating at ISP-A and handed off to ISP-B for delivery would be balanced by a packet moving in the opposite direction. ISPs often entered into no-cost agreements to carry one another's traffic, each figuring that the amount of traffic it carried for another ISP would be matched by that other ISP carrying its own traffic. Network neutrality prevailed naturally since ISPs, compensated for bandwidth used, did not differentiate one packet from another. The more packets of any kind, the more profits for all ISPs, an economic situation that aligned nicely with customers' interest in having an expanding supply of bandwidth.

Internet economics have changed considerably in recent years with the rise of behemoth for-profit content providers such as Facebook, Google, Amazon, and Netflix. Their modeling shows the appearance of these for-profit content providers did two things: it changed the Internet from a symmetric network to an asymmetric one where the vast preponderance of traffic flows from content providers to customers. And it introduced a new revenue stream, one outside the Internet and generated by advertising, online merchandizing, or payments for gaming, streaming video, and financial and other services. (Revenues from online video services alone grew 175% from $1.86 to $5.12 billion between 2010 and 2013. Source: FCC NPRM on the Open Internet.)

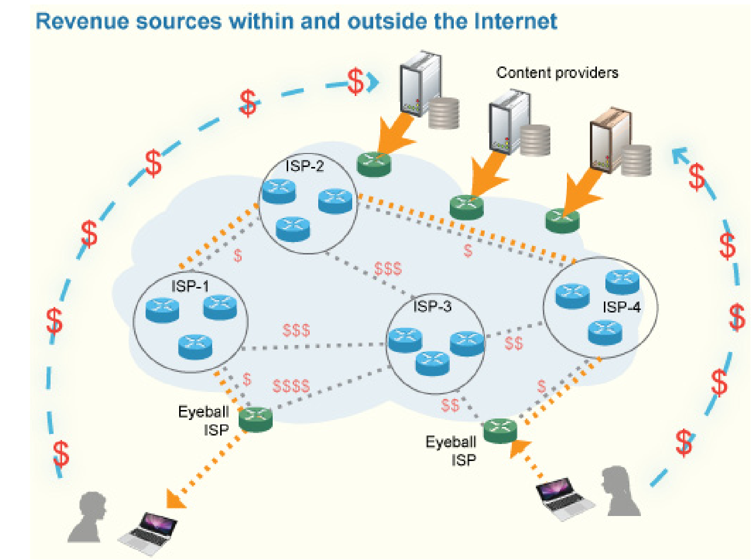
The ISPs themselves were changing and becoming more specialized. The paper subdivides ISPs into different but overlapping conceptual types, the main ones being eyeball ISPs such as Time Warner, Comcast, and AT&T that specialize in the last-mile content delivery to paying consumers, and the transit ISPs such as Level 3 and Cogent that own the global infrastructure and provide transit services for other ISPs. Where eyeball ISPs serve people, transit ISPs serve the content providers and earn revenue by delivering content to consumers on behalf of the content providers. Since transit ISPs don't have direct access to consumers, they arrange with the eyeball ISPs for the last-mile delivery of content to customers.

ISPs are cut off from the growing revenue source happening outside the Internet. The resulting imbalance in profit sharing reduces the investment incentives for ISPs to build out the network.

According to Misra, "In a coalition where one participant doesn't feel adequately compensated or at least compensated to the same degree as other participants, this gives rise to conflict. Mathematically, we call the situation being outside the core, where some participants have the incentive to leave the coalition since they will be better off without it. In schoolyard parlance, it is the equivalent of taking your ball and going home."

With an imbalance in the direction of traffic and no mechanism for appropriate compensation, the previous no-cost (or zero-dollar) bilateral arrangements broke down and were replaced by paid-peering arrangements where ISPs pay one another to carry one another's traffic. Each ISP adopts its pricing policies to maximize profit, and these pricing policies play a role in how ISPs cooperate with one another, or don't cooperate. Profit-seeking and cost-reduction objectives often induce selfish behaviors in routing—ISPs will avoid links considered too expensive for example—thus contributing to Internet inefficiencies.

Paid-peering is one ISP strategy to gain profits. For the eyeball ISPs that control access to consumers, there is another way. Charge higher prices by creating a premium class of service with faster speeds. Creating a two-class system is currently prohibited by network neutrality; however the FCC proposed in May new rules that would allow such a two-tier Internet, effectively ending at least one aspect of net neutrality: prohibiting the discriminations on delivering different contents on the network.



The Internet has evolved from a simple, symmetric network where light email and web traffic flowed between academics and researchers and the only revenues were from selling bandwidth, to an asymmetric one where traffic flows from content providers to consumers, generating massive revenues for content providers.

An imbalance of power

The eyeball ISPs, however, are in a power position because, unlike transit ISPs, eyeball ISPs have essentially no competition. Misra cites data released by FCC chairman Tom Wheeler showing that more than 19 percent of US residents have no broadband provider offering 25Mbps service, and another 55 percent have only one such provider. The ISPs that provide decent-speed service, for all intents and purposes are monopolies. While wireless Internet (including Wi-Fi) is often cited by ISPs as competition, the comparison is not fair. Unwired Internet is not fast as the wired service that is available only through the eyeball ISPs.

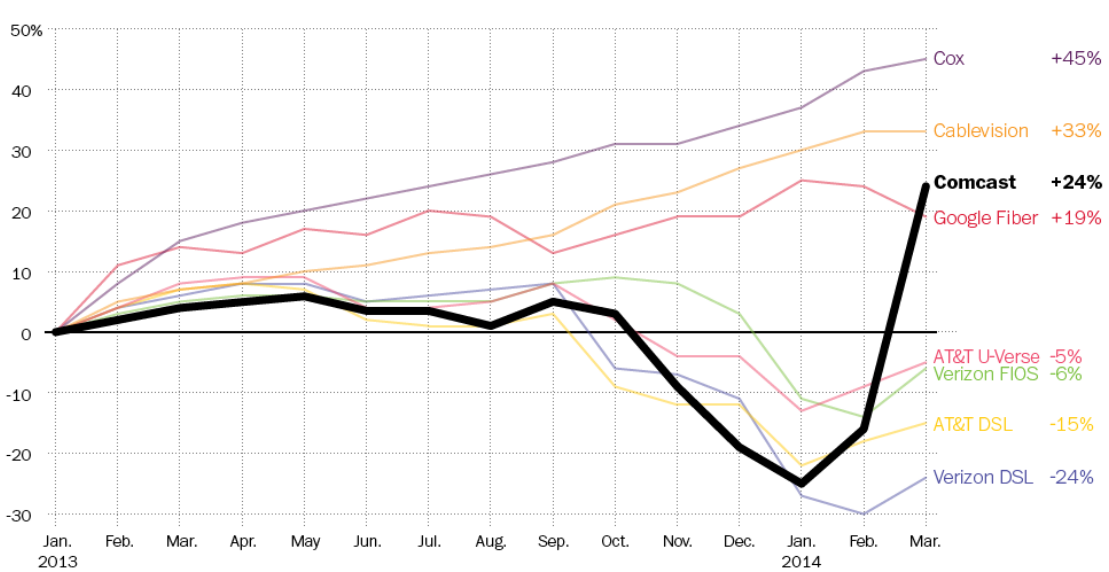
Content providers like Netflix are in a much weaker position than the eyeball ISPs. Content providers need ISPs much more than the ISPs need them. If Netflix were to disappear, other streaming services would rush to fill in the gap. For the ISPs, it matters little whether it's Amazon, Hulu, or another service (and worryingly, services run by the eyeball ISPs themselves) providing streaming services.

If ISPs don't need Netflix, neither do customers. Customers unhappy with Netflix's service can simply choose another such service. They can't, however, normally choose a different ISP.

The monopolistic power of the eyeball ISPs may soon be made stronger. The proposed Comcast takeover of Time Warner would create a single company that controls more than 40% of the US high-speed Internet service market. Anti-trust legislation would seem to prohibit such a takeover, but both companies argue that they don't compete. In this they are correct. The markets in which Time Warner and Comcast operate do not overlap. Seemingly by design, Comcast operates in San Francisco and Time Warner in Los Angeles, while Time Warner operates in New York City and Comcast in Philadelphia. Comcast Chairman and CEO Brian L. Roberts uses this argument to support his company's takeover of Time Warner, but Misra sees it differently. "Time Warner and Comcast claim they are not killing competition because they don't compete in the same markets. That is not an argument; that is a question. Why aren't they competing?"

Occupying a position of power and knowing customers are stuck, the eyeball ISPs can and do play hardball with content providers. This was effectively illustrated in the recent Netflix-vs-Comcast standoff when Comcast demanded Netflix pay additional charges (above what Netflix was already paying for bandwidth). When Netflix initially refused, Comcast customers with Netflix service started reporting download speeds so slow that some customers quit Netflix. These speed problems seemed to resolve themselves right around the time Netflix agreed to Comcast's demands. (Netflix has since signed a similar deal with Time Warner so that Netflix now has special arrangements with the four major ISPs, including AT&T and Verizon.)

% change in Netflix download speed since Jan. 2013, by I.S.P.



SOURCE: Netflix. GRAPHIC: reprinted from The Washington Post. Published April 24, 2014.

It would have been relatively inexpensive for Comcast to add capacity. But why should it? Monopolies such as Comcast have no real incentive to upgrade their networks. There is in fact an incentive to **not** upgrade since a limited commodity commands a higher price than a bountiful one. By limiting bandwidth, Comcast can force Netflix and other providers to pay more or opt into the premium class.

While the eyeball ISPs such as Comcast and Time Warner are monopolistic, the transit ISPs are not and must compete with one another. Any transit ISP that charges higher prices for the same global connectivity as another ISP is going to lose business. As reported by Vox.com (from data collected from DrPeering.com), average transit prices have fallen by a factor of 1000 since 1998, a far different story than what is happening in the world of monopolistic eyeball ISPs.

If customers notice the eyeball ISPs' failure to upgrade service through slower download speeds, transit ISPs notice it through blocked or over-utilized interchange points. Transit ISPs have long noticed that certain interface points are regularly over-utilized to the point they cannot accept additional traffic and start dropping packets. The interesting point is that blocked interchanges occur only in those markets monopolized or dominated by a single eyeball ISP and nowhere else. Internet exchange points, which are essentially line cards on routers, are not expensive; at a cost of $30,000, they represent a rounding error to Comcast's daily profits. Adding new interchanges would easily accommodate increased data traffic from Netflix or any other content provider, but for profit reasons such upgrades are resisted. (For more about the perspective from the transit side, see *blog.level3.com/open-internet/observations-internet-middleman.*)

Besides charging more for fast Internet lanes, ISPs have other ways to extract revenues from content providers. What Netflix paid for in its deal with Comcast was not a fast lane in the Internet, but a special arrangement whereby Comcast connects directly to Netflix's servers to speed up content delivery. It is important to note that this arrangement is not currently covered under conventional net neutrality, which bans fast lanes over the Internet backbone. In the Netflix-Comcast deal, Netflix's content is being moved along a private connection and never reaches the global Internet. (Tom Wheeler, chairman of the FCC, has acknowledged that these interconnection arrangements do not fall within the scope of any net-neutrality protections.)

According to Misra, net neutrality loses its meaning and becomes irrelevant when ISPs and content providers arrange private pathways that avoid the global links. "Technically you restrict fast lanes on a public highway, but you are freely allowing private highways for paying content providers. The effect is much the same."

In the face of the monopolistic power of ISPs and their stiff resistance to regulation (which they are increasingly able to avoid in any case) and with perverse incentives to not increase bandwidth, the hope for maintaining a free and open Internet would seem to be a lost cause.

A little competition would help

Susan Crawford, a professor of Law who served as President Barack Obama's Special Assistant for Science, Technology, and Innovation Policy and author of *Captive Audience, The Telecom Industry and Monopoly Power in the New Gilded Age* and *The Responsive City*, has extensively researched the issue from a legal and consumer perspective. She has been a vocal proponent of municipal broadband, arguing that competition will keep prices in line and improve services. It's an argument that makes sense and it forms the basis for US anti-trust legislation. But will it work?

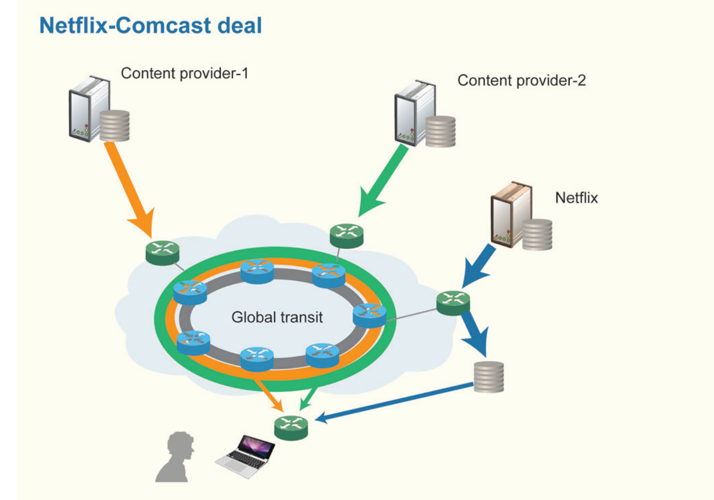
According to Ma and Misra, it will. In *The Public Option: A Non-regulatory Alternative to Network Neutrality* (originally published 2012) they again use game-theory economics to model how various participants would act in different scenarios, including a monopolistic one regulated by net neutrality as well as a scenario that introduces a public-option ISP.

The mathematical models they built more realistically simulate actual traffic flow on the Internet, taking into account the complexities of TCP, specifically how TCP traffic behaves in congestion. The previous models used by economists had assumed open-loop traffic (i.e., the bandwidth demand of the flows ignore the level of congestion) moving through what is called in queueing theory an M/M/1 queue. This seemingly sophisticated mathematical model view ignores that TCP is closed loop (meaning it reacts to congestion) and that TCP packets move through a complex network of routers (packets from the same origin might take different paths). Closed- and open-loop traffic flows behave very differently under congestion, and economic models that fail to account for this and other differences often arrive at incorrect conclusions about the quality of service delivered by the network.

Using more sophisticated modeling, Ma and Misra are able to show that even under tightly regulated net-neutrality, monopolies increase their profits when bandwidth is scarce. However, the introduction of a public-option ISP changes the equation. Under a scenario that includes a public-option ISP, profits become tied to market share, thus incentivizing ISPs to increase bandwidth to make themselves more attractive to customers. If the existing ISP employs some non-neutral strategies, consumers have the option to move to a network-neutral public ISP.

According to Crawford "(lack of) competition for high-capacity Internet access is a big issue here in the US. A public-option in the form of municipal fiber is something that I have long advocated as a possible solution to the issue. Vishal's work validates technically what a lot of us have been arguing for from a policy perspective."

This is borne out by real-world examples in those cities that have municipal broadband. In Stockholm where the city laid fiber (raising money through municipal bonds), customers have 100 Mbps Internet at one quarter the price of what customers pay in the US for a connection that is 1/10th the speed. And the city makes money. In the UK market, content providers can choose from at least four ISPs that compete for their business. It's a market where there is plenty of content, prices are low, and speeds are fast, all with zero network neutrality regulations. Zero.



A backdoor around net neutrality. Netflix paid Comcast to connect directly to its servers, bypassing global transit where net neutrality is currently enforced (and bypassing also the transit ISPs).

The pattern holds in the US. Susan Crawford uses the example of Santa Monica's 10-Gbps broadband service City Net, which charges businesses a third of what a private operator would charge. (A second, important benefit of municipal broadband is ensuring Internet access to low-income families and those in rural areas, populations not well-served by for-profit companies.)

So why aren't other US cities doing more to promote municipal broadband? Because the powerful ISPs and their lobbyists have persuaded politicians in 19 states to enact legislation to prohibit public-funded municipal networks, this despite the 1996 Telecommunications Act, which prevents states from blocking any entity looking to provide telecommunications services. The same ISPs that resist regulations needed to enforce net neutrality apparently have no problem with legislation or regulation when it suits their purposes or maintains their monopolistic status.

ISPs like any business entity will always seek and find profits in any scenario. The goal of inserting a public-option ISP is not to deny profits to ISPs but to align the profit motives of broadband ISPs with customer interests. By forcing ISPs to compete for customer subscriptions on the merits of better service and abundant bandwidth, a municipal ISP does just that.

It is a view that the FCC is now coming around to embrace. In remarks made just recently (October 1), FCC Chairman Wheeler cited the importance of competition and local choice to increase broadband and Internet speed. More than net-neutrality, as the mathematical modeling by Ma and Misra shows, it is competition that will ensure both a free and open Internet and that the US remains competitive in the global, knowledge-based economy.

By Linda Crane

Automatically Describing Disasters

An interview with **Kathy McKeown**

"The goal is to build a system for automatically generating a comprehensive description of a disaster, one that includes objective, factual information—what is the specific sequence of events of what happened when—with compelling, emotional first-person accounts of people impacted by the disaster."

Henry and Gertrude Rothschild Professor of Computer Science
Kathy McKeown

A deck of playing cards honors 54 notable women in computing. Produced by Duke University and Everwise (in conjunction with CRA-W and Anita Borg Institute Wikipedia Project), the first card decks were distributed at this year's Grace Hopper conference. (A Kickstarter campaign is raising money for a second printing.)

*One of the women to be honored with her own card is **Kathy McKeown**, a computer scientist working in the field of natural language processing. The first female full-time professor in Columbia's school of engineering, she was also the first woman to serve as department chair. Currently the Henry and Gertrude Rothschild Professor of Computer Science, she is also the inaugural director of Columbia's multidisciplinary Data Science Institute. And now she is the seven of clubs.*

Congratulations on being included among the Notable Women of Science card deck. What does it feel like to see your picture on the seven of clubs?

It is really exciting to be part of this group, to be part of such a distinguished group of computer scientists.

You started at the Columbia Engineering school in 1982 and were for a time the only full-time female professor in Columbia's school of engineering. From your experience seeing other women follow soon after, what's most helpful for women wanting to advance in computer science?

Just having other women around makes a big difference. Women can give one another advice and together support issues that women in particular care about. Having a woman in a senior position is especially helpful. When I was department chair, several undergraduate women approached me about starting a Women in Computer Science group. As a woman, I understood the need for such a group and in my position it was easier for me to present the case.

Of course, getting women into engineering and computer science requires making sure girls in school remain interested in these subjects as they approach college, and I think one way that is done is by showing them how stimulating and gratifying it can be to design and build something that helps others.

Talking of interesting work that helps others, you recently received an NSF grant for "Describing Disasters and the Ensuing Personal Toll." What is the work to be funded?

The goal is to build a system for automatically generating a comprehensive description of a disaster, one that includes objective, factual information—what is the specific sequence of events of what happened when—with compelling, emotional first-person accounts of people impacted by the disaster. We will use natural language processing techniques to find the relevant articles and stories, and then weave them together into a single, overall resource so people can query the information they need, both as the disaster unfolds or months or years later.

It might be emergency responders or relief workers needing to know where to direct their efforts first; it might be journalists wanting to report current facts or researching past similar disasters. It might be urban planners wanting to compare how two different but similar neighborhoods fared under the disaster, maybe discovering why one escaped heavy damage while the other did not. Or it might be someone who lived through the event coming back in years after to remember what it was like.

It's not a huge grant—it's enough for me and two graduate students for three years, but it's a very appealing one, and it's one that allows undergraduates to work with us and do research. Already we have seven students working on this, the two funded graduate students and five undergraduates.

It makes sense that journalists and emergency responders would want an objective account of a disaster or event. Why include personal stories?

Because it's the personal stories that let others understand fully what is going on. Numbers tell one side of the story, but emotions tell another. During Katrina, there were reports of people marooned in the Superdome in horrendous conditions—reports that were initially dismissed by authorities. It wasn't until reporters went there and interviewed people and put up their stories with pictures and videos that the rest of us could actually see the true plight and desperation of these people, and you realize at the same time it could be you. It wasn't possible to ignore what was happening after that, and rescue efforts accelerated.

What was the inspiration?

For us it was Hurricane Sandy, since many of us were here in New York or nearby when Sandy struck. One student, whose family on the Jersey Shore was hard-hit, was particularly motivated to look at what could be done to help people.

But more importantly, Sandy as an event has interesting characteristics. It was large scale and played out over multiple days and multiple areas, and it triggered other events. Being able to provide a description of an event at this magnitude is hard and poses an interesting problem. The project is meant to cover any type of disaster—earthquakes, terror attacks, floods, mass shootings—where the impact is long-lasting and generates sub-events.

How will it work?

The underlying system will tap into streaming social media and news to collect information, using techniques of natural language processing and artificial intelligence to find those articles and stories pertinent to a specific disaster and the sub-events it spawns. Each type of disaster is associated with a distinct vocabulary and we'll build language models to capture this information.

Obviously we'll look at established news sites for factual information. To include first-person stories, it's not yet entirely clear where to look since there aren't well-defined sites for this type of content. We will be searching blogs and discussion boards and wherever else we can discover personal accounts.

For users—and the intent is for anyone to be able to use the system—we envision currently some type of browser-type interface. It will probably be visual and may be laid out by location where things happened. Clicking on one location will present descriptions and give a timeline about what happened at that location at different times, and each sub-event will be accompanied by a personal account.

Newsblaster is already finding articles that cover the same event. Will you be building on top of Newsblaster?

Yes, after all, Newsblaster represents 11 years of experience in auto-generating summaries and it contains years of data, though we will modernize it to include social media, which Newsblaster doesn't currently do. We also need to also expand the scope of how Newsblaster uses natural language processing. Currently it relies on common language between articles to both find articles of the same event and then to produce summaries. I'm simplifying here, but Newsblaster works by extracting nouns and other important words from articles and then measuring statistical similarity of the vocabulary in these articles to determine which articles cover the same topic.

In a disaster covering multiple days with multiple sub-events, there is going to be a lot less common language and vocabulary among the articles we want to capture. A news item about flooding might not refer directly to Sandy by name; it may describe the flooding only as "storm-related" but we have to tie this back to the hurricane itself even when two articles don't share a common language. There's going to be more paraphrasing also as journalists and writers, to avoid being repetitive after days of writing about the same topic, change up their sentences. It makes it harder for language tools that are looking for the same phrases and words.

Determining semantic relatedness is obviously the key, but we're going to need to build new language tools and approaches that don't rely on the explicit presence of shared terms.

How will "Describing Disasters" find personal stories?

That's one question, but the more interesting question is how do you recognize a good, compelling story people would want to hear? Not a lot of people have looked at this. While there is work on what makes scientific writing good, recognizing what makes a story compelling is new research.

We're starting by investigating a number of theories drawn from linguistics and from literature on what type of structure or features are typically found in narratives. We'll be looking especially at the theory of the sociolinguist William Labov of the University of Pennsylvania who has been looking at the language that people use when telling stories. There is often an orientation in the beginning that tells you something about location, or a sequence of complicating actions that culminates in an event, which Labov calls the most reportable event—something shocking or involving life and death for which you tend to get a lot of evaluative material.

One student is now designing a most-reportable-event classifier but in a way that is not disaster-specific. We don't want to have to enumerate it for different events, to have to explicitly state that the most reportable event for a hurricane is flooding, and that it's something different for a tornado or a mass shooting.

What are the hard problems?

Timelines that summarize changes over time, that describe how something that happened today is different from what we knew yesterday. In general that's a hard problem. On one hand, everything is different, but most things that are different aren't important. So how do we know what's new and different that is important and should be included? Some things that happen on later days may be connected to the initial event, but not necessarily. How do we tell?

Being able to connect sub-events triggered by an initial event can be hard for a program to do automatically. Programs can't see the correlation necessarily. We're going to have to insert more intelligence for this to happen.

Here's one example. There was a crane hanging over mid-town Manhattan after Hurricane Sandy came in. If we were using our normal expectations of hurricane-related events, we wouldn't necessarily think of a crane dangling over a city street from a skyscraper. This crane sub-event spawned its own sub-event, a terrible traffic jam. Traffic jams happen in New York city. How would we know that this one traffic jam was a result of Hurricane Sandy and not some normal everyday event in New York?

It does seem like an interesting problem. How will you solve it?

We don't know, at least not now. Clearly geographic and temporal information is important, but large-scale disasters can encompass enormous areas and the effects go on for weeks. We need to find clues in language to draw the connections between different events.

We're just getting started. And that's the fun of research, the challenge of the difficult. You start with a question or a goal without an easy answer and you work toward it.

There are other problems, of course. The amount of data surrounding a large-scale disaster is enormous. Somehow we'll have to wade through massive amounts of data to find just what we need. Validation is another issue. How do we know if we've developed something that is correct? We'll need metrics to evaluate how well the system is working.

Another issue will be reining in everything we will want to do. There is so much opportunity with this project, particularly for multidisciplinary studies. We can easily see pulling in journalism students, and those working in visualization. To be able to present the information in an appealing way will make the system more usable to a wider range of people, and it may generate new ways of looking at the data. We see this project as a start; there is a lot of potential for something bigger.

Want the attention of the audience? **Introduce a new gesture.**

Professor of Computer Science **John R. Kender**

John Zhang

Correlating Speaker Gestures in Political Debates with Audience Engagement Measured via EEG, *a paper presented earlier this month at ACM Multimedia, describes the work done by Columbia University researchers to identify what gestures are most effective at getting an audience to pay attention. Using EEG data of study participants watching clips of the 2012 presidential debates, researchers found atypical, or extremal gestures to be the strongest indication of listeners' engagement. This finding not only benefits speakers but may lead to methods for automatically indexing video.*

When people talk, they often gesture to add emphasis and help clarify what they are saying. People who speak professionally—lecturers, teachers, politicians—know this either instinctively or learn it from experience, and it's supported by past studies showing that particular gestures are used for various semantic purposes.

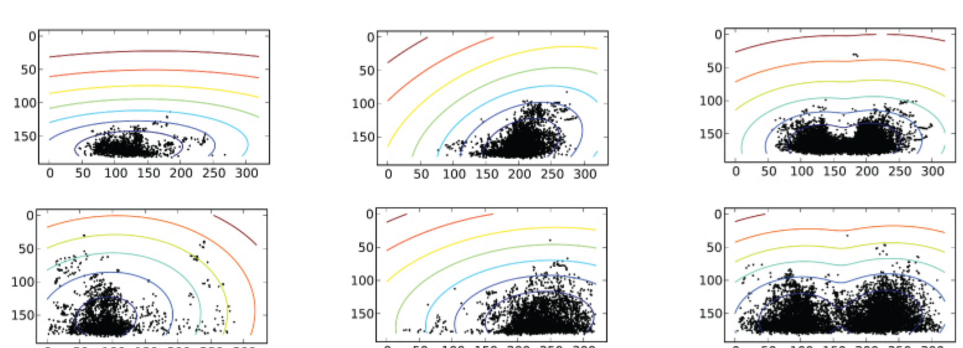
But what gestures are most effective in getting listeners to pay close attention to what a speaker is saying?

To find out, Columbia University computer science and biomedical engineering researchers motivated by **John Zhang**, a PhD student of Prof. **John R. Kender**, set up an experiment to correlate the level of audience interest with specific hand gestures made by a speaker. The researchers asked 20 participants—equally divided along gender and political party lines—to view video clips of the 2012 presidential debates between Obama and Romney. The intent was to track the candidates' gestures with the level of interest exhibited by the audience.

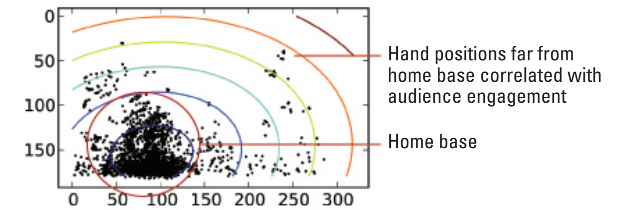
Gestures were automatically extracted from the video using an algorithm that incorporated computer-vision techniques to identify and track hand motions and distinguish left and right hands even when the hands were clasped. The algorithm also detected gestural features the researchers suspected beforehand might correlate with audience attention, specifically velocity but also change in direction and what is called an invitation feature, where the hands are spread in an open manner. (In previous work studying teacher gestures, the authors were the first to identify this type of gesture.)

Another feature was likelihood of hand position. Though gestures vary greatly from person to person, individuals have their own habits. They may move their hands up and down or left and right, but they tend to do within a defined area, or home base, only occasionally moving their hands outside the home base. These non-habitual, or *extremal*, poses became an additional feature.

Mapping hand locations, first debate



Each point indicates a hand location (sampling was done every 1 second). Curves indicate the probability of a gesture occurring at a specific distance (in pixels) from the candidate's home base.



Points far from "home base" correlated with heightened levels of listener attention (here, Romney's left hand in the first debate).

From tracking gestures, it was immediately apparent that Romney used his hands more and made more use of extremal gestures. This was true in both debates 1 and 3. (The study did not look at debate 2 where the format—candidates stood and walked while holding a microphone—could bias the gestures.) Obama, widely criticized for his lackluster performance in debate 1, gestured less and made zero extremal gestures in the first debate. If gestures alone decided the debate winner, Obama lost debate 1. (He would improve in the third debate, probably after coaching.)

For gauging the level of interest, researchers chose EEG data since it is a direct, non-intrusive measure of brain activity already known to capture information related to attention. Electrodes were attached to the scalp of the 20 study participants to record brain activity while they watched debate clips. The data capture was carried out in the lab of Paul Sajda, a professor in the Biomedical department, who also helped interpret the data. (EEG data is not easy to work with. The signals are weak and noisy, and the capture process itself requires an electrostatically shielded room. The data was also voluminous—46 electrodes for 20 participants tracked for 47 minutes with 2000 data points per second—and was reduced using an algorithm written for the purpose.)

While the EEG data showed many patterns of activity, there was a lot of underlying signal, from which researchers identified three main components. Two corresponded to specific areas of the brain: the first to electrodes near the visual cortex, with more activity suggesting that audience members were actively watching a candidate. The second component was generated near the prefrontal cortex, the site of executive function and decision-making, indicating that audience members were thinking and making decisions. (It was not possible to specify a source for the third component.)

Once time-stamped, the EEG data, averaged across subjects, was aligned with gestures in the video so researchers could locate statistically significant correlations between a gesture feature (direction change, velocity, and extremal pose) and a strong EEG component. Moments of engagement were defined as a common neural response across all subjects. Responses not shared with other subjects might indicate lack of engagement (i.e., boredom), as each participant would typically focus on different stimuli.

Extremal gestures turned out to be the strongest indication of listeners' engagement. No matter how researchers subdivided the participants—Democrats, Republicans, females, males, all—what really triggered people's attention was something new and different in the way a candidate gestured.

This finding that extremal poses correlate with audience engagement should help speakers stress important points.

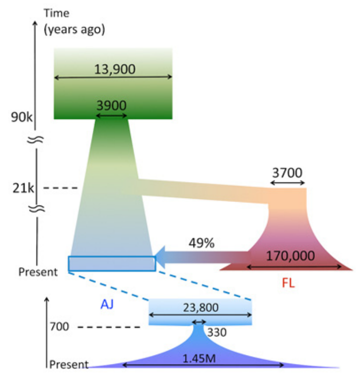
And it may also provide an automatic way to index video, an increasingly necessary task as the amount of video continues to explode. Video is hard to chunk meaningfully, and video indexing today often relies on a few extracted images and standard fast-forwarding and reversing. An algorithm trained to find extremal speaker gestures might quickly and automatically locate video highlights. Students of online courses, for example, could then easily skip to the parts of a lecture needing the most review.

More work is planned. The researchers looked only at correlation, leaving for future work the task of prediction, where some EEG data is set aside to see if it's possible to use gestures to predict where there is engagement. Whether certain words are more likely to get audience reaction is another area to explore. In a small step in this direction, the researchers looked at words with at least 15 occurrences, finding that two words, business and (interest) rate, were unusual in their ability to draw attention. Though the data is insufficient for any definite conclusions, it does suggest potentially interesting results.

Nor did researchers look at what the hand was doing, whether it was pointing, forming a fist, or posed in another manner. The current work focused only on hand position, but this was enough to show that gesturing can effectively engage and influence audiences.

By Linda Crane

Mapping the DNA Sequence of Ashkenazi Jews



Reconstruction history of Ashkenazi Jews and Europeans indicates the former to descend from a small, "bottleneck" group, which is a mix between Europeans and Middle Easterners.

Led by **Itsik Pe'er**, associate professor of computer science at Columbia Engineering, a team of researchers has created a data resource that will improve genomic research in the Ashkenazi Jewish population and lead to more effective personalized medicine. The team, which includes experts from 11 labs in the New York City area and Israel, focused on the Ashkenazi Jewish population because of its demographic history of genetic isolation and the resulting abundance of population-specific mutations and high prevalence of rare genetic disorders. The Ashkenazi Jewish population has played an important role in human genetics, with notable successes in gene mapping as well as prenatal and cancer screening. The study was published online on *Nature Communications* today.

"Our study is the first full DNA sequence dataset available for Ashkenazi Jewish genomes," says Pe'er, who is also a co-chair of the Health Analytics Center at Columbia's Institute for Data Sciences and Engineering, as well as a member of its Foundations of Data Science Center. "With this comprehensive catalog of mutations present in the Ashkenazi Jewish population, we will be able to more effectively map disease genes onto the genome and thus gain a better understanding of common disorders. We see this study serving as a vehicle for personalized medicine and a model for researchers working with other populations."

To help in his hunt for disease genes, Pe'er founded The Ashkenazi Genome Consortium (TAGC) in September 2011 with **Todd Lencz**, an investigator at The Feinstein Institute for Medical Research, director of the Laboratory of Analytic Genomics at the Zucker Hillside Hospital, and associate professor of molecular medicine and psychiatry at the Hofstra North Shore-LIJ School of Medicine. The other TAGC members, who are providing expertise in the diseases they are studying, are:

- **Gil Atzmon**, associate professor of medicine and genetics, Albert Einstein College of Medicine (genetics of longevity and diabetes);
- **Lorraine Clark**, associate professor of clinical pathology and cell biology and co-director, Personalized Genomic Medicine Laboratory, Columbia University Medical Center, **Laurie Ozelius**, associate professor at Icahn School of Medicine at Mount Sinai, and **Susan Bressman**, chair of neurology at Mount Sinai Beth Israel (Parkinson's disease and related neurological phenotypes);
- **Harry Ostrer**, professor of pathology, genetics, and pediatrics, Albert Einstein College of Medicine (radiogenomics, cancers and rare genetic disorders);
- **Ken Offit**, chief of clinical genetics at Memorial Sloan Kettering Cancer Center (breast, ovarian, colon and prostate cancers, lymphoma);
- **Inga Peter,** associate professor of genetics and genomic sciences, and **Judy Cho**, professor of medicine and professor of genetics and genomic sciences, both at The Mount Sinai Hospital (inflammatory bowel disease); and
- **Ariel Darvasi**, vice-dean of The Faculty of Life Sciences at The Hebrew University of Jerusalem (multiple diseases).

Before the TAGC study, data was available for a limited number of DNA markers (only approximately one in every 3000 letters of DNA) that are mostly common in Europeans. The TAGC researchers performed high-depth sequencing of 128 complete genomes of Ashkenazi Jewish healthy individuals. They compared their data to European samples, and found that Ashkenazi Jewish genomes had significantly more mutations that had not yet been mapped. Pe'er and his team analyzed the raw data and created a comprehensive catalog of mutations present in the Ashkenazi Jewish population.

The TAGC database is already proving useful for clinical genomics, identifying specific new mutations for carrier screening. Lencz explains: "TAGC advances the goal of bringing personal genomics to the clinic, as it tells the physician whether a mutation in a patient's genome is shared by healthy individuals, and can alleviate concerns that it is causing disease. Without our work, a patient's genome sequence is much harder to interpret, and more prone to create false alarms. We have eliminated two thirds of these false alarms."

The TAGC study further enables more effective discovery of disease-causing mutations, since some genetic factors are observable in Ashkenazi individuals but essentially absent elsewhere. Moreover, the demography of the Ashkenazi population, the largest isolated population in the U.S., enables large-scale recruitment of study patients and hence more genetic discoveries than in other well-known isolated populations like the Amish and Hutterites locally, or the Icelanders overseas. The researchers expect that medical insights from studies of specific populations will also be relevant to general populations as well.

The TAGC team's findings also shed light on the long-debated origin of Ashkenazi Jews and Europeans. The genetic data indicates that the Ashkenazi Jewish population was founded in the late medieval times by a small number, effectively only hundreds of individuals, whose descendants expanded rapidly while remaining mostly isolated genetically.

"Our analysis shows that Ashkenazi Jewish medieval founders were ethnically admixed, with origins in Europe and in the Middle East, roughly in equal parts," says Shai Carmi, a post-doctoral scientist who works with Pe'er and who conducted the analysis. "TAGC data are more comprehensive than what was previously available, and we believe the data settle the dispute regarding European and Middle Eastern ancestry in Ashkenazi Jews. In addition to illuminating medieval Jewish history, our results further pave the way to better understanding European origins, millennia before. For example, our data provides evidence for today's European population being genetically descendant primarily from late mid-eastern migrations that took place after the last ice age, rather than from the first humans to arrive to the continent, more than 40,000 years ago."

The researchers note that their results also raise a number of hypotheses regarding Ashkenazi origin, which they have now begun exploring. They are looking at, for example, the identity of the Ashkenazi founders: who are the modern populations most similar to them in Europe and in the Middle East? When did the admixture process take place and to what extent was that process sex-biased, where men with one origin mated with women from another?

Pe'er adds, "Equally important to us is making our data available to the entire research community. We've released it to public-access databases and fully expect the creativity of the scientific world to come up with additional uses for the data. What's especially gratifying is the idea that our work will pave the way for personalized genomics in other populations as well."

Pe'er and his TAGC colleagues are already working on their next step: to study specific diseases in the Ashkenazi population, including schizophrenia, Parkinson's, Crohn's, diabetes, and cancer, as well as other inherited traits such as longevity. They are also working with additional collaborators, including the New York Genome Center, to sequence approximately 500 more Ashkenazi genomes to make the catalog of mutations even more comprehensive.

The study was funded by:

- Columbia Engineering: Human Frontier Science Program
- Collaborating labs: NIH, NSF, Glenn Foundation for the Biology of Aging; Rachel and Lewis Rudin Foundation; North Shore-LIJ Health System Foundation, Brain & Behavior Foundation; US-Israel Binational Science Foundation; New York Crohn's Disease Foundation; Edwin and Caroline Levy and Joseph and Carol Reich; the Parkinson's Disease Foundation; the Sharon Levine Corzine Cancer Research Fund; and the Andrew Sabin Family Research Fund.

By Holly Evarts

Professors Nayar Elected to the **National Academy of Inventors**

Shree Nayar, T.C. Chang Professor of Computer Science, has been elected a fellow of the National Academy of Inventors (NAI) for demonstrating "a highly prolific spirit of innovation in creating or facilitating outstanding inventions that have made a tangible impact on quality of life, economic development, and the welfare of society." He is among 170 new fellows who will be inducted during the NAI's 4th Annual Conference on March 20, 2015, at the California Institute of Technology in Pasadena, CA.

"It is wonderful to be an academic and be recognized as an inventor," Nayar says. "I am honored to be a part of such a remarkable group of innovators."

Nayar, who directs the Computer Vision Laboratory, has published over 200 scientific articles and has been awarded more than 40 patents for his inventions related to digital imaging, computer vision, human-computer interfaces, and robotics. Today, his research results are widely used in cameras for smart phones, industrial vision systems for factory automation, and rendering engines for computer graphics. He is the recipient of the prestigious David Marr Prize in 1990 and 1995, the David and Lucile Packard Fellowship in 1992, the National Young Investigator Award in 1993, and the NTT Distinguished Scientific Achievement Award in 1994. He was elected to the National Academy of Engineering in 2008 and the American Academy of Arts and Sciences in 2011 for his pioneering work on computational imaging and his seminal contributions to physics-based computer vision. In 2006, he received the Columbia Great Teacher Award from the Society of Columbia Graduates.

By Holly Evarts

Changxi Zheng **Wins NSF CAREER Award**

Assistant Professor
Changxi Zheng

Changxi Zheng is the recipient of a National Science Foundation (NSF) CAREER Award for his proposal to create realistic, computer-generated sounds for immersive virtual realities.

The five-year, $500,000 grant will fund his proposal "Simulating Nonlinear Audiovisual Dynamics for Virtual Worlds and Interactive Applications."

"I am honored by the award and excited about the research it supports," says Zheng, who is co-directing the Columbia Computer Graphics Group (C2G2) in the Columbia Vision + Graphics Center. "The algorithms and tools used today for constructing immersive virtual realities are inherently appearance-oriented, modeling only geometry or visible motion. Sound is usually added manually and only as an afterthought. So there is a clear gap."

While computer-generated imagery has made tremendous progress in recent years to attain high levels of realism, the same efforts have not yet been applied to computer-generated sound. Zheng is among the first working in the area of dynamic, physics-based computational sound for immersive environments, and his proposal will look to tightly integrate the visual and audio components of simulated environments. It will represent a change from today's practices where digital sound is usually created and recorded separately of the action and then dropped in when appropriate. In Zheng's proposal, computational sound will be automatically generated using physics-based simulation methods and fully synchronized with the associated motion.

"What I propose to do about sound synthesis is analogous to the existing image rendering methods for creating photorealistic images," says Zheng. "We would like to create audiovisual animations and virtual environments using computational methods."

"In addition to the realism, sound that is not tightly coupled with motion loses some of its impact and detracts from the overall immersive experience," says Zheng. "Attaining any type of realism in virtual worlds requires synchronizing the audio and visual components. What the user hears should spring directly and naturally from what the user sees."

It will take new mathematical models and new computational methods. Sound is a physical phenomenon, and creating sound by computer requires understanding and simulating all the motions and forces that go into producing sound. Computational methods will have to replicate everything from the surface vibrations on an object that produce the pressure waves we hear as sound, while taking into account how frequency, pitch, and volume are affected by the object's size, shape, weight, surface textures, and countless other variables. The way sound propagates differs also, depending on whether sound waves travel through air or water and what obstructions, or other sound waves, get in the way. It's a dynamic situation where a slight adjustment to one variable produces nonlinear changes in another.

Zheng's system will have to tackle not only the physical phenomena but do so in a computationally efficient way so there is no delay between an action and the resulting sound. "Realistic virtual environments require a great many complex computations. We will need fast, cost-effective algorithms. We need also to identify what calculations can be approximated while still maintaining realism."

The computational sound is just the beginning. Zheng foresees his proposal as laying the groundwork for building interactive applications and spurring new techniques for efficient digital production and engineering design. Education outreach for disseminating the information learned from the research is also part of the proposal, with workshops planned for high-school students to encourage their interest in STEM fields, and with college courses for teaching undergraduate and graduate researchers how to build audiovisual computational technologies.

For now, numerous challenges exist toward computationally synthesizing realistic sound but if the research is successful, it will enable new virtual-world applications with fully synchronized audiovisual effects, lead to new ways of creating and editing multimedia content, and bring computer-generated sound into the future.

Roxana Geambasu **Scores a "Brilliant 10"**

Assistant Professor
Roxana Geambasu

Roxana Geambasu has been named one of *Popular Science*'s "Brilliant Ten", the magazine's 13th annual list of the "brightest young minds in science and engineering."

Noting that these top 10 researchers are already changing the world, *Popular Science* cites Geambasu for her research that "teaches the cloud to forget personal data" and builds "software that allows the public to see where the information they share goes."

"It is very exciting and humbling to get this honor," says Geambasu. "I am flattered to be included in such an inspiring group."

Geambasu's work is focused on ensuring data security and privacy in an era of cloud computing and ubiquitous mobile devices, all technologies upon which billions of users rely to access and host sensitive data—emails, documents, financial information, etc.—and which have become easy targets for theft, espionage, hacking, and legal attacks. As she has pointed out, the mobile devices we carry with us everywhere 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 notes that this has become an untenable state and 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.

"My research aims to forge a new, more privacy-conscious future where users are better aware of the implications of their online actions and Web services and applications are built with users' privacy in mind," Geambasu adds.

She is the second professor at the School to make the magazine's annual list. In 2011, fellow Computer Science Associate Professor **Eitan Grinspun** was named to the "Brilliant 10".

Geambasu, who joined Columbia Engineering in 2011, has won a number of awards, including a prestigious Microsoft Research Faculty Fellowship and an NSF CAREER award, both in 2014, and the first Google Fellowship in Cloud Computing in 2009. She received her BS from Polytechnic University of Bucharest (Romania) in 2005; her MS (2007) and PhD (2011) from the University of Washington.

In order to arrive at the "Brilliant Ten", *Popular Science* combs through hundreds of nominations from around the country and vets the most promising candidates with experts working in those fields.

"*Popular Science* has always looked optimistically towards the future, and there's no more promising sign of what that future will bring than the inspired work of young scientists and engineers across the country," says Editor-in-Chief Cliff Ransom. "The 'Brilliant Ten' is a celebration of the best and brightest."

By Holly Evarts

Henning Schulzrinne Named **ACM Fellow** and Recognized by **ITC** with **Service Award**

Julian Clarence Levi
Professor of
Mathematical Methods
and Computer Science
Henning Schulzrinne

The two awards are testament to his lasting achievements and the immense impact of VoIP and related technologies on society

Henning Schulzrinne has been named a 2014 ACM (Association for Computing Machinery) Fellow in recognition of "contributions to the design of protocols, applications, and algorithms for Internet multimedia." He is one of 47 to be so recognized in 2014 for contributions to computing.

"I'm deeply honored to be named ACM Fellow. Indirectly, this also recognizes the contributions of all the IRT lab members over the years," says Schulzrinne, in referring to the Internet Real-Time Lab (IRT) lab, which he directs and which conducts research in the areas of Internet and multimedia services.

It is the second honor accorded Schulzrinne in as many months. In December he received an Outstanding Service Award by the Internet Technical Committee (ITC), a joint committee of the Internet Society and the IEEE Communications Society. The award was given in recognition of excellent and extended service rendered to the fields of the ITC, of which Schulzrinne was the founding chair. He served from 1994 to 2000 and helped shape the agenda of the committee in its early years.

The ITC award, being given for the first time, acknowledges both Schulzrinne's service to the ITC as well as service to the community through his foundational work on the key protocols Session Initiation Protocol (SIP) and Real-Time Transport Protocol (RTP) that led to the development of Voice-over-IP (VoIP).

The two awards are testament to his lasting achievements and the immense impact of VoIP and related technologies on society, and they follow on Schulzrinne's induction into the Internet Hall of Fame in 2013. Previous awards include the New York City Mayor's Award for Excellence in Science and Technology and the VON Pioneer Award. He is also an IEEE Fellow.

In addition to his research, Schulzrinne is active in public policy and in serving the broader technology community. From 2011 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. He continues to serve as a technical advisor to the FCC. Schulzrinne is also 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 70 Internet Requests for Comment.

Currently a professor within Columbia's Computer Science department, Schulzrinne continues to work on VoIP and other multimedia applications. Ongoing projects look to create an overall architecture for the Internet of Things, limit phone spam ("robocalls"), make it easier to diagnose network problems, and contribute to a better understanding between engineers and public policy.

Bjarne Stroustrup Awarded **AITO Dahl-Nygaard Prize**, and Named a Fellow of the **Computer History Museum**

Visiting Professor in
Computer Science
Bjarne Stroustrup

Bjarne Stroustrup, a visiting professor at Columbia Engineering, is the senior award winner of the 2015 AITO (Internationale pour les Technologies Objets) Dahl-Nygaard Prize for the design, implementation, and evolution of the C++ programming language. For this same achievement, Stroustrup has also been made a Fellow of the Computer History Museum.

This prize, established in 2005 and one of the most prestigious in the area of software engineering, is named for Ole-Johan Dahl and Kristen Nygaard for their work in creating Simula, the first object-orientated language.

"I feel particularly honored by this award because I knew Ole-Johan Dahl and Kristen Nygaard," says Stroustrup. "While still a student in the University of Aarhus, I learned object-oriented programming from Kristen Nygaard through long discussions."

Stroustrup was influenced by the object-oriented model of Simula when he first began work on C++ in 1979. Then at Bell Labs, he needed a language that would provide hardware access and high performance for systems programming tasks while also handling complexity. Since no such language then existed, he designed one by essentially building on top of C to add support for object-oriented programming, data abstraction, and other capabilities.

In 1985, C++ was commercially released and spread rapidly, becoming the dominant object-oriented programming language in the 1990s and one of the most popular languages ever invented.

Fittingly Stroustrup is the recipient of many awards, honorary degrees, and other recognitions. He is a member of the National Academy of Engineering and was previously made a Fellow of both the IEEE and the ACM.

Just recently, he was named also a Fellow of the Computer History Museum to recognize the impact and importance of C++ in computer history.

He is a prolific writer and contributor. His publications include several books—*The C++ Programming Language* (Fourth Edition, 2013), *A Tour of C++* (2013), *Programming: Principles and Practice using C++* (2014)—as well as a long list of academic and general-interest publications that Stroustrup maintains here.

Stroustrup continues to update and add functionality to C++. Even as new languages have been created for the rapidly shifting programming landscape, C++ remains widely used, particularly in large-scale and infrastructure applications such telecommunications, banking, and embedded systems.

By Linda Crane

Steve Bellovin Elected to the **National Cyber Security Hall of Fame**

Percy K. and Vida L. W.
Hudson Professor of
Computer Science
Steve Bellovin

Computer Science Professor **Steve Bellovin**, one of the foremost experts on Internet security and privacy, has been elected to the National Cyber Security Hall of Fame for his research on networks, security, and public policy issues. Bellovin joins a crop of top industry leaders as the 2014 class of inductees, to be honored at a banquet on October 30th. Bellovin coauthored one of the first books on firewalls, *Firewalls and Internet Security: Repelling the Wily Hacker*, and holds a number of patents on cryptographic and network protocols. A graduate of Columbia College, he joined the Columbia Engineering faculty in 2005 following a distinguished career at Bell Labs and AT&T Research Labs, where he was an AT&T Fellow. Bellovin was elected a fellow of the National Academy of Engineering and for many years has served as a member of specialized technology advisory committees, including for the Department of Homeland Security, Election Assistance Commission, and the National Academies. From 2012 to 2013, he held the post of chief technologist at the Federal Trade Commission.

In addition to this honor, Bellovin is now the Percy K. and Vida L. W. Hudson Professor of Computer Science. The Hudson chair was established by Mr. Hudson, a descendent of Henry Hudson, who graduated from Columbia School of Mines in 1899.

Julia Hirschberg Elected to the **American Philosophical Society**

Percy K. and Vida L. W. Hudson Professor of Computer Science and Department Chair **Julia Hirschberg**

Julia Hirschberg, Percy K. and Vida L. W. Hudson Professor of Computer Science and Department Chair, has been elected to the American Philosophical Society (APS), in recognition of her contributions to spoken language processing. She is one of 33 scientists, humanists, social scientists, and leaders in civic and cultural affairs elected to APS this year, including Jill Lepore, James Levine, Alice Waters, and Rem Koolhaas.

"I'm delighted and very honored to be joining such a distinguished group," says Hirschberg. "I am definitely looking forward to the APS meetings.

Founded by Benjamin Franklin in 1743, APS's mission is to promote useful knowledge in the sciences and humanities through excellence in scholarly research, professional meetings, publications, library resources, and community outreach. Only 5,474 members have been elected since its founding.

APS members are nominated by current members and have included a broad range of illustrious scholars and national leaders, from George Washington, John Adams, and Thomas Jefferson to John James Audubon, Charles Darwin, Louis Pasteur, Thomas Edison, Marie Curie, Linus Pauling, Margaret Mead, and Robert Frost. Contemporary luminaries include Warren Buffet, Juan Carlos of Spain, Jimmy Carter, and Neil Armstrong, and Columbia members University President Lee C. Bollinger and Professors Joseph Stiglitz and Eric Kandel.

By Holly Evarts

Xi Chen Named 2015 Recipient of the **Presburger Award**

Assistant Professor **Xi Chen**

Xi Chen was named this year's recipient of the Presburger Award. Since 2010, the award has been bestowed by the European Association for Theoretical Computer Science (EATCS) to "a young scientist for outstanding contributions in theoretical computer science, documented by a published paper or a series of published papers." The award is named after Mojzesz Presburger, who while still a student in 1929 invented the field of Presburger arithmetic, whose simplicity—it involves only addition of natural numbers—means Gödel's Incompleteness Theorem does not apply to it.

Chen was selected for his fundamental contributions in a variety of areas within theoretical computer science. "His work in algorithmic game theory and computational economics includes the answer to the long-standing question about the computational complexity of Nash equilibria for two-player games, showing PPAD-completeness. For classes of markets and types of utility functions widely used in economics, he settled the complexity of market equilibria, again showing PPAD-completeness. His work on complexity theory includes a complete dichotomy theorem for partition function computation, showing it to be either polynomial or #P-complete, as well as for counting constraint satisfaction problems with complex weights, general concepts that include e.g. counting graph homomorphisms. His work on algorithms includes a proof that isomorphism of strongly regular graphs, a well-known hard case for the graph isomorphism problem, can be tested in time exponential in $n^{1/5}$—the first significant progress in more than a decade."

It is unusual for someone so young (Chen was born in 1982) to have already written papers that contributed to many different aspects of theoretical computer science.

"I feel honored to receive this award and be included in the company of the previous winners whose work I know and respect," said Chen.

Chen joined the Columbia Computer Science Department in 2011. He attended Tsinghua University where he received a BS degree in Physics/Math (2003) and a PhD in Computer Science (2007).

By Linda Crane

Professor **Al Aho** received the Alumni Association Distinguished Faculty Teaching Award for outstanding teaching. The award was presented on Class Day, May 19, 2014. Alfred will also receive an honorary degree from the University of Toronto during convocation ceremonies in 2015 as a world-leading computer scientist for his groundbreaking research on programming languages, compilers and algorithms.

Professor **Steven Feiner** was a keynote speaker at SIGGRAPH Asia 2014 in Shenzhen, China. He talked about Augmented Reality (AR) and its ability to change our lives. He discussed how AR could augment a world beyond our own first-person perspective, assist us as we work and play together, and create hybrid user interfaces for multiple display and interaction devices that complement each other.

Professor **Jonathan Gross** was awarded a 5-year Simons Foundation Collaboration Grant in support of his research on computational aspects of low-dimensional topology. The Simons Foundation is a private foundation that funds research in mathematics and the physical sciences. It was founded by James Simons, a well-known mathematician, hedge-fund manager, and philanthropist.

Jonathan Bell

Prof. **Gail Kaiser**

PhD student **Jonathan Bell** and Professor **Gail Kaiser** received an ACM SIGSOFT Distinguished Paper Award at the 36th International Conference on Software Engineering (ICSE). The paper, "Unit Test Virtualization with VMVM", describes a technique for significantly speeding up the software testing process, which allows test cases to be executed more frequently during software development, and hence allowing bugs to be detected and repaired earlier. The technique was shown to decrease the amount of time needed to run tests by an average of 62%, significantly reducing the time and computational cost of testing software.

EE PhD student Hyunwoo Nam, EE Adjunct Professor Doru Calin (Director, High Performance Wireless Technologies and Networks at Alcatel Lucent Bell Labs), and Professor **Henning Schulzrinne** received the best paper award at IEEE WCNC 2015 for their paper, "Intelligent Content Delivery over Wireless via SDN". The paper describes how end-to-end Internet services can be achieved in mobile networks such as 4G using software-defined networking (SDN). They introduced high-level architecture, possible use-cases, and experimental results in Wi-Fi networks to show the feasibility of their approaches.

Professor **Simha Sethumadhavan** was appointed by FCC as a member of the Downloadable Security Technology Advisory Committee (DSTAC) on January 27, 2015, which will develop interoperability standards for TV navigation devices.

Dr. Lisa Wu

Andrea Lottarini

Timothy Paine

Prof. **Martha Kim**

Prof. **Kenneth Ross**

A paper by Dr. **Lisa Wu**, PhD student **Andrea Lottarini**, MS student **Timothy Paine**, Professor **Martha Kim** and Professor **Kenneth Ross** was selected as one of IEEE Micro's 12 Top Picks from across all conferences in computer architecture. The paper is titled "Q100: The Architecture and Design of a Database Processing Unit", and describes an improved processor for relational data analytics.

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